

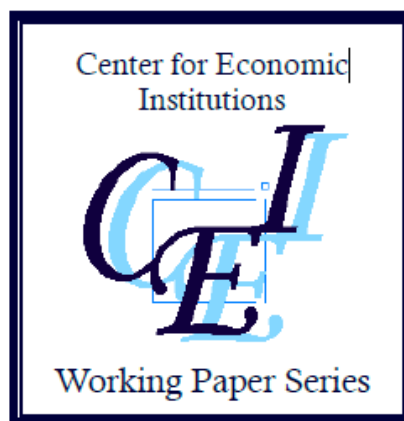
**Center for Economic Institutions**  
**Working Paper Series**

No. 2025-4

**“Product Market Competition, Labor Mobility, and Firm-Sponsored Training: New Implications of Market Power”**

**Arghya Ghosh, Hodaka Morita and Susumu Sato**

March, 2026



Institute of Economic Research  
Hitotsubashi University  
2-1 Naka, Kunitachi, Tokyo, 186-8603 JAPAN  
<https://cei.ier.hit-u.ac.jp/English/index.html>  
Tel:+81-42-580-8405/Fax:+81-42-580-8333

# Product Market Competition, Labor Mobility, and Firm-Sponsored Training: New Implications of Market Power\*

Arghya Ghosh<sup>†</sup>

Hodaka Morita<sup>‡</sup>

Susumu Sato<sup>§</sup>

## Abstract

Firms compete in both product and labor markets by making decisions about hiring, training, and poaching workers. We develop a theoretical model in which firm-sponsored training links product and labor market competition. Changes in labor, product, or overall market power influence firms' incentives to invest in training, which can lead to clear-cut welfare improvements benefiting all relevant parties. The distinction between under- vs. overinvestment in training, a unique feature that emerges from the interaction between the two markets, plays a critical role in determining whether or not such welfare improvements are possible, enriching welfare and policy implications.

**Keywords:** product market competition, oligopoly, market power, labor mobility, training, welfare, antitrust implications

**JEL Codes:** D21, L13, L40, M50

---

\*We thank Simon Anderson, Jed DeVaro, Wei Li, Kieron Meagher, Volker Nocke, Mike Waldman, Yijiang Wang, and Gang Zhang as well as conference and seminar participants at Asia-Pacific Industrial Organization Conference 2022, Organization Economics Workshop 2022, Cheung Kong Graduate School of Business, Fukushima University, Hitotsubashi University, Keio University, Shinshu University, and the University of Tokyo for helpful comments. This work was supported by JSPS KAKENHI Grant Numbers 16K21741 and 22H00838. Ghosh gratefully acknowledges Chiaki Moriguchi and the Center for Economic Institutions (CEI) at the Institute of Economic Research, Hitotsubashi University, for their hospitality and financial support.

<sup>†</sup>School of Economics, UNSW Business School, University of New South Wales, Sydney, New South Wales, Australia. E-mail: [a.ghosh@unsw.edu.au](mailto:a.ghosh@unsw.edu.au)

<sup>‡</sup>Cheung Kong Graduate School of Business, China; Institute of Economic Research, Hitotsubashi University, Japan. E-mail: [hodakamorita@gmail.com](mailto:hodakamorita@gmail.com)

<sup>§</sup>Institute of Economic Research, Hitotsubashi University, Tokyo, Japan. E-mail: [susumusato.econ@gmail.com](mailto:susumusato.econ@gmail.com)

# 1 Introduction

Firms compete not only in product markets but also in labor markets, where they enjoy market power as both sellers of products and buyers of labor. The effect of firms' product market power on consumers has been the main focus of scholars and practitioners. Attention to labor market power has rapidly increased recently, with growing evidence on firms' wage-setting power (Card, 2022) and the increasing prevalence of contractual restraints on worker mobility, including noncompete and no-poaching agreements. Nevertheless, their competition in the product and labor markets has been mostly analyzed separately; hence, the link between them has remained underexplored.

In this study, we explore the effects of firms' market power on consumers, workers, and firms from new perspectives. To this end, we develop a model that captures the link between product and labor market competition, which is established through firms' incentives to train workers. Firm-sponsored training plays a crucial role in both markets in practice. On the product market side, firm-sponsored training is an important source of firms' productivity and competitive advantage.<sup>1</sup> On the labor market side, provision of training to employees improves their human capital and wages, and firms' incentive to provide training crucially depends on the prospect of retaining workers. Facing these two markets, firms train workers, retain and poach them, and sell their products to consumers.

We consider an industry in which firms offer differentiated products and jobs to consumers and workers, respectively. Workers have idiosyncratic preferences for jobs and face switching costs when they change jobs. Firms train workers to accumulate industry-specific human capital and then compete in the labor market by setting poaching and retention wages. Simultaneously, firms compete in the product market by setting prices. Human capital accumulated by workers lowers production costs and thereby improves firms' profitability in the product market. Switching costs and differential job preferences lend labor market power to firms, while product differentiation underpins firms' product market power.

In our model, training level is determined by product and labor market competition, and firms may either under- or overinvest in training from an industry perspective. Under labor market competition, some workers at a firm get poached, and their human capital contributes to competitors' profits. When firms make training decisions, they ignore this positive externality, which works in the direction of underinvestment in training.<sup>2</sup> Under

---

<sup>1</sup>Almeida and Carneiro (2009) find a substantially high return (8.6%) on firm investments in human capital in the form of formal job training, using a panel of large firms with detailed information on the duration of training, the direct costs of training, and several firm characteristics. See also, for example, Bapna, Langer, Mehra, Gopal and Gupta (2013), Riley, Michael and Mahoney (2017), Mehra, Langer, Bapna and Gopal (2014), Martins (2021), and Morikawa (2021) for related empirical findings.

<sup>2</sup>In the labor economics literature, it is well known that firms tend to underinvest in training in the

product market competition, lower production costs achieved by training induce firms to reduce product prices, which lowers competitors' profits. Each firm ignores the negative impacts of lower prices driven by more intense training. Firms overinvest in training when this negative externality exceeds the positive externality, which happens when the intensity of product market competition, captured by the aggregate diversion ratio in our model, is high relative to labor mobility, and underinvest otherwise.

We explore the welfare implications of firms' labor market power, product market power, and overall market power (which affects both). Labor market power is represented by workers' switching costs in our model. Regarding product market power, we study the effects of the intensity of product market competition, as an inverse measure of the product market power, modeling the intensity of competition in a general form following [Vives \(2008\)](#). As a source of overall market power that influences both labor and product markets, we examine the effects of common or cross ownership—hereafter referred to as *overlapping ownership*—between competing firms.

One might expect that an increase in firms' market power, be it labor, product, or overall market power, would hurt at least one of consumers and workers, implying that a change in market power cannot lead to a clear-cut welfare improvement in the sense of benefiting all relevant parties (consumers, workers, and firms). Our model yields predictions that are consistent with this conventional wisdom *when the level of training is fixed*. An increase in labor market power hurts workers, leaving consumers unaffected as product prices remain unchanged. Regarding product and overall market power, whenever a change in market power induces firms to reduce their outputs due to weaker product market competition, it increases industry profit but hurts consumers, workers, or possibly both.

The endogenous determination of the training level enriches the welfare implications of our model, indicating that a change in market power can lead to a clear-cut welfare improvement and hence can be pro-competitive. Although the detailed mechanism differs by source of market power, the common driving force is its effect on training incentives. When a change in market power increases firms' incentives to train their workers, more intense training reduces production costs, which, in turn, induces firms to produce more. Higher output and lower prices benefit consumers. At the same time, more intense training and higher output together make workers more attractive to firms, increasing both retention and poaching wages. We show that the possibility of a clear-cut welfare improvement, arising from endogenous determination of training, hinges critically on whether the equilibrium training level reflects over- or underinvestment.

---

presence of labor mobility. [Chang and Wang \(1996\)](#), for example, show that firms underinvest in human capital, where the information asymmetry on workers' productivity between their current employer and potential employers can cause an externality distortion in human capital investment.

Regarding labor market power, an increase in switching costs reduces labor mobility, which induces firms to train their workers more intensively. Although the direct effect of higher switching costs hurts workers, the indirect effect of more intense training puts upward pressure on wages. We find that the positive wage effect can overshadow the negative effect of higher switching costs for workers; hence, higher labor market power can benefit both consumers and workers. This finding suggests the possibility of clear-cut welfare improvement, where the distinction between under- and overinvestment plays a critical role in determining whether such an improvement can indeed be achieved. Our model shows that a clear-cut welfare improvement is possible if more intense training induced by higher labor market power mitigates underinvestment. In contrast, when more intense training worsens overinvestment, it is unlikely. Because product market competition is the source of overinvestment, our result highlights the importance of accounting for product market competition when evaluating the effects of labor market power.

Regarding product and overall market power, a change in market power benefits both consumers and workers when it increases industry output. Although the direct effect of higher output reduces firms' profits, more intense training induced by higher output increases profits. We find that the positive effect on profits can overshadow the negative effect, so that a change in market power can lead to a clear-cut welfare improvement. Here, again, the necessary condition for such an improvement is that more intense training induced by a change in market power mitigates underinvestment in training.

Contractual restraints on worker mobility, including noncompete and no-poaching agreements, have become increasingly prevalent recently. Regarding noncompete agreements, as the length of a noncompete agreement increases, workers' switching costs increase, since workers who switch employers must wait longer before they can be employed by competitors of their current employers. In this sense, the switching cost in our model can be interpreted as a reduced-form representation of noncompete agreements. Under this interpretation, our model predicts that stronger enforcement of noncompete agreements would induce firms to train their workers more intensively, and more intensive training can lead to a clear-cut welfare improvement if it mitigates underinvestment in training but such an improvement is unlikely when it worsens overinvestment. Thus, our model indicates, in contrast to the recent final rule issued by the Federal Trade Commission to ban noncompete agreements nationwide, that noncompete agreements should be subject to a case-by-case analysis, wherein product market competition plays a critical role in determining whether noncompete agreements benefit workers or not.

The rest of the paper is organized as follows. Section 2 discusses our paper's relationship to the literature. Section 3 presents our model that captures the link between product and labor market competition. Section 4 analyzes the model to characterize equi-

librium and then shows that firms may underinvest or overinvest in training depending on parameterizations. Section 5 explores the welfare implications of labor market power, product market power, and overall market power. In Section 6, we discuss our model’s policy implications on noncompete agreements and empirical relevance. We also consider two extensions of our model; one incorporates initial hiring, and the other incorporates firm-specific human capital that lowers indirect costs alongside industry-specific human capital that reduces direct costs. Section 7 summarizes and concludes.

## 2 Relationship to the literature

We examine the effects of firms’ market power in both product and labor markets. Product market power lies at the heart of industrial organization economics and antitrust discussions (see, e.g., [Whinston, 2007](#)), whereas labor market power had received less attention in academic literature and policy debates until recently ([Card, 2022](#)).<sup>3</sup>

Recognition of firms’ labor market power has led to increased adoption of imperfect competition models of labor markets. [Bhaskar and To \(1999\)](#) show how imperfect competition models can explain the apparently puzzling labor market phenomenon. [Berger et al. \(2022\)](#) and [Azar and Vives \(2021\)](#) embed oligopsonistic labor markets in a general equilibrium setting. [Berger et al. \(2022\)](#) develop a quantitative model of oligopsony and measure welfare losses due to firms’ oligopsony power. [Azar and Vives \(2021\)](#) combine oligopsony in the labor market with oligopoly in the product market and diversified ownership of firms. They find that an increase in market concentration due to fewer firms or increased diversification depresses the economy by reducing employment, output, real wages, and labor share unless there is an increasing return to scale.

We contribute to the study of market power by analyzing product and labor market competition in a unified framework. A distinguishing feature of ours is endogenous firm-sponsored training, which offers new perspectives on market power by linking product and labor markets. Our model allows us to analyze changes that directly affect firms’ market power in one of the two markets, product or labor, while keeping their power in the other market constant. At the same time, we also consider the effect of overlapping ownership that affects the market power in both markets.

Our work is related to [Bassanini and Brunello \(2011\)](#) and [Gersbach and Schmutzler \(2012\)](#). Like us, these studies also consider product market competition and firm-sponsored training. [Bassanini and Brunello \(2011\)](#) examine the relationship between product market regulation and workplace training in the European context. Monopolistic

---

<sup>3</sup>See [Card, Cardoso, Heining and Kline \(2018\)](#), [Lamadon, Mogstad and Setzler \(2022\)](#), and [Berger, Herkenhoff and Mongey \(2022\)](#) for recent studies that document substantial wage-setting power of firms.

competition prevails in the product market, while the labor market involves collective wage bargaining. [Gersbach and Schmutzler \(2012\)](#) also consider product market competition in their analysis of industry-specific training. Their work provides an explanation as to why firms might provide their workers with industry-specific training even though doing so may intensify wage competition. Labor mobility does not arise in the equilibrium of the models considered in [Bassanini and Brunello \(2011\)](#) and [Gersbach and Schmutzler \(2012\)](#). In contrast, labor mobility does arise in the equilibrium of our model, playing an important role by affecting firms' incentives to train their workers.

We also contribute to recent studies that analyze labor market power arising from noncompete agreements (e.g., [Shi, 2023](#); [Shy and Stenbacka, 2023](#)). [Shy and Stenbacka \(2023\)](#) study the effects of noncompete agreements where firms invest in training junior workers, and trained workers choose whether to remain with their initial employers or switch to a competing employer. [Shi \(2023\)](#) studies optimal regulation of noncompete agreements in a model of labor market competition with on-the-job search and endogenous noncompete agreements. Product market competition, which does not feature in [Shi \(2023\)](#) and [Shy and Stenbacka \(2023\)](#), plays a critical role in our model, which reveals the effects of labor market power on consumers. It also enriches the effects of labor market power on workers by associating them with product market competition.

In our model, labor market power can enhance welfare as it induces firms to invest more intensively in training. This result parallels findings in the R&D literature, where greater product market power—arising, for example, from increased product differentiation or longer patent protection—can promote welfare by encouraging higher R&D investment ([Vives, 2008](#); [Tirole, 1988](#), Ch 10). A distinctive feature of our framework is that labor market power affects not only workers but also consumers, and product market power affects not only consumers but workers, owing to the connection between labor and product market competition mediated by labor turnover and training.

### 3 Model

We consider a model of oligopoly in which firms compete for consumers and workers at the same time. There is a set  $\mathcal{N} = \{1, \dots, n\}$  of single-product firms that produce horizontally differentiated products. Each firm is initially endowed with a unit mass of workers who can perform a job to improve productivity in supplying each product.<sup>4</sup> Firms train their initial employees and then compete in product and labor markets.

---

<sup>4</sup>In Section 6, we consider an extension of our model that incorporates initial hiring.

**Consumers** Consumer demand is represented by a symmetric demand function  $D(\mathbf{p}) = \{D_i(\mathbf{p})\}_{i \in \mathcal{N}}$  with a given price profile  $\mathbf{p} = \{p_i\}_{i \in \mathcal{N}} \in \mathbb{R}^n$ . We assume that  $\partial D_i(\mathbf{p})/\partial p_i < 0$  for all  $i \in \mathcal{N}$  and  $\partial D_i(\mathbf{p})/\partial p_j \geq 0$  for all  $j \neq i$  whenever  $D_i(\mathbf{p}) > 0$ . Consumer surplus is given by the function  $V(\mathbf{p}) = \bar{V} + \text{CS}(\mathbf{p}) \in \mathbb{R}$  that satisfies the envelope condition  $\partial \text{CS}(\mathbf{p})/\partial p_i = -D_i(\mathbf{p})$ . Let  $Q(p) := D_i(p, p, \dots, p)$  be the individual output under symmetric price  $p \in \mathbb{R}$ , and  $Q_p(p) := \partial D_i(p, p, \dots, p)/\partial p_i$  be the own-price derivative of the demand under symmetric price  $p$ . We assume that  $Q'(p) = \sum_{j \in \mathcal{N}} [\partial D_i(p, p, \dots, p)/\partial p_j] \leq 0$  and  $\lim_{p \rightarrow \infty} Q(p) = 0$ , which implies that a simultaneous increase in the prices of all products does not increase industry demand and that demand vanishes as prices rise to infinity. We also assume that the elasticity of demand faced by individual firms,  $\varepsilon_Q(p) := -pQ_p(p)/Q(p)$ , is increasing in  $p$ , which could be viewed as a symmetric oligopoly analog of Marshall's Second Law of Demand.

Table 1 shows the examples of demand functions that satisfy the properties assumed in the model. It includes independent goods, linear demand model of [Shubik and Levitan \(2013\)](#), and Logit demand systems.

|                   | $D_i(\mathbf{p})$  | $\text{CS}(\mathbf{p})$  |
|-------------------|--|--|
| Independent goods | $q(p_i)$   | $\int_{p_i}^{\infty} q(x) dx$  |
| Shubik-Levitan    | $\frac{1}{n} \left[ v - p_i + \frac{\gamma}{1-\gamma} (\bar{p} - p_i) \right]$                       | $\frac{1}{n} \left[ \sum_{i=1}^n \left( -vp_i + \frac{1}{2(1-\gamma)} p_i^2 \right) - \frac{n\gamma}{2(1-\gamma)} \bar{p}^2 \right]$ |
| Logit             | $\frac{\exp\left(\frac{v-p_i}{\mu}\right)}{V_0^C + \sum_{i=1}^n \exp\left(\frac{v-p_i}{\mu}\right)}$ | $\mu \log \left[ V_0^C + \sum_{i=1}^n \exp\left(\frac{v-p_i}{\mu}\right) \right]$  |

Table 1: Examples of demand and indirect utility functions. In Shubik-Levitan specification  $\bar{p} := \frac{\sum_{i=1}^n p_i}{n}$  denotes the average price.

**Workers** Each worker  $l$  who is initially employed by firm  $j$  can choose whether to stay at firm  $j$  or move to firm  $i$  ( $\neq j$ ). Any utility that a worker in firm  $j$  obtains from moving to firm  $i$  is determined by the wage that firm  $i$  offers to workers in firm  $j$ , denoted by  $w_{ij}$ , the worker's cost of moving from firm  $j$  to firm  $i$  ( $\neq j$ ), denoted by  $\sigma_{ji}$ , and the worker's idiosyncratic preference with firm  $i$ , denoted by  $\varepsilon_{jil}$ . Similarly, the utility of each worker  $l$  who stays with the initial employer  $j$  is determined by  $w_{jj}$  and the worker's idiosyncratic preference  $\varepsilon_{jjl}$  with firm  $j$ .

Let  $v_{ji} = w_{ij} - \sigma_{ji}$ , where  $\sigma_{ji} = 0$  if  $i = j$ , and  $\sigma(\geq 0)$  if  $(i \neq j)$ . Here, we assume that every worker incurs the same cost  $\sigma$  to move from one firm to another. We call  $\sigma$  the switching cost and interpret an increase in  $\sigma$  as an increase in firms' labor market power. Assuming, for simplicity, that the moving cost and idiosyncratic preferences can

be measured in monetary terms,  $v_{ji} + \varepsilon_{jil}$  represents the monetary value enjoyed by worker  $l$  who initially belongs to firm  $j$  and works for firm  $i$ . We assume that workers have no outside options so that every worker works for a firm in this industry. Then, worker  $l$  who is initially employed by firm  $j$  works for firm  $i$  if and only if working for firm  $i$  maximizes the worker's utility, that is,  $v_{ji} + \varepsilon_{jil} \geq v_{jk} + \varepsilon_{jkl}$  holds for all  $k \in \mathcal{N}$ .

We assume that  $\varepsilon_{jl} = (\varepsilon_{jkl})_{k \in \mathcal{N}} \in \mathbb{R}^n$  is a random variable that has a symmetric joint distribution function. We then have a real-valued function with  $n$  arguments, denoted by  $L(\cdot)$ , with the following property: Given the vector  $v_j := (v_{jk})_{k \in \mathcal{N}}$ , the probability that the worker in firm  $j$  switches to firm  $i \neq j$  is given by

$$L_{ji} = L(v_{ji}, v_{j-i}) := \Pr(v_{ji} + \varepsilon_{jil} \geq v_{jk} + \varepsilon_{jkl} \text{ for all } k \in \mathcal{N}),$$

where  $v_{j-i} = (v_{jk})_{k \in \mathcal{N} \setminus \{i\}} \in \mathbb{R}^{n-1}$ , and the probability that the worker stays at firm  $j$  is given by

$$L_{jj} = L(v_{jj}, v_{j-j}) := \Pr(v_{jj} + \varepsilon_{jjl} \geq v_{jk} + \varepsilon_{jkl} \text{ for all } k \in \mathcal{N}),$$

where  $v_{j-j} = (v_{jk})_{k \in \mathcal{N} \setminus \{j\}} \in \mathbb{R}^{n-1}$ .

By definition, the function  $L(\cdot)$  is increasing in the first argument and decreasing in all other arguments. Also, because of the symmetry of the distribution of  $\varepsilon_{jl}$ , we have that  $L(\cdot)$  is symmetric. Furthermore, since we assume that there is no outside option, we have  $\sum_{k \in \mathcal{N}} L_{jk} = 1$  for all  $j \in \mathcal{N}$ . The additive separability of  $v_{ji}$  and  $\varepsilon_{jil}$  in worker  $l$ 's utility also implies that for  $v_j \in \mathbb{R}^n$  and  $v'_j \in \mathbb{R}^n$  such that  $v'_{jk} = v_{jk} + x$  for all  $k \in \mathcal{N}$  where  $x \in \mathbb{R}$ ,  $L(v_j) = L(v'_j)$  holds.

We assume that the joint distribution function of  $\varepsilon_{jl}$  satisfies conditions such that  $L(\cdot)$  is twice continuously differentiable in every argument. Furthermore, we make the following assumption regarding the shape of  $L(\cdot)$  function.<sup>5</sup>

**Assumption 1.** *Let  $y := \sigma + w_I - w_O$  be the relative attractiveness of staying at the initial employer and define the functions  $X_I, X_O$  as*

$$X_I(y) := \frac{L_{ii}}{\partial L_{ii} / \partial w_{ii}} \Bigg|_{v_{ii} - v_{ik} = y \text{ for all } k \neq i},$$

$$X_O(y) := \frac{L_{ji}}{\partial L_{ji} / \partial w_{ij}} \Bigg|_{v_{ij} - v_{ii} = -y, \text{ and } v_{ij} - v_{ik} = 0 \text{ for all } k \neq i}.$$

*Then,  $X_I(y)$  is increasing in  $y$ , and  $X_O(y)$  is decreasing in  $y$ .*

To understand the idea behind this assumption, consider the situation in which each firm  $k \in \mathcal{N}$  offers  $w_{kk} = w_I$  to its own initial employees to retain  $L_{kk} = L_I$  of them, and

---

<sup>5</sup>This assumption is part of sufficient conditions for a symmetric equilibrium to be unique and interior.

$w_{kj} = w_O$  ( $j \neq k$ ) to firm  $j$ 's initial employees to poach  $L_{jk} = L_O$  of them. In order for firm  $i$  to retain one additional worker, the firm must increase its retention wage  $w_{ii}$  by  $1/(\partial L_{ii}/\partial w_{ii})$ . Since such an increase in wage applies to all the  $L_{ii}$  retained workers, the extra wage cost for retaining an additional worker is  $L_{ii} \times 1/(\partial L_{ii}/\partial w_{ii})$ . Similarly, the extra wage cost for poaching an additional worker is  $L_{ji} \times 1/(\partial L_{ji}/\partial w_{ij})$ . Assumption 1 implies that, as  $w_I$  increases, retaining an additional worker becomes more costly and poaching an additional worker becomes less costly. Similarly, Assumption 1 also implies that, as  $w_O$  increases, retaining an additional worker becomes less costly and poaching an additional worker becomes more costly. These properties guarantee the uniqueness of the symmetric equilibrium wages.<sup>6</sup>

Regarding worker surplus, we have a real-valued function  $WS(\cdot)$  with  $n$  arguments such that the aggregate surplus of workers who initially belong to firm  $j$  is given by

$$WS(v_j) = \mathbb{E} \left[ \max_{i \in \mathcal{N}} \{v_{ji} + \varepsilon_{jil}\} \right].$$

Due to the continuity of  $L(\cdot)$ , the distribution of  $\varepsilon_{jil}$  has no mass point, which implies that  $WS(\cdot)$  is continuously differentiable.

**Production technology** Our model links product and labor markets as follows: Let  $H_i$  denote the aggregate human capital accumulated for product  $i$ , which is given by  $H_i = \sum_{j \in \mathcal{N}} L_{ji} h_j$ , where  $h_j$  denotes the level of human capital possessed by each worker initially hired by firm  $j$ . Each firm  $j$  must incur the cost of  $\tau(h_j)$  per employer for human capital provision, where  $\tau'(h_j) > 0$  holds for all  $h_j > 0$ .

Suppose that each firm  $i$ 's constant marginal cost is given by a strictly decreasing and weakly convex function  $c(H_i)$  of human capital, where  $c'(H_i) < 0$  and  $c''(H_i) \geq 0$  hold for all  $H \geq 0$ . Furthermore, we assume that  $c(\cdot)$  is not "too convex" in the sense that  $-c'(H)H$  is increasing in  $H$  for all  $H \geq 0$  as stated in the following assumption.

**Assumption 2.** Define the curvature of marginal cost function by  $\chi(H) := -[c''(H)H]/[c'(H)] \geq 0$ . Then  $\chi(H) \leq 1$  for all  $H \geq 0$ .

Assumption 2 says that the reduction in the marginal cost made by an additional worker increases with the level of human capital per worker. This assumption is not critical for the qualitative nature of our results, but it helps us simplify the presentation of results by reducing the number of cases to be considered. A leading example of the

---

<sup>6</sup>Assumption 1 can be regarded as one concerning the distribution of idiosyncratic preferences. Similar assumptions are common in the industrial organization literature (e.g., [Rhodes and Zhou, 2024](#)). In Online Appendix A.2.1, we show that a variety of popular discrete-choice models (e.g., [Anderson, De Palma and Thisse, 1992](#)) satisfy this property.

function  $c(H)$  that satisfies Assumption 2 is  $c(H) = \bar{C} - \frac{\eta}{\alpha}H^\alpha$ , where  $\eta > 0$  and  $\alpha \in (0, 1]$ . In this case,  $\chi(H) = 1 - \alpha < 1$ .

**Strategies and timing of moves** Firm  $i \in \mathcal{N}$  chooses  $\{(w_{ji})_{j \in \mathcal{N}}, h_i, p_i\}$  to maximize its profit, which is given by product-market profit minus wage bill and training cost:

$$\Pi_i = (p_i - c(H_i))D_i(\mathbf{p}) - \sum_j w_{ij}L_{ji} - \tau(h_i). \quad (1)$$

The timing of the moves is given as follows:

1. At the beginning of the game, each firm  $i \in \mathcal{N}$  independently chooses  $h_i$ , which is unobservable by other firms.
2. Without observing  $h_j$  for  $j \neq i$ , each firm  $i \in \mathcal{N}$  chooses  $(w_{ij})_{j \in \mathcal{N}}$  and  $p_i$ .
3. Workers decide whether and which firm to move, and consumers choose whether and which product to purchase.

We adopt the subgame-perfect equilibrium as the solution concept of the model, and we focus on interior symmetric equilibria. To ensure that a symmetric equilibrium, when it exists, is unique and interior, we make the following regularity assumptions.

**Assumption 3** (Regularity assumptions).

- (i)  $Q(c(h)) > 0$  for all  $h > 0$ , and  $\lim_{p \rightarrow \infty} \varepsilon_Q(p) > 1$ .
- (ii)  $\lim_{h \rightarrow 0} \{\tau'(h)/[-c'(h)]\} = 0$ ,  $\tau'(h)/[-c'(h)]$  is increasing in  $h$ , and the inequality  $\lim_{h \rightarrow \infty} \{\tau'(h)/[-c'(h)] - Q(c(h))\} > 0$  holds.
- (iii) The inequality  $[Q'_p(p - c) + Q'_p + Q_p][c''QL_I + \tau''] + [c']^2L_IQ_pQ' < 0$  holds for any possible value of  $L_I$  in  $[1/n, 1)$ .

Parts (i) and (ii) of Assumption 3 guarantee the interior levels of equilibrium outputs and training levels, respectively. Part (iii) guarantees the uniqueness of the symmetric equilibrium. The following is a leading example of our model, which will be used in later sections.

**Example 1** (Shubik-Levitan and Chen-Riordan specification). Consider the case where the demand function  $D_i(\cdot)$  is given by *Shubik and Levitan (2013)* specification and labor supply function  $L_{ji}(\cdot)$  is given by *Chen and Riordan (2007)* specification:

$$D_i(\mathbf{p}) = \frac{1}{n} \left[ v - p_i + \frac{\gamma}{1 - \gamma} \sum_{j=1}^n \left( \frac{p_j - p_i}{n} \right) \right], \quad L_{ji}(v^j) = \frac{1}{n} + \frac{\sum_{k \neq i} (v_{ji} - v_{jk})}{n(n-1)\theta},$$

where  $\gamma \in [0, 1)$  and  $\theta > 0$ . Furthermore, suppose that marginal cost function  $c(\cdot)$  and training cost function  $\tau(\cdot)$  are given as follows:

$$c(H) = \bar{C} - \frac{\eta}{\alpha_c} H^{\alpha_c}, \quad \tau(h) = \frac{\kappa}{\alpha_\tau} h^{\alpha_\tau},$$

where  $\eta, \kappa, \alpha_c > 0$ , and  $\alpha_\tau > \alpha_c$ . Then, Assumptions 1 and 2 are always satisfied. Assumption 3-(i) is satisfied when  $v > \bar{C}$ , Assumption 3-(ii) is satisfied when  $\alpha_\tau \geq 2\alpha_c$ , and Assumption 3-(iii) is satisfied when  $\kappa$  is sufficiently large relative to  $\eta$ .

### 3.1 Discussion

**Switching cost and idiosyncratic preferences** Our worker specification features switching costs and idiosyncratic preferences. Manning (2003) states, citing Robinson (1933), that it is ignorance, heterogeneous preferences, and mobility costs that are plausible sources of frictions in the labor market. On top of inherent mobility costs, switching cost can also arise from a number of different sources, including contractual restraints on worker mobility such as noncompete agreements.

Our modelling choice on idiosyncratic preferences is consistent with Card et al. (2018), who provide a microeconomic foundation for imperfect market competition in the labor market by allowing workers to have heterogeneous preferences over the work environments of different potential employers. In our model, each worker has different idiosyncratic preferences with his/her current employer and other potential employers. When a worker's idiosyncratic preference with the current employer turns out to be relatively low, the worker may end up moving to another firm, even though the worker must incur the switching cost. Without idiosyncratic preferences, no workers switch employers.

**Timing and observability assumptions** Although we frame the model as a three-stage game, stages 1 and 2 can be viewed as one stage, since the firms cannot observe the other firms' training levels. Unobservability of training levels is often assumed in the literature on human capital investment in labor markets (e.g., Chang and Wang, 1996). An alternative assumption of observable training would substantially complicate the analysis because equilibria of all stage 2 subgames under arbitrary levels of training  $h_i$  need to be identified in order to characterize subgame-perfect equilibria. Furthermore, the observability of training gives rise to strategic commitment effects, as it induces other firms to adjust their wages and prices (e.g., Gersbach and Schmutzler, 2012). To focus on primary (non-strategic) incentives for firms to train workers, we opt for unobservable training, as often done in the R&D literature (e.g., Vives, 2008; López and Vives, 2019).

**Cost specification and its interpretation** In our model, the marginal cost is constant with respect to outputs but is decreasing in the level of aggregate human capital. Through their human capital, workers directly contribute to reducing marginal cost rather than increasing production, although indirectly, lower marginal cost prompts firms to produce more. This mechanism pertains to non-production workers (e.g., engineers) who enhance technology rather than engage in direct production. We can interpret the cost specification that production workers and their wages are embedded in the unit cost function  $c(H_i)$ . For example,  $c(H_i)$  can arise from a technology in which firm  $i$  requires  $c(H_i)$  production workers to produce one unit of output, with the reservation wage of these workers normalized to one. An increase in training or employment of non-production workers—that raises  $H_i$ —reduces this unit labor requirement and, hence, the unit production cost.<sup>7</sup>

## 4 Analysis of the model

### 4.1 Characterization of the equilibrium

In the second stage, each firm  $i$  chooses  $w_{ij}$ ,  $j \in \mathcal{N}$  and  $p_i$  to maximize the profit  $\Pi_i$  given by equation (1). The first-order condition for  $p_i$ , and  $w_{ij}$  for  $i, j \in \mathcal{N}_f$  is given by

$$\frac{\partial \Pi_i}{\partial p_i} = D_i + \frac{\partial D_i}{\partial p_i} [p_i - c(H_i)] = 0, \quad (2)$$

$$\frac{\partial \Pi_i}{\partial w_{ij}} = -L_{ji} - \frac{\partial L_{ji}}{\partial w_{ij}} [c'(H_i) D_i h_j + w_{ij}] = 0. \quad (3)$$

Next, we consider the choice of the training level  $h_i$  by firms in the first stage. Each firm  $i$  chooses  $h_i$  to maximize profit  $\Pi_i$  taking other firms' training levels, wages, and prices as given. Thus, the first-order condition for  $h_i$  is given by

$$\frac{\partial \Pi_i}{\partial h_i} = -c'(H_i) D_i L_{ii} - \tau'(h_i) = 0. \quad (4)$$

At the symmetric equilibrium, (2), (3), and (4) hold, and we have  $p_i = p^*$ ,  $D_i = Q(p^*)$ ,  $L_{ii} = L_I^*$ ,  $L_{ij} = L_O^*$ ,  $h_i = h^*$ ,  $w_{ii} = w_I^*$  and  $w_{ij} = w_O^*$  for all  $i$  and  $j \neq i$ .<sup>8</sup>

Suppose that there exists a symmetric equilibrium in which a strictly positive measure of workers are retained and poached (that is,  $L_I^* > 0$  and  $L_O^* > 0$ ). We call it an *interior*

---

<sup>7</sup>Even if the marginal cost is not constant in the output, our main insights would carry over as long as production cost is submodular in human capital and outputs—that is, the cost reduction achieved by human capital becomes more important as output expands.

<sup>8</sup>Our model does not rule out possibilities of asymmetric equilibria in which firms employ different mass of workers and workers earn different retention and poaching wages.

*symmetric equilibrium.* In the interior symmetric equilibrium, wages  $w_I^*$  and  $w_O^*$  are determined as the solution to the simultaneous equations (5) and (6):

$$w_I^* + \frac{L_{ii}}{\partial L_{ii}/\partial w_{ii}} = -c'(h^*)Q(p^*)h^*, \quad (5)$$

$$w_O^* + \frac{L_{ji}}{\partial L_{ji}/\partial w_{ij}} = -c'(h^*)Q(p^*)h^*, \quad (6)$$

where the second terms of the left-hand side of equations (5) and (6) are evaluated at  $w_{ii} = w_I^*$ , and  $w_{ij} = w_O^*$  for  $j \neq i$ .

We will now complete the characterization of the equilibrium labor market allocation  $(L_I^*, L_O^*)$ . First, the following lemma establishes that the wage difference  $\Delta w := w_O^* - w_I^*$  and labor market allocation  $(L_I^*, L_O^*)$  are determined independently of  $(p, h)$ .

**Lemma 1.** *Suppose that a symmetric equilibrium exists. Then, under Assumption 1, there exists a critical value  $\bar{\sigma} > 0$  such that the symmetric equilibrium is interior (i.e.,  $L_O^* > 0$ ) if and only if  $\sigma < \bar{\sigma}$ . When  $\sigma < \bar{\sigma}$ , there exists a unique value  $\Delta w^*$  that is independent of  $(p, h)$  and satisfies  $w_O^* - w_I^* = \Delta w^*$ , where*

$$\Delta w^* = X_I(\sigma - \Delta w^*) - X_O(\sigma - \Delta w^*). \quad (7)$$

*Furthermore, equilibrium retention and poaching probability,  $(L_I^*, L_O^*)$ , are also determined independently of  $(p, h)$  and satisfy  $L_I^* \in [1/n, 1)$  and  $L_I^* + (n - 1)L_O^* = 1$ .*

Lemma 1 tells us that the equilibrium wage difference and labor market allocation are independent of other equilibrium variables. This is because training is industry specific and workers have no outside option.<sup>9</sup> Because workers have no outside option, only wage differences, not wage levels, determine the labor market allocation.<sup>10</sup>

Furthermore, because training is industry specific, changes in symmetric training levels uniformly affect wages, leaving wage differences unchanged. Similarly, changes in symmetric prices also uniformly affect wages. Hence, prices  $p$  and training levels  $h$  have no effect on the equilibrium wage difference  $\Delta w^*$ . Consequently, the equilibrium labor market allocation is determined solely by the labor market conditions, independent of  $(p, h)$ .

Next, we identify the equilibrium condition for  $p^*$  and  $h^*$ , given the equilibrium labor market allocation  $(L_I^*, L_O^*)$ . The necessary condition for the symmetric equilibrium price

<sup>9</sup>In Section 6, we consider an extension of our model that incorporates firm-specific human capital that lowers indirect costs alongside industry-specific human capital that reduces direct costs.

<sup>10</sup>When workers have outside options and choose to leave the industry in the equilibrium, the equilibrium labor market allocation depends on wage levels as well, making  $L_I$  and  $L_O$  dependent on  $(p, h)$ .

$p^*$  and the training level  $h^*$  are given by the simultaneous equations (8) and (9):

$$Q_p(p)[p - c(h)] + Q(p) = 0, \quad (8)$$

$$-c'(h)Q(p)L_I^* - \tau'(h) = 0. \quad (9)$$

Solving the system of equations, we find  $p^*$  and  $h^*$ , where Assumption 3 ensures the existence of a unique solution  $(p^*, h^*)$ .

We are now ready to characterize the symmetric equilibrium.

**Proposition 1.** *Suppose that a symmetric equilibrium exists, where  $\sigma < \bar{\sigma}$  holds. Then, under Assumptions 1 and 3, the symmetric equilibrium is interior and unique.*

Proposition 1 presumes that a symmetric equilibrium exists. Typically, sufficient conditions for such an equilibrium existence rely on the quasi-concavity of profit functions (e.g., [Glicksberg, 1952](#)). In our setting, it is hard to establish a general condition under which each firm's profit function is quasi-concave because of the high dimensionality of the strategy space. We establish the existence of the symmetric equilibrium under several functional specifications (see Online Appendix A.2.2). In the rest of the paper, we proceed with our analysis under the assumption that the model has a unique symmetric equilibrium.

## 4.2 Underinvestment vs. overinvestment in training

As mentioned in the Introduction, in our model, firms can either underinvest or overinvest in training in the sense that the equilibrium training level can be above or below the level of training that maximizes industry profit plus wage bill, which is equivalent to product market profit minus training cost. R&D literature uses a similar measure—product market profits minus the cost of investment in R&D—to determine under- or overinvestment in R&D. Distinction between under and overinvestments plays a critical role in our model because, as we show in the next section, underinvestment in training is a necessary condition for an increase in firms' market power to result in a clear-cut welfare improvement. Furthermore, the likelihood of under- and overinvestments is connected to the relative strength of product and labor market competition, thereby illustrating the role of the link between the two markets in obtaining useful policy implications.

Consider a variant of the model in which the level of training is fixed at  $h$  rather than determined endogenously. For a given  $h$ , let  $\tilde{p}(h)$  denote the equilibrium price and  $\bar{\Pi}(h)$  denote each firm's equilibrium profit plus wage bill. We then have that

$$\bar{\Pi}(h) = \underbrace{[\tilde{p}(h) - c(h)]Q(\tilde{p}(h))}_{\text{product market profit}} - \underbrace{\tau(h)}_{\text{training costs}}.$$

We say that  $h$  is an overinvestment when  $\bar{\Pi}'(h) < 0$  — a marginal increase in the training level from  $h$  reduces each firm's profit plus wage bill as well as the industry profit plus wage bill in equilibrium. Similarly,  $h$  reflects underinvestment when  $\bar{\Pi}'(h) > 0$ .

The derivative  $\bar{\Pi}'(h)$  captures how an individual firm's profit plus wage bill changes when all firms marginally increase their training level from a given  $h$ . Although each firm chooses its own training level optimally in equilibrium, it does not internalize the impacts on other firms' gross profits. These impacts, as we explain later in this section, can be positive or negative. Since all firms are symmetric, industry profit plus wage bill is  $n$  times individual firm's profits plus wage bill, and a marginal increase in training level increases or decreases it if and only if there is under- or overinvestment, respectively.

We have that

$$\bar{\Pi}'(h^*) = \tilde{p}'(h^*)Q(p^*)\psi(p^*) - c'(h^*)Q(p^*)[1 - L_I^*], \quad (10)$$

where

$$\psi(p) := \frac{\sum_{j \neq i} \frac{\partial D_i}{\partial p_j}}{-\frac{\partial D_i}{\partial p_i}} = \frac{Q_p(p) - Q'(p)}{Q_p(p)} \in [0, 1]$$

denotes the aggregate diversion ratio at  $p$ .<sup>11</sup> The denominator  $-\partial D_i/\partial p_i$  represents a marginal increase of firm  $i$ 's sales when it unilaterally reduces its price  $p_i$ , and the numerator  $\sum_{j \neq i} (\partial D_i/\partial p_j)$  represents the sales firm  $i$  loses when other firms also decrease their prices  $p_j$ . Hence, the ratio  $\psi(p)$  represents the strength of rivalry between firms in the industry. Because  $\tilde{p}'(h^*) < 0$ , the equilibrium training level is more likely to be overinvestment when the rivalry represented by  $\psi(p^*)$  is stronger. This observation leads to the following proposition. Note that for expositional simplicity we assume in Proposition 2-(iii) that the pass-through rate times the aggregate diversion ratio,  $\rho(h)\psi(\tilde{p}(h))$  (where  $\rho(h) := \tilde{p}'(h)/c'(h)$  is the cost pass-through), is increasing in  $h$ . We show in the Online Appendix A.2.3 that this condition holds under various functional specifications.

**Proposition 2.** *The equilibrium training level can be overinvestment or underinvestment depending on the strength of firms' rivalry in the product market.*

- (i) *Suppose that there is no product market competition in the sense that  $\psi(p) = 0$  for all  $p$ . Then the equilibrium training level is always underinvestment.*
- (ii) *Suppose that  $\psi(p) = 1$  for all  $p$  so that the industry-level demand is fixed. Then the equilibrium training level is always overinvestment.*

---

<sup>11</sup>Generally, the local derivative of  $\bar{\Pi}(h)$  at  $h = h^*$  does not imply under- or overinvestment in the strict sense that  $h^*$  exceeds  $\arg \max_h \bar{\Pi}(h)$  or not. In Online Appendix A.2.3, we provide a sufficient condition under which such a local derivative is sufficient for determining whether  $h^*$  exceeds  $\arg \max_h \bar{\Pi}(h)$  or not.

(iii) Suppose  $0 < \psi(p) < 1$  holds for some  $p$  and  $\rho(h)\psi(\tilde{p}(h))$  is increasing in  $h$ . Then, there exists a critical value  $\sigma^\dagger \geq 0$  such that the equilibrium training level is overinvestment (resp. underinvestment) if and only if  $\sigma > \sigma^\dagger$  (resp.  $\sigma < \sigma^\dagger$ ).

The logic here is as follows. When all firms marginally increase their training level at  $h = h^*$ , it mitigates underinvestment in training in terms of each firm's insufficient direct cost-reduction incentive due to labor turnover. At the same time, such an increase in training indirectly affects firms' profits because the corresponding reduction in each firm's constant marginal cost reduces product prices. The indirect effect works in the direction of decreasing industry profit by worsening overinvestment in training. This is because, when each firm unilaterally reduces its price, it does not take into account the negative externality that the lower price imposes on other firms' profits. The equilibrium training level becomes overinvestment when the indirect effect dominates the direct effect.

The condition for under- or overinvestment can be presented in empirically measurable terms. Using equation (10), the condition  $\bar{\Pi}'(h^*) \leq 0$  can be expressed as

$$\rho(h^*)\psi(p^*) \geq 1 - L_I^*. \quad (11)$$

Hence, firms in equilibrium overinvest in training if and only if the aggregate diversion ratio multiplied by cost pass-through exceeds the labor turnover rate. Because all three objects are estimable, equation (11) provides a simple condition in terms of observable industry characteristics under which firms are likely to underinvest or overinvest in training.<sup>12</sup>

In order to obtain a sharper characterization of conditions for under- or overinvestment, below we consider a specific demand function presented in Example 1.

**Proposition 3.** *Suppose that the demand function is given by the Shubik-Levitan specification presented in Example 1. Then  $\rho(h^*)\psi(p^*) = \frac{\gamma \frac{n-1}{1-\gamma} \frac{n}{n-1}}{2 + \frac{\gamma}{1-\gamma} \frac{n-1}{n}}$ , which is increasing in the product substitutability  $\gamma$ . For any given  $\sigma$ , there exists a value  $\hat{\gamma}(\sigma)$  such that the equilibrium training level is overinvestment if and only if  $\gamma > \hat{\gamma}(\sigma)$ . Furthermore, the function  $\hat{\gamma}(\sigma)$  is decreasing in  $\sigma$ , implying that the range of the values of  $\gamma$  that leads to overinvestment expands with the switching cost  $\sigma$ .*

Proposition 3 can be understood as follows. Stronger rivalry in product market competition, represented by higher  $\gamma$ , works in the direction of inducing firms to overinvest in training, while weaker labor market power, represented by lower  $\sigma$ , works in the direction of inducing firms to underinvest in training by increasing labor turnover. It follows that,

<sup>12</sup>See Fabra and Reguant (2014) and Miller, Osborne and Sheu (2017) for examples of estimation of cost pass-through, and Conlon and Mortimer (2021) for empirical properties of diversion ratio.

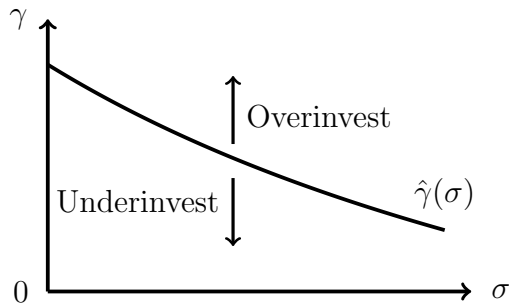


Figure 1: Over/underinvestment

holding  $\sigma$  fixed, higher  $\gamma$  makes overinvestment more likely, and, holding  $\gamma$  fixed, lower  $\sigma$  makes underinvestment more likely, as illustrated by Figure 1.

We end this section by making two final points. First, investment in industry-specific training with labor turnover featured in our model is comparable to investment in cost-reducing R&D with technology spillovers in models of R&D competition (e.g., [d'Aspremont and Jacquemin, 1988](#); [Suzumura, 1992](#); [Leahy and Neary, 1997](#)), in the sense that cost-reducing “technology” embedded in a worker through training “spills over” from one firm to the other firm through labor turnover in our model.<sup>13</sup> Firms tend to underinvest (reps. overinvest) in R&D when the rate of technology spillovers are high (resp. low) in standard models of R&D competition, when product market profit minus R&D cost is used as the metric ([Leahy and Neary, 1997](#)). Similarly, in our model firms tend to underinvest (reps. overinvest) in training when the labor turnover rate is relatively high (resp. low), when product market profit minus training cost is used as the metric.

A unique feature of labor turnover as a key determinant of under- or overinvestment is that labor turnover establishes the link between labor and product market competition, and hence under- or overinvestment is determined by the interaction between firms’ labor and product market power in our model, as illustrated by Figure 1. In contrast, models of R&D competition do not capture the roles played by labor market power.

Our second point concerns what would happen in our model when total surplus, an alternative metric often used in the R&D literature, is adopted as the metric for under- vs. overinvestment in training. In terms of total surplus, firms always underinvest in training as they ignore two positive externalities. First, each firm ignores  $-c'(h^*)Q(p^*)(1 - L_I^*)$ , which is the benefit of training-induced cost reduction that increases rival firms’ profits. Second, a small increase in training from  $h = h^*$  lowers prices and increases consumer surplus by  $-n\tilde{p}'(h^*)Q(p^*)$ . Part of that increment,  $-n\tilde{p}'(h^*)\psi(p^*)Q(p^*)$ , is offset by an

<sup>13</sup>Notice that they are comparable but different in the sense that cost-reducing technology embedded in a worker contributes to the cost-reduction of one firm, either the firm that has retained the worker *or* a firm that has poached the worker, whereas technology spillovers typically contribute to the cost-reduction of a firm that has invested in the technology *and* other firms that have not invested in it.

equivalent reduction in profits. The remaining part,  $-n\tilde{p}'(h^*)[1 - \psi(p^*)]Q(p^*)$ , adds to total surplus, but is ignored by firms when they decide training levels.

## 5 Effects of market power

In this section, we explore the welfare implications of market power. In each section, we first discuss the effects of firms' market power (labor, product, or overall market power) holding the level of training fixed, in which case our model predicts that a change in market power cannot lead to a clear-cut welfare improvement. We then show that the endogenous determination of the training level enriches the welfare implications, indicating that a clear-cut welfare improvement is possible. A necessary condition for such an improvement is that a change in market power induces firms to train more, and more intense training mitigates underinvestment.

### 5.1 Effects of labor market power

**Direct effect** To clarify the role played by the endogenous determination of the training level, we first study the direct effect of  $\sigma$ , holding the level of training fixed. To this end, we consider a variant of the model in which the level of training is fixed at  $h$ . Suppose that the model has an interior symmetric equilibrium. Let  $\tilde{w}_I(\sigma, h)$ ,  $\tilde{w}_O(\sigma, h)$  and  $\tilde{L}_I(\sigma, h)$  respectively denote the equilibrium retention wage, the poaching wage, and the retention rate, and let  $\widetilde{CS}(\sigma, h)$ ,  $\widetilde{WS}(\sigma, h)$ ,  $\widetilde{IP}(\sigma, h)$ , and  $\widetilde{TS}(\sigma, h)$  respectively denote consumer surplus, worker surplus, industry profit, and total surplus in equilibrium, where  $\widetilde{TS}(\sigma, h) = \widetilde{CS}(\sigma, h) + \widetilde{WS}(\sigma, h) + \widetilde{IP}(\sigma, h)$ . The following lemma summarizes the direct effects.

**Lemma 2.** *Consider a variant of the model in which the level of training is fixed at  $h$ , and suppose that the model has an interior symmetric equilibrium. Then, as  $\sigma$  increases,*

- (i)  $\tilde{w}_I(\sigma, h)$  decreases,  $\tilde{w}_O(\sigma, h)$  increases, and  $\tilde{w}_O(\sigma, h) - \sigma$  decreases. Furthermore,  $\tilde{w}_O(\sigma, h) - \sigma$  decreases faster than the decrease in  $\tilde{w}_I(\sigma, h)$  as depicted in Figure 2.
- (ii)  $\tilde{L}_I(\sigma, h)$  increases,
- (iii)  $\widetilde{CS}(\sigma, h)$  remains unchanged,  $\widetilde{WS}(\sigma, h)$  decreases, and  $\widetilde{IP}(\sigma, h)$  may increase or decrease.

Furthermore, (a)  $\widetilde{TS}(\sigma, h)$  attains its maximum at  $\sigma = 0$ , and (b) there exists a value  $\tilde{\sigma}(> 0)$  such that  $\widetilde{TS}(\sigma, h)$  is strictly decreasing in  $\sigma$  for all  $\sigma < \tilde{\sigma}$ , where  $\tilde{\sigma} = \infty$  under certain functional specifications.

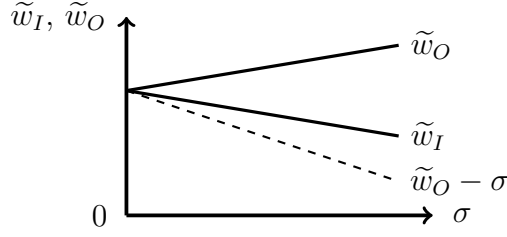


Figure 2: Behaviors of wages ( $\tilde{w}_I$  and  $\tilde{w}_O$ ) with respect to switching costs ( $\sigma$ )

Lemma 2 can be understood as follows. As  $\sigma$  increases, each firm takes advantage of the higher switching cost by retaining more workers at a lower wage, thus  $\tilde{w}_I(\sigma, h)$  decreases. Each firm mitigates, to some extent, the negative impacts that the higher switching cost imposes on poached workers by raising its poaching wage, thus  $\tilde{w}_O(\sigma, h)$  increases. However, the increase in  $\tilde{w}_O(\sigma, h)$  does not fully offset the increase in  $\sigma$  so that switching employers becomes a less attractive option than staying with the initial employer. This means that  $\tilde{w}_O(\sigma, h) - \sigma$  decreases faster than the decrease in  $\tilde{w}_I(\sigma, h)$ , which implies that the retention rate  $\tilde{L}_I(\sigma, h)$  increases. This explains (i) and (ii).

Let us turn to (iii). Since the level of training is fixed, each firm's equilibrium production cost remains unchanged at  $c(h)$ , implying that each firm's output, the industry output, and consumer surplus  $\widetilde{CS}(\sigma, h)$ , all remain unchanged. Regarding the effect on workers, each worker chooses whether to remain with his/her initial employer by receiving  $\tilde{w}_I(\sigma, h)$  or to move to another firm by receiving  $\tilde{w}_O(\sigma, h)$  and incurring the switching cost  $\sigma$ . Since both  $\tilde{w}_I(\sigma, h)$  and  $\tilde{w}_O(\sigma, h) - \sigma$  decrease as  $\sigma$  increases, worker surplus  $\widetilde{WS}(\sigma, h)$  decreases. As to the effect on industry profit, industry profitability in our model is represented by industry profit plus wage bill, which remains unchanged as  $c(h)$  remains unchanged. Regarding the wage bill, an increase in  $\sigma$  decreases the retention wage but increases the poaching wage, implying that the total wage bill may increase or decrease. This implies that industry profit  $\widetilde{IP}(\sigma, h)$  may increase or decrease depending on whether the total wage decreases or increases.

Lastly, consider the effect on total surplus, which can be disaggregated into consumer surplus, industry profit plus wage bill, and workers' nonmonetary utility (NMU). Since both consumer surplus and industry profit plus wage bill remain unchanged when  $h$  is fixed,  $\sigma$  affects total surplus only through NMU, which consists of idiosyncratic preferences and switching costs. As  $\sigma$  increases, poached workers incur higher switching costs, which works in the direction of decreasing NMU. However, an increase in  $\sigma$  mitigates inefficiency associated with excessive turnover by reducing it, and this effect increases NMU.<sup>14</sup> When

<sup>14</sup>With respect to staying with or leaving current employers, workers make efficient decisions in the sense of maximizing NMU when retention and poaching wages are equal. However, for all  $\sigma > 0$ , the poaching wage is strictly higher than the retention wage in equilibrium, resulting in turnover that exceeds the efficient level.

$\sigma$  is below a threshold, denoted  $\tilde{\sigma}$ , the former effect dominates the latter, as stated in Lemma 2-(b). Despite this nonmonotonicity, total surplus is maximized at  $\sigma = 0$  because, when  $\sigma = 0$ , there is no inefficiency arising from wage differentials and workers incur no switching costs.

**Total effect** Let us now consider the total effect, taking into account the effect of  $\sigma$  on the training level. As relevant welfare measures, let  $CS^*$ ,  $WS^*$ , and  $IP^*$  denote the equilibrium consumer surplus, worker surplus, and industry profit, respectively, and  $TS^* = CS^* + WS^* + IP^*$  denote the equilibrium total surplus. As  $\sigma$  increases, each firm retains a larger fraction of its workers at a lower wage. Anticipating the higher retention rate  $L_I^*$ , each firm trains its initial employees more intensively. An increase in training level  $h^*$  reduces each firm's constant marginal cost  $c(h^*)$ , which increases each firm's equilibrium output. Hence, industry output  $Q^*$  increases and price  $p^*$  decreases, implying that consumer surplus  $CS^*$  increases. These results are presented in Lemma 3-(i) and (ii).

**Lemma 3.** *As  $\sigma$  increases,*

- (i)  $L_I^*$  and  $h^*$  increase,
- (ii)  $Q^*$  and  $CS^*$  increase, and
- (iii)  $WS^*$  and  $IP^*$  may increase or decrease depending on parameterizations.

*Furthermore, while  $TS^*$  can increase or decrease with  $\sigma$ , unlike the case of fixed  $h$ ,  $TS^*(\sigma)$  is not necessarily maximized at  $\sigma = 0$ .*

Regarding worker surplus, when the level of training is endogenously determined, an increase in  $\sigma$  can increase worker surplus due to the indirect effect through two channels. First, as the higher  $h^*$  induced by an increase in  $\sigma$  increases  $Q^*$ , each firm can apply per unit cost reduction to a larger amount of its output. Second, the marginal cost reduction by retaining or hiring a worker from another firm,  $-c'(h^*)h^*$ , is increasing in  $h^*$  under Assumption 2. Both of the two indirect effects work in the same direction of increasing the value of a worker for firms, where retention and poaching wages increase by the same amount since training makes a worker equally valuable to all employers. This implies that labor mobility remains unchanged as  $h^*$  increases, and hence workers' nonmonetary utility (NMU) also remains unchanged. Since NMU remains the same while wages increase for all workers, it follows that the indirect effect increases worker surplus. Then, given that the direct effect decreases worker surplus, the total effect of an increase in  $\sigma$  can be positive or negative depending on the relative magnitude of the two effects.

Regarding industry profit, higher  $h^*$  works in the direction of increasing wages and hence decreasing industry profit. At the same time, the lower production cost induced

by higher  $h^*$  works in the direction of increasing (resp. decreasing) industry profit if it mitigates underinvestment (resp. worsens overinvestment). Along with the direct effect, the total effect on industry profit can also be positive or negative.

Finally, turning to total surplus, we find that, compared to the case of  $\sigma = 0$ , a positive  $\sigma$  can increase total surplus because the higher training induced by  $\sigma$  increases total surplus, which may offset the decrease in NMU under fixed  $h$  as shown in Lemma 2.

|                     | Output | Price | Retention rate | Retention wage | Poaching wage | Consumer surplus | Worker surplus | Industry profit |
|---------------------|--------|-------|----------------|----------------|---------------|------------------|----------------|-----------------|
| Exogenous training  | —      | —     | ↑              | ↓              | ↑             | —                | ↓              | ↑↓              |
| Endogenous training | ↑      | ↓     | ↑              | ↑↓             | ↑             | ↑                | ↑↓             | ↑↓              |

Table 2: Effect of an increase in  $\sigma$  on equilibrium outcomes

Table 2 summarizes the effect of an increase in  $\sigma$  on equilibrium outcomes under exogenous and endogenous training. These entries in the cells of the table tell us the possible direction of change in equilibrium outcomes due to an increase in  $\sigma$  based on Lemmas 2 and 3.

Lemma 3 suggests that there is a possibility for an increase in  $\sigma$  to result in a clear-cut welfare improvement. Consider a range of parameterizations in which industry profit increases as  $\sigma$  increases. Firms can then increase their profits by jointly increasing  $\sigma$ —for example, through collusive agreements or by lobbying for relaxing regulations. This can be interpreted as a reduced form of coordination between firms in jointly raising the strength and effectiveness of noncompete agreements with their employees (see Section 6 for a discussion of policy implications on noncompete agreements). Higher  $\sigma$  benefits not only firms but also consumers. Can this coordination then be regarded as pro-competitive action? The answer depends on the effect of  $\sigma$  on workers, which in turn crucially depends on whether the equilibrium level of training is overinvestment or underinvestment, as stated in Proposition 4. Note that, for expositional simplicity, Proposition 4 is stated for the case of  $\sigma < \tilde{\sigma}$  where  $\tilde{\sigma}$  is as defined in Lemma 2.

**Proposition 4.** *Suppose that a local increase in  $\sigma (< \tilde{\sigma})$  increases industry profit.*

- (i) *If the equilibrium training level is overinvestment, the increase in  $\sigma$  can never result in a clear-cut welfare improvement because it decreases equilibrium worker surplus.*
- (ii) *If the equilibrium training level is underinvestment, the increase in  $\sigma$  may result in a clear-cut welfare improvement because it may increase equilibrium worker surplus.*

To understand Proposition 4, first suppose that an increase in  $\sigma$  worsens overinvestment in training and hence decreases industry profit plus wage bill. Then, since an increase

in  $\sigma$  increases industry profit, it decreases wages. This implies that firms' coordinated action of increasing  $\sigma$  decreases worker surplus whenever it lowers workers' non-monetary utility (i.e.  $\sigma < \tilde{\sigma}$ ). Next suppose that an increase in  $\sigma$  mitigates underinvestment in training. Then, even though an increase in  $\sigma$  increases industry profit, it can also increase wages and worker surplus. We find that an increase in  $\sigma$  can in fact increase worker surplus. In such a case, coordination in  $\sigma$  can be viewed as pro-competitive in the sense that it benefits all three relevant parties: firms, consumers, and workers.

## 5.2 Effects of product market power

There are a number of ways to model product market competition, where the intensity of competition is often captured by product substitutability (e.g., [Vives, 1999](#)), or the value of outside options that may capture the competitive pressures from outside markets (e.g., [Nocke and Schutz, 2018](#)). We take a general approach in which these specific modelling choices are nested. Following [Vives \(2008\)](#), we parameterize the intensity of product market competition by  $\zeta$  so that we can write the demand function as  $D_i(\mathbf{p}) = \tilde{D}_i(\mathbf{p}; \zeta)$ . Let  $\tilde{Q}(p; \zeta) := \tilde{D}_i(p, \dots, p; \zeta)$ , and  $\tilde{Q}_p(p; \zeta) := \partial \tilde{D}_i(p, \dots, p; \zeta) / \partial p_i$ . We assume that  $\zeta$  represents the intensity of product market competition in the following sense:

**Assumption 4.** *The following conditions hold:*

(i) *Demand scale  $\tilde{Q}(p; \zeta)$  weakly decreases with  $\zeta$ .*

(ii) *Price elasticity  $\varepsilon_Q(p; \zeta) := -\tilde{Q}_p(p; \zeta)p/\tilde{Q}(p; \zeta)$  strictly increases with  $\zeta$ .*<sup>15</sup>

A unique feature of our model is that it enables us to study the ways in which workers are affected by product market power. The effect of product market power on workers, in turn, crucially depends on its effect on the industry output.

In a variant of the model in which the level of training is fixed at  $h$ , Assumption 4 implies that the equilibrium price  $\tilde{p}(h; \zeta)$  and each firm's equilibrium profit plus wage bill  $\bar{\Pi}(h; \zeta)$  are both decreasing in  $\zeta$ . As  $\zeta$  increases, the lower equilibrium price works in the direction of increasing industry output, whereas the direct effect on demand scale (Assumption 4 (i)) works in the opposite direction. Hence the equilibrium industry output increases if and only if the former effect dominates the latter effect. Higher output makes workers more valuable to firms, and hence an increase in  $\zeta$  increases wages and worker surplus if and only if it increases the industry output. Higher industry output increases

---

<sup>15</sup>Assumption 4(i) holds for a wide range of specific demand functions analyzed in [Vives \(2008\)](#) and [Nocke and Schutz \(2018\)](#) including Shubik-Levitan, Hotelling, and Logit demand function where  $\zeta$  captures respectively the substitutability parameter, transport cost, and the value of outside option. Assumption 4(ii) is made by [Vives \(2008\)](#).

worker surplus but decreases firms' profits, which implies that an increase in  $\zeta$  cannot result in a clear-cut welfare improvement.

When the training level is endogenously determined, output and the training have mutually reinforcing effects.<sup>16</sup> On the one hand, higher output induces each firm to train its workers more intensively because it can apply the unit cost reduction to a larger amount of its output. On the other hand, a higher level of training reduces each firm's production cost and hence it increases each firm's output. Thus, a change in  $\zeta$  increases  $Q^*$  if and only if it increases  $h^*$ . Higher output and more intense training make workers more valuable, leading to higher wages and worker surplus. Thus, we have Proposition 5.

**Proposition 5.** *An increase in  $\zeta$  increases the equilibrium wages and the equilibrium training level if and only if it increases the equilibrium industry output. That is, the following relationships hold:*

$$\text{sign} \left\{ \frac{dQ^*}{d\zeta} \right\} = \text{sign} \left\{ \frac{dw_I^*}{d\zeta} \right\} = \text{sign} \left\{ \frac{dw_O^*}{d\zeta} \right\} = \text{sign} \left\{ \frac{dh^*}{d\zeta} \right\}.$$

*Furthermore, an increase in  $\zeta$  increases the equilibrium worker surplus  $WS^*$  if and only if it increases the equilibrium industry output.*

This result is consistent with the recent views of several antitrust scholars, who argue that output serves as a measure of worker welfare. For example, [Hovenkamp \(2019\)](#) argues that "... just as consumers benefit from high output because it produces lower prices in product markets, so too labor benefits from high output because it increases the demand for jobs and, in the process, boosts wages." Proposition 5 supports this argument insofar as changes in outputs are driven by product market competition.

An increase in  $\zeta$  may, in fact, increase or decrease the equilibrium worker surplus, as demonstrated by the following examples.

**Example 2.** *(i) In the Shubik-Levitan example of Table 1, the equilibrium worker surplus increases with product substitutability  $\gamma$ . (ii) In the Logit example of Table 1, the equilibrium worker surplus decreases with the value of outside option  $V_0^C$ .*

The effect of  $\zeta$  on consumers is complex because a change in  $\zeta$  can directly affect consumer surplus even when prices are given. If an increase in  $\zeta$  has a positive direct effect on consumers and increases the equilibrium industry output, it unambiguously increases consumer surplus. In contrast, if an increase in  $\zeta$  has a negative direct effect, it increases consumer surplus if and only if the negative direct effect is dominated by the

---

<sup>16</sup>The firm-sponsored training in our model is analogous to cost-reducing R&D in [Vives \(2008\)](#)'s model, and hence the effect of product market competition on training in our model is similar to the effect of competition on R&D in [Vives \(2008\)](#)'s model. In our model, a link between product and labor market competition leads to mutually reinforcing effects of a change in  $\zeta$  on product and labor market outcomes.

positive indirect effect. We find that an increase in  $\zeta$  can in fact increase consumer surplus under some parameterizations and decrease it in others.

An increase in  $\zeta$  benefits workers if it increases the industry output, and it can also benefit consumers. So, it can result in a clear-cut welfare improvement if the higher industry output benefits firms. When the level of training is fixed, an increase in  $\zeta$  decreases the profit of the firms if it increases the industry output, which implies that a clear-cut welfare improvement is not possible. This, however, is not necessarily the case in our full model because the higher industry output induces each firm to train its workers more intensively. If the equilibrium level of training is overinvestment, a higher training level further reduces each firm's profit and the industry profit. In contrast, if the equilibrium level of training is underinvestment, a higher training level works in the direction of increasing each firm's profit by mitigating underinvestment and so the industry profit may increase or decrease as  $\zeta$  increases. This leads us to Proposition 6.

**Proposition 6.** *Suppose that the demand function is given by the Shubik-Levitan specification presented in Example 1. As product substitutability  $\gamma$  increases, equilibrium training level  $h^*$ , aggregate output ( $Q^*$ ), consumer surplus ( $CS^*$ ), and worker surplus ( $WS^*$ ) all increase. Industry profit ( $IP^*$ ) decreases when  $h^*$  represents overinvestment, but increases under some parameterizations when  $h^*$  reflects underinvestment.*

Proposition 6 tells us that an increase in  $\zeta$  ( $\gamma$  in the Shubik-Levitan example) can benefit all three parties involved; workers, consumers, and firms, if the higher training level mitigates underinvestment in training. We find that such a clear-cut welfare improvement can indeed be the case under a range of parameterizations.

### 5.3 Effects of overall market power: Overlapping ownership

Thus far, we have studied the effects of labor (resp. product) market power holding product (resp. labor) market power constant. In this subsection, we study the effects of a change in firms' overall market power that changes their labor and product market power simultaneously, and explore how labor and product market power would interact with each other in such a setup to amplify or dampen welfare implications. To this end, we consider the effects of changes in conduct that may arise from partial collusion or common/cross ownership, where we call it "overlapping ownership".

Consider an extension of the model in which the objective function of each firm  $i \in \mathcal{N}$  is given by the weighted sum of firms' profits  $\Omega_i = \pi_i + \lambda \sum_{j \neq i} \pi_j$ , where the weight  $\lambda \in [0, 1)$  is often called the Edgeworth coefficient of effective sympathy among firms that arises from overlapping ownership (e.g. [López and Vives, 2019](#); [Vives and Vravosinos,](#)

forthcoming). For a moment, let  $\sigma = 0$  to focus on the effects of market power associated with  $\lambda$ . The analytical details of the extension are provided in the Online Appendix A.4.

We first study the direct effect of  $\lambda$  holding the training level fixed. As  $\lambda$  increases, each firm  $i$  cares more about other firms' profits and thus has less incentive to lower its product price and raise its retention and poaching wages, because doing so decreases other firms' profits. In other words, higher  $\lambda$  weakens both product and labor market competition, thereby reducing both consumer and worker surplus. In terms of effects on firms, weaker competition increases industry profit. Thus, a change in  $\lambda$  cannot result in a clear-cut welfare improvement.

The endogenous determination of the training level yields richer welfare implications due to its two effects of  $\lambda$  that work in opposite directions on training incentives. On the one hand, the direct effect of higher  $\lambda$  is to decrease industry output by weakening product market competition. The lower output, in turn, reduces the incentive of firms to train workers. On the other hand, as  $\lambda$  increases, each firm cares more about other firms' profits, thereby increasing each firm's training incentive, as it can capture a larger fraction of the returns from training provided to workers who are ultimately poached by other firms.

If the latter effect dominates the former so that higher  $\lambda$  increases  $h^*$ , the production costs decrease, which tends to increase the aggregate output ( $Q^*$ ) and lower the market price ( $p^*$ ). Provided that the extent of overlapping ownership is not too high, the training-induced reduction in  $p^*$  can dominate the upward pressure on price arising from reduced competition, thereby increasing consumer surplus. Worker surplus can also increase, as higher values of  $Q^*$  and  $h^*$  increase the value of workers to firms, incentivizing firms to offer higher retention and poaching wages.

Higher  $\lambda$ , if it increases both consumer and worker surplus, would lead to a clear-cut welfare improvement because it also increases industry profit by weakening competition. Using the functional forms from Example 1, we find that consumers and workers tend to benefit from overlapping ownership when  $\lambda$  and  $\theta$  are small, respectively, as shown in the following proposition.

**Proposition 7.** *Suppose that  $\sigma = 0$ , and consider an increase in  $\lambda$  under the parametric specification of Example 1.*

1. *There exists a critical value  $\hat{\lambda} \in (0, 1)$  such that consumer surplus increases with the increase in  $\lambda$  if and only if  $\lambda < \hat{\lambda}$ .*
2. *For any given  $\lambda$ , there exists a critical value  $\hat{\theta}(\lambda) (> 0)$  such that an increase in  $\lambda$  increases worker surplus if and only if  $\theta < \hat{\theta}(\lambda)$ .*

Although our analysis identifies the potentially procompetitive effects of overlapping ownership, such procompetitive effects disappear when the switching cost  $\sigma$  is so high that the labor turnover rate is low. This is because the beneficial effects of overlapping ownership stem from firms internalizing the positive externalities of labor turnover when making training decisions. When switching costs are high, the labor turnover rate is low, reducing the positive externalities to internalize. This argument is confirmed in the parametric setting of Example 1; In Proposition A.7 of Online Appendix A.4, we show that there exists a critical value  $\sigma_h(> 0)$  such that if  $\sigma > \sigma_h$ , an increase in  $\lambda$  decreases training, thereby decreasing consumer and worker surplus as well.

## 6 Discussion and extension

### 6.1 Discussion

**Policy implications on noncompete agreements** The widespread use of noncompete agreements has prompted nearly 70 state and federal policies in the US, which ban or limit the use of noncompete agreements for all or some of the workforce (Rothstein and Starr, 2022). The use of noncompete agreements is widespread also in Europe (OECD, 2022). Vuorenkoski (2018) finds that 45% of workers belonging to the Finnish trade union Akava are bound by noncompete agreements. Streefkerk, Elshout and Cuelenaere (2015) note that about 19% of employees were covered by noncompete agreements. Young (2021) reports that over 35% of private sector workers in Austria were bound by noncompete agreements in 2005-2006.

In response to the increasing prevalence of contractual restraints on worker mobility including noncompete agreements and no-poaching agreements, the Department of Justice (DOJ) and Federal Trade Commission (FTC) in the United States jointly issued Antitrust Guidance for Human Resource Professionals in 2016.<sup>17</sup> On April 23, 2024, the FTC issued a final rule to ban noncompete agreements nationwide, where the Commission thus adopts a comprehensive ban on new non-competes with all workers.<sup>18</sup>

Against the final rule, two commissioners, Andrew N. Ferguson and Melissa Holyoak, have presented dissenting statements. Their concern stems from the rule's treatment of non-compete agreements as categorical—or per se—violations of Section 5. Based on several empirical evidences that are available to date, they assert that noncompete agreements can be anticompetitive or procompetitive depending on circumstances. Both statements point out that noncompete agreements reduce labor mobility, which in turn

<sup>17</sup>Available at <https://www.justice.gov/atr/file/903511/download> (visited on July 18, 2025).

<sup>18</sup>The FTC's 2024 final rule has been vacated by a federal district court and therefore has not taken effect. The decision is currently under appeal.

induces employers to train their employees more intensively. Commissioner Ferguson then argues that more intense training is critical to the success of employers, employees, and the general economy.<sup>19</sup> Commissioner Holyoak continues to point out that, even when noncompete agreements lower wages, that may benefit consumers by lowering product prices.<sup>20</sup> Both commissioners contend that noncompete agreements should be subject to a case-by-case analysis—the rule of reason—rather than the *per se* rule.

Our model yields implications that are consistent with their concerns. In our model, the switching cost  $\sigma$  can be interpreted to represent noncompete agreements in a reduced form. This is because as the length of a noncompete agreement increases, workers’ switching cost increases since workers who switch their employers have to wait for a longer period of time until they can be employed by competitors of their current employers.

Under this interpretation, our model predicts that stronger enforcement of noncompete agreements reduces labor mobility, which in turn induces firms to train their workers more intensively. More intense training reduces production costs, inducing firms to produce more and benefiting consumers. It can also benefit both firms and workers, thereby leading to a clear-cut welfare improvement, if it mitigates underinvestment in training. This is the case when labor market competition is more intense relative to product market competition, and thus labor mobility is relatively high. In contrast, when product market competition is more intense than labor market competition, it worsens overinvestment in training and thus hurts workers whenever it benefits firms, making such a welfare improvement impossible. Furthermore, we identify an empirically measurable condition under which stronger enforcement of noncompete agreements leads to underinvestment vs. overinvestment in training, which authorities may use to judge whether or not noncompetes are likely to benefit workers.

**Empirical relevance** Our theoretical predictions accord well with recent empirical findings of noncompete agreements (NCAs). Our model predicts that stronger enforcement of NCAs reduces labor mobility, which in turn induces firms to train workers more intensively and produce, increasing consumer surplus. As for wages, stronger NCAs may increase or decrease wages because the direct effect and the indirect effect may work in opposite directions.

Using data from the Survey of Income and Program Participation, [Starr \(2019\)](#) shows

---

<sup>19</sup>Ferguson points out, “Investments in employee training and research and development are neither unfair nor anticompetitive. They are critical to the success of employers, employees, and the general economy. And some noncompete agreements promote both. These findings ought to be fatal to a categorical rule that rests on the proposition that every noncompete agreement is unfair and anticompetitive.” ([Ferguson, 2024](#), p.38).

<sup>20</sup>Holyoak points out, “... even if non-compete clauses do lower wages, that may lower the marginal cost for the employer, and a lower marginal cost will, in competitive situations, be passed through as lower prices to consumers.” ([Holyoak, 2024](#), p.17).

that an increase in enforceability of NCAs is associated with a 14% increase in training. The estimation results of [Kodama, Kambayashi and Izumi \(2025\)](#), which use a firm-level survey in Japan, indicate that firms that use NCAs are more likely to invest in their workers. Empirical findings on the effects of NCAs on wages are rather mixed. Although [Kodama et al. \(2025\)](#) show that NCAs are positively associated with average wages, opposite findings are presented by several studies, including [Starr \(2019\)](#) and [Johnson, Lavetti and Lipsitz \(2023\)](#). [Starr \(2019\)](#) finds that, despite the increases in training, an increase in enforceability of NCAs is associated with a 4% decrease in hourly wages. Similarly, [Johnson et al. \(2023\)](#) find, using newly constructed panel data, that higher NCA enforceability diminishes workers' earning and job mobility.<sup>21</sup> Regarding effects on consumers, [Kodama et al. \(2025\)](#) find that NCAs are positively associated with firm sales, suggesting that NCAs increase consumer surplus.

## 6.2 Extension

**Initial hiring** Our model begins with a situation where each firm is endowed with a unit mass of workers. We have considered an extension of the model that incorporates initial hiring. In what follows, we outline the extended model and its main results, with analytical details and explanations provided in the Online Appendix A.5.

There is a continuum of workers with unit density, where each worker is characterized by  $x \in [0, \bar{X}]$ . Let  $\phi(x)$  denote worker  $x$ 's opportunity cost to work for any of the  $n$  firms in the industry. We assume that  $\phi(x)$  is increasing in  $x$ , and that  $\lim_{x \rightarrow \bar{X}} \phi(x) = \infty$  holds. Each firm  $i$  chooses the mass of workers  $z_i$  it hires, which is assumed to be unobservable to other firms and workers. Let  $\mathbf{z} := (z_i)_{i=1}^n$  represent the vector of the mass of workers initially hired by all  $n$  firms. Workers initially hired by firm  $i$  receive an initial wage  $w_i^0$ , where a vector of initial wages  $\mathbf{w}^0 := (w_i^0)_{i=1}^n$  is determined to clear the market. After initial employment is settled, each firm  $i$  chooses its training level  $h_i$ . Then, without observing  $(z_j, h_j)_{j \neq i}$ , each firm  $i$  chooses  $p_i$  and the vector of its retention wage and poaching wages  $(w_{ji})_{j=1}^n$ . The vector  $\mathbf{w}^1 := ((w_{ji})_{j=1}^n)_{i=1}^n$  represents the vector of retention and poaching wages offered by all  $n$  firms.

In our analysis, we focus on symmetric perfect Bayesian equilibrium with passive belief, and find that the model has such an equilibrium. Regarding effects of labor market power, we consider the following two cases: (i) After the initial hiring is settled and before each firm  $i$  chooses its training level  $h_i$ , there is an unanticipated increase in  $\sigma$  in the sense that such an increase is not expected when each firm  $i$  chooses  $z_i$ , and (ii) an increase in  $\sigma$  is anticipated when each firm  $i$  chooses  $z_i$ .

---

<sup>21</sup>[Starr \(2019\)](#) points out that greater enforceability of NCAs may not lead to wage increases even though training rises, despite the fact that the receipt of training is associated with greater wages.

We find that an unanticipated increase in  $\sigma$  affects the equilibrium outcomes in the same way as in the main analysis. When an increase in  $\sigma$  is anticipated, the qualitative nature of the effects becomes different, depending on whether it increases or decreases employment. When an increase in  $\sigma$  increases employment, it increases not only consumer surplus but also worker surplus, leading to a clear-cut welfare improvement whenever it increases industry profit. Furthermore, a clear-cut welfare improvement is possible even when there is overinvestment in training. By contrast, when an increase in  $\sigma$  decreases employment, it always lowers worker surplus and may even decrease consumer surplus, implying that a clear-cut welfare improvement is impossible.

**Imperfect overlap of product and labor markets** Although our model assumes that the sets of product market competitors and labor market competitors perfectly overlap, this is often not the case in reality—for example, due to geographic segmentation in labor markets. To investigate such a scenario, we analyze the imperfect overlap between product and labor market competition in Online Appendix A.3. Specifically, we consider a setting with  $M \geq 1$  distinct local labor markets, each served by  $N \geq 2$  competing firms. Those  $n = M \times N$  firms compete in a common product market. We interpret an increase in  $M$  (holding  $n$  fixed) as greater labor market segmentation. The labor market segmentation effectively serves as an increase in switching costs—for a worker initially hired by a firm in one labor market, the cost of switching to a firm in another labor market can be viewed as prohibitively large. Using the parametric specification of Example 1, we find that a greater labor market segmentation increases the equilibrium retention rate and leads to a higher training level. As a result, the equilibrium training level is more likely to be overinvestment. In light of Proposition 4, this implies that a clear-cut welfare improvement arising from increased switching costs becomes less likely as the extent of labor market segmentation increases.

**Industry-specific versus firm-specific training** In our model, we focus on investment in industry-specific human capital, which lowers direct production costs. In practice, firms also invest in relation-specific human capital, which reduces not only direct but also indirect costs, including administrative costs (e.g., human resource management, accounting, legal affairs) and selling costs. By fostering stronger relationships between production workers and employees in administrative and sales departments, firms can further reduce non-production costs. Recognizing the firm-specific nature of relation-specific investments, we have extended our model to incorporate firm-specific human capital that reduces indirect costs alongside industry-specific human capital that reduces direct costs. For analytical tractability, we model indirect costs as fixed costs that decrease with the level of firm-specific human capital.

In the extension, in addition to  $h_i$  units of industry-specific human capital, each firm  $i$  provides  $s_i$  ( $\geq 0$ ) units of firm-specific human capital with convex training costs  $\Psi(s_i)$  per worker. Investment in  $s_i$  reduces firm  $i$ 's indirect costs from  $K$  to  $K - \beta S_i$  where  $S_i = s_i L_{ii}$ . At the beginning of the game, each firm  $i$  chooses  $h_i$  and  $s_i$  at the same time, where both  $h_i$  and  $s_i$  are unobservable to other firms. Everything else remains the same as in the base model. We find that in this extension, the qualitative nature of equilibrium characterization results and comparative statics results with respect to  $\sigma$  are the same as in the base model. We also find that a clear-cut welfare improvement driven by an increase in industry-specific training induced by higher labor market power tends to be less likely in an industry or a firm in which the importance of firm-specific human capital, captured by  $\beta$ , is relatively high. See Online Appendix A.6 for details on the analysis of this extension.

## 7 Summary and Conclusion

We have developed a model that captures the link between firms' product and labor market competition, and have explored the model to study the effects of their market power on consumers, workers, and firms themselves. A change in market power affects labor mobility, which affects firms' incentives to train their workers. The resulting changes in firms' production costs, in turn, affect the extent of product market competition. Our model yields novel welfare implications of firms' labor market power, product market power, and overall market power, where the distinction between under- vs. overinvestment in training plays a critical role in determining the nature of welfare implications.

Our model predicts that, when the level of training is fixed, an increase in firms' market power cannot yield a clear-cut welfare improvement, consistent with the conventional view that greater market power adversely affects at least one of consumers or workers. The endogenous determination of the training level, however, enriches the welfare implications of our model, indicating that such a welfare improvement is possible.

An increase in labor market power reduces labor mobility and increases firms' incentives to train workers, thereby intensifying product market competition to benefit consumers. It can also benefit workers and firms when more intense training mitigates underinvestment in training, making a clear-cut welfare improvement possible. In contrast, if more intense training worsens overinvestment, workers are unlikely to be better off as higher switching cost reduces wages whenever they increase industry profit. To assess whether there is under- or overinvestment in training, we derive an empirically testable condition involving the pass-through rate, diversion ratio, and labor turnover.

Turning to product market power, more intense product market competition (i.e.,

lower market power) can benefit not only consumers but also workers and firms. An increase in the intensity of product market competition, when measured by product substitutability in the Shubik-Levitan specification, can result in such a welfare improvement. It induces firms to train their workers more intensively, benefiting not only workers but also consumers. Furthermore, if more intense training mitigates underinvestment, it can also increase industry profit, whereas it necessarily decreases industry profit in the case of worsening overinvestment. The distinction between under- vs. overinvestment in training plays a critical role here as well.

Regarding an increase in overall market power, a higher degree of overlapping ownership works in the direction of providing more intense training because it induces each firm to care more about an increase in its rival firms' profits due to the higher productivity of poached workers. If this effect dominates the negative effect of lower output due to weaker product and labor market competition, an increase in overall market power increases the equilibrium training level, benefiting consumers, workers, and firms, leading to a clear-cut welfare improvement. This can occur when labor turnover is relatively high, but it cannot occur when labor turnover is low.

Exploration of labor market power in our model yields policy implications on non-compete agreements. Noncompete agreements are still permitted in most European jurisdictions. However, in the US, the FTC has recently issued a nationwide ban. In our model, workers' switching costs can be interpreted to represent noncompete agreements in a reduced form. Under this interpretation, our analysis suggests that noncompete agreements can have positive impacts on society by inducing employers to train their employees more intensively, which can benefit both employees and consumers. Such benefits, however, might not materialize when labor mobility is low. Furthermore, worker surplus and consumer surplus do not necessarily move in the same direction. Thus, our analysis suggests that rather than a uniform ban on noncompetes, a case-by-case approach would be more prudent, especially if both consumers' and workers' interests feature in the decision-making of antitrust authorities.

## References

- Almeida, Rita and Pedro Carneiro**, “The Return to Firm Investments in Human Capital,” *Labour Economics*, 2009, 16 (1), 97–106.
- Anderson, Simon P, Andre De Palma, and Jacques-Francois Thisse**, *Discrete Choice Theory of Product Differentiation*, MIT Press, 1992.
- Azar, José and Xavier Vives**, “General Equilibrium Oligopoly and Ownership Structure,” *Econometrica*, 2021, 89 (3), 999–1048.
- Bapna, Ravi, Nishtha Langer, Amit Mehra, Ram Gopal, and Alok Gupta**, “Human Capital Investments and Employee Performance: An Analysis of IT Services Industry,” *Management Science*, 2013, 59 (3), 641–658.
- Bassanini, Andrea and Giorgio Brunello**, “Barriers to Entry, Deregulation and Workplace Training: A Theoretical Model with Evidence from Europe,” *European Economic Review*, 2011, 55 (8), 1152–1176.
- Berger, David, Kyle Herkenhoff, and Simon Mongey**, “Labor Market Power,” *American Economic Review*, 2022, 112 (4), 1147–1193.
- Bhaskar, V and Ted To**, “Minimum Wages for Ronald McDonald Monopsonies: A Theory of Monopsonistic Competition,” *Economic Journal*, 1999, 109 (455), 190–203.
- Card, David**, “Who Set your Wage?,” *American Economic Review*, 2022, 112 (4), 1075–1090.
- , **Ana Rute Cardoso, Joerg Heining, and Patrick Kline**, “Firms and Labor Market Inequality: Evidence and Some Theory,” *Journal of Labor Economics*, 2018, 36 (S1), S13–S70.
- Chang, Chun and Yijiang Wang**, “Human Capital Investment under Asymmetric Information: The Pigovian Conjecture Revisited,” *Journal of Labor Economics*, 1996, 14 (3), 505–519.
- Chen, Yongmin and Michael H Riordan**, “Price and Variety in the Spokes Model,” *Economic Journal*, 2007, 117 (522), 897–921.
- Conlon, Christopher and Julie Holland Mortimer**, “Empirical Properties of Diversion Ratios,” *RAND Journal of Economics*, 2021, 52 (4), 693–726.
- d’Aspremont, Claude and Alexis Jacquemin**, “Cooperative and Noncooperative R & D in Duopoly with Spillovers,” *American Economic Review*, 1988, 78 (5), 1133–1137.
- Fabra, Natalia and Mar Reguant**, “Pass-through of Emissions Costs in Electricity markets,” *American Economic Review*, 2014, 104 (9), 2872–2899.
- Ferguson, Andrew N.**, “Dissenting Statement of Commissioner Andrew N. Ferguson Joined by Commissioner Melissa Holyoak In the Matter of the Non-Compete Clause Rule Matter Number P201200,” 2024.

- Gersbach, Hans and Armin Schmutzler**, “Product Markets and Industry-specific Training,” *RAND Journal of Economics*, 2012, 43 (3), 475–491.
- Glicksberg, Irving L**, “A Further Generalization of the Kakutani Fixed Theorem, with Application to Nash Equilibrium Points,” *Proceedings of the American Mathematical Society*, 1952, 3 (1), 170–174.
- Holyoak, Melissa**, “Dissenting Statement of Commissioner Melissa Holyoak Joined by Commissioner Andrew N. Ferguson In the Matter of the Non-Compete Clause Rule Matter Number P201200,” 2024.
- Hovenkamp, Herbert**, “Competition Policy for Labour Markets,” *University of Pennsylvania, Institute for Law & Economics Research Paper*, 2019, (19-29).
- Johnson, Matthew S, Kurt J Lavetti, and Michael Lipsitz**, “The labor market effects of legal restrictions on worker mobility,” Technical Report, National Bureau of Economic Research 2023.
- Kodama, Naomi, Ryo Kambayashi, and Atsuko Izumi**, “Non-compete Agreements: Human Capital Investments or Compensated Wages?,” *IZA Discussion Paper*, 2025.
- Lamadon, Thibaut, Magne Mogstad, and Bradley Setzler**, “Imperfect Competition, Compensating Differentials, and Rent Sharing in the US Labor Market,” *American Economic Review*, 2022, 112 (1), 169–212.
- Leahy, Dermot and J Peter Neary**, “Public Policy Towards R&D in Oligopolistic Industries,” *American Economic Review*, 1997, 642–662.
- López, Ángel L and Xavier Vives**, “Overlapping Ownership, R&D Spillovers, and Antitrust Policy,” *Journal of Political Economy*, 2019, 127 (5), 2394–2437.
- Manning, Alan**, *Monopsony in Motion: Imperfect Competition in Labor Markets*, Princeton University Press, 2003.
- Martins, Pedro S**, “Employee Training and Firm Performance: Evidence from ESF Grant Applications,” *Labour Economics*, 2021, 72 (C), 102056.
- Mehra, Amit, Nishtha Langer, Ravi Bapna, and Ram Gopal**, “Estimating Returns to Training in The Knowledge Economy: A Firm-Level Analysis of Small and Medium Enterprises,” *MIS Quarterly*, 2014, 38 (3), 757–772.
- Miller, Nathan H, Matthew Osborne, and Gloria Sheu**, “Pass-through in a Concentrated Industry: Empirical Evidence and Regulatory Implications,” *RAND Journal of Economics*, 2017, 48 (1), 69–93.
- Morikawa, Masayuki**, “Employer-Provided Training and Productivity: Evidence from A Panel of Japanese Firms,” *Journal of the Japanese and International Economies*, 2021, 61, 1–8.

- Nocke, Volker and Nicolas Schutz**, “Multiproduct-firm Oligopoly: An Aggregative Games Approach,” *Econometrica*, 2018, 86 (2), 523–557.
- OECD**, *OECD Employment Outlook 2022* 2022.
- Rhodes, Andrew and Jidong Zhou**, “Personalized pricing and competition,” *American Economic Review*, 2024, 114 (7), 2141–2170.
- Riley, Shawn M, Steven C Michael, and Joseph T Mahoney**, “Human Capital Matters: Market Valuation of Firm Investments in Training and The Role of Complementary Assets,” *Strategic Management Journal*, 2017, 38 (9), 1895–1914.
- Robinson, Joan**, *Economics of Imperfect Competition*, Macmillan, 1933.
- Rothstein, Donna and Evan Starr**, “Noncompete Agreements, Bargaining, and Wages: Evidence from The National Longitudinal Survey of Youth 1997,” *Monthly Labor Review*, 2022.
- Shi, Liyan**, “Optimal Regulation of Noncompete Contracts,” *Econometrica*, 2023, 91 (2), 425–463.
- Shubik, Martin and Richard Levitan**, *Market Structure and Behavior*, Harvard University Press, 2013.
- Shy, Oz and Rune Stenbacka**, “Noncompete Agreements, Training, and Wage Competition,” *Journal of Economics and Management Strategy*, 2023, 32 (2), 328–347.
- Starr, Evan**, “Consider This: Training, Wages, and the Enforceability of Covenants not to Compete,” *ILR Review*, 2019, 72 (4), 783–817.
- Streefkerk, Maarten, Suzan Elshout, and Boukje Cuelenaere**, “Concurrentiebeding. Dataverzameling bij het LISS panel (Non-compete clause: Data collection via the LISS panel),” 2015.
- Suzumura, Kotaro**, “Cooperative and Noncooperative R&D in an Oligopoly with Spillovers,” *American Economic Review*, 1992, 1307–1320.
- Tirole, Jean**, *The theory of industrial organization*, MIT press, 1988.
- Vives, Xavier**, *Oligopoly Pricing: Old Ideas and New Tools*, MIT press, 1999.
- , “Innovation and Competitive Pressure,” *Journal of Industrial Economics*, 2008, 56 (3), 419–469.
- **and Orestis Vravosinos**, “Free entry in a Cournot market with overlapping ownership,” *American Economic Journal: Microeconomics*, forthcoming.
- Vuorenkoski, Vesa**, “Kilpailukielto: ja salassapitoehtojen ongelmat ja ratkaisumallit (Non-compete clauses: Problems and solutions),” *AkavaWorks Artikkelit*, 2018.
- Whinston, Micheal D**, “Antitrust Policy toward Horizontal Mergers,” in Mark Armstrong and Robert H Porter, eds., *Handbook of Industrial Organization*, Vol. 3, Amsterdam: North-Holland, 2007, chapter 36, 2369–2440.

**Young, Samuel**, “Noncompete clauses, job mobility, and job quality: Evidence from a low-earning noncompete ban in austria,” *SSRN Electron J*, 2021.

# A Online Appendix for “Product Market Competition, Labor Mobility, and Firm-Sponsored Training”

## Contents

|        |   |     |
|--------|---|-----|
| A.1    | Proofs . . . . .  | A1  |
| A.1.1  | Proof of Lemma 1 . . . . .                                      | A1  |
| A.1.2  | Proof of Proposition 1 . . . . .                                | A2  |
| A.1.3  | Proof of Proposition 2 . . . . .                                | A4  |
| A.1.4  | Proof of Proposition 3 . . . . .                                | A5  |
| A.1.5  | Proof of Lemma 2 . . . . .                                      | A6  |
| A.1.6  | Proof of Lemma 3 . . . . .                                      | A9  |
| A.1.7  | Proof of Proposition 4 . . . . .                                | A11 |
| A.1.8  | Proof of Proposition 5 . . . . .                                | A12 |
| A.1.9  | Proof of Example 2 . . . . .                                    | A12 |
| A.1.10 | Proof of Proposition 6 . . . . .                                | A13 |
| A.2    | Further details on the model and analysis . . . . .             | A15 |
| A.2.1  | Examples that satisfy Assumption 1 . . . . .                    | A15 |
| A.2.2  | Full characterizations of models used for proofs . . . . .      | A20 |
| A.2.3  | Under- versus overinvestment . . . . .                          | A23 |
| A.2.4  | Condition for switching cost to lower equilibrium NMU . . . . . | A25 |
| A.3    | Segmented labor markets . . . . .                               | A26 |
| A.4    | Model with overlapping ownership . . . . .                      | A28 |
| A.4.1  | Setting . . . . .   | A28 |
| A.4.2  | Equilibrium analysis . . . . .                                  | A29 |
| A.4.3  | Effects of overlapping ownership . . . . .                      | A31 |
| A.5    | Model with initial hiring . . . . .                             | A35 |
| A.5.1  | Setting . . . . .   | A35 |
| A.5.2  | Equilibrium analysis . . . . .                                  | A37 |
| A.5.3  | Effects of labor market power . . . . .                         | A40 |
| A.6    | Model with firm-specific training . . . . .                     | A47 |

## A.1 Proofs

### A.1.1 Proof of Lemma 1

Let  $\Delta v_{j-i} := (v_{ji} - v_{jk})_{k \in \mathcal{N} \setminus \{i\}} \in \mathbb{R}^{n-1}$ . Then, because  $L(v_j) = L(v'_j)$  holds for  $v_j \in \mathbb{R}^n$  and  $v'_j \in \mathbb{R}^n$  such that  $v'_{jk} = v_{jk} + x$ , we have

$$L(v_{ji}, v_{j-i}) = L(0, -\Delta v_{j-i}).$$

At symmetric  $(p, h, w_I, w_O)$ , equations (5) and (6) can be written as

$$w_I = -c'(h)Q(p)h - X_I(\sigma - \Delta w), \tag{A.1}$$

$$w_O = -c'(h)Q(p)h - X_O(\sigma - \Delta w). \tag{A.2}$$

Subtracting equation (A.1) from equation (A.2), we obtain equation (7). Because the expression

$$\Delta w - \Delta X(\sigma - \Delta w)$$

is monotonically increasing in  $\Delta w$  under Assumption 1, equation (7) has a unique solution in  $\Delta w$  whenever it exists.

The retention rate  $L_I^*$  and poaching rate  $L_O^*$  are given by

$$\begin{aligned} L_I^* &= L(w_I^*, w_O^* - \sigma, w_O^* - \sigma, \dots, w_O^* - \sigma) \\ &= L(0, w_O^* - \sigma - w_I^*, w_O^* - \sigma - w_I^*, \dots, w_O^* - \sigma - w_I^*) \\ &= L(0, -y^*, -y^*, \dots, -y^*), \end{aligned}$$

and

$$\begin{aligned} L_O^* &= L(w_O^* - \sigma, w_O^* - \sigma, w_O^* - \sigma, \dots, w_I^*, \dots, w_O^* - \sigma) \\ &= L(0, 0, \dots, \sigma - w_O^* + w_I^*, \dots, 0) \\ &= L(0, 0, \dots, y^*, \dots, 0). \end{aligned}$$

Finally, we show the condition under which the solution to equation (7) with  $L_O^* > 0$  (or equivalently,  $L_I^* < 1$ ) exists. In the proof of Lemma 3, we show that  $L_O^*$  is decreasing in  $\sigma$ . Furthermore, at  $\sigma = 0$ ,  $\Delta w^* = 0$  and  $y^* = 0$  hold, in which case  $L_I^* = L_O^* = L(0, 0, \dots, 0) = 1/n$ . Put together, there exists a critical value  $\bar{\sigma} > 0$  (with the possibility that  $\bar{\sigma} = \infty$ ) such that the solution to equation (7) such that  $L_O^* > 0$  (or equivalently,  $L_I^* < 1$ ) exists if and only if  $\sigma < \bar{\sigma}$ .  $\square$

### A.1.2 Proof of Proposition 1

I show that under regularity assumptions (Assumption 3), there exists a unique solution to the system of equations (8) and (9).

We first show that for any given  $h$ , there exists a unique solution to equation (8), denoted by  $\tilde{p}(h) \in (c(h), \infty)$ , such that  $Q(\tilde{p}(h)) > 0$ . Defining

$$g_p(p, c) = \frac{Q(p)}{-Q_p(p)p} - \frac{p - c}{p},$$

$\tilde{p}(h)$  can be characterized as the solution to equation  $g_p(p, c(h)) = 0$ . Because we assume that  $Q(c(h)) > 0$  for any  $h > 0$ ,  $g_p(c(h), c(h)) = \{Q(c(h))/[Q_p(c(h))c(h)]\} > 0$ . Furthermore, because we assume that  $\lim_{p \rightarrow \infty} \varepsilon_Q(p) > 1$ , we have

$$\lim_{p \rightarrow \infty} g_p(p, c) = \lim_{p \rightarrow \infty} \left( \frac{1}{\varepsilon_Q(p)} - \frac{p - c}{p} \right) < 0.$$

Hence, there exists a solution to the equation  $g_p(p, c(h)) = 0$  for any  $h$ . Furthermore, because  $\varepsilon'_Q(p) > 0$ , we have

$$\frac{\partial g_p(p, c)}{\partial p} = \frac{-\varepsilon'_Q(p)}{[\varepsilon_Q(p)]^2} - \frac{c}{p^2} < 0.$$

Hence, the solution to  $g_p(p, c(h))$  is unique. Equivalently, we can also show that whenever  $g_p(p, c) = 0 \Leftrightarrow \varepsilon_Q(p)(p - c)/p = 1$ , we have

$$\begin{aligned}
& Q'_p(p)(p - c) + Q_p(p) + Q'(p) \\
& < \frac{p - c}{p} \left[ -Q_p(p) + \frac{Q_p(p)pQ'(p)}{Q(p)} \right] + Q_p(p) + Q'(p) \\
& = \frac{c}{p}Q_p(p) + Q'(p) \left[ 1 - \varepsilon_Q(p)\frac{p - c}{p} \right] \\
& = \frac{c}{p}Q_p(p) < 0,
\end{aligned}$$

which follows from the fact that

$$\varepsilon'_Q(p) = \frac{[-Q_p(p) - Q'_p(p)p]Q(p) + Q_p(p)pQ'(p)}{[Q(p)]^2} > 0.$$

Hence, by applying the implicit function theorem to equation (8), we have

$$\tilde{p}'(h) = \frac{Q_p(\tilde{p}(h))c'(h)}{Q'_p(\tilde{p}(h))(\tilde{p}(h) - c) + Q_p(\tilde{p}(h)) + Q'(\tilde{p}(h))} < 0.$$

Next, we consider equation (9) evaluated at  $p = \tilde{p}(h)$ . The equation becomes

$$F(h) := -c'(h)Q(\tilde{p}(h))L_I^* - \tau'(h).$$

Because we assume that  $\lim_{h \rightarrow 0}[\tau'(h)/c'(h)] = 0$  and  $Q(c(h)) > 0$  for any  $h > 0$ , we have

$$\begin{aligned}
\lim_{h \rightarrow 0} F(h) &= \lim_{h \rightarrow 0} -c'(h) \left[ Q(\tilde{p}(h))L_I^* - \frac{\tau'(h)}{-c'(h)} \right] \\
&= \lim_{h \rightarrow 0} -c'(h)Q(\tilde{p}(h))L_I^* \\
&> 0.
\end{aligned}$$

Furthermore, because we assume that  $\lim_{h \rightarrow \infty}\{\tau'(h)/[-c'(h)] - Q(c(h))\} > 0$ , we have

$$\begin{aligned}
\lim_{h \rightarrow \infty} F(h) &= \lim_{h \rightarrow \infty} -c'(h) \left[ Q(\tilde{p}(h))L_I^* - \frac{\tau(h)}{-c'(h)} \right] \\
&< \lim_{h \rightarrow \infty} -c'(h) \underbrace{[Q(\tilde{p}(h))L_I^* - Q(c(h))]}_{(-)} \\
&< 0.
\end{aligned}$$

Finally, Assumption 3-(iii) implies that

$$\begin{aligned}
F'(h) &= -\tilde{p}'c'Q'L_I^* - c''QL^I - \tau'' \\
&= \frac{-1}{Q'_p(p - c) + Q_p + Q'} \left[ [c']^2L_IQ_pQ' + [Q'_p(p - c) + Q_p + Q'] [c''QL_I + \tau''] \right] \\
&< 0.
\end{aligned}$$

Therefore, given  $L_I^*$ , there exists a unique solution to the system of equations (8)-(9). Together with Lemma 1, there exists a unique solution to the system of equations (8), (9), and (7) such that  $L_O^* > 0$ , if and only if  $\sigma < \bar{\sigma}$ .  $\square$

### A.1.3 Proof of Proposition 2

We first show that  $\tilde{p}(h)$  is decreasing in  $h$ . Let  $\tilde{p}(h)$  be the solution to equation (8). Then, by the implicit function theorem, we have

$$\begin{aligned}\tilde{p}'(h^*) &= \frac{Q_p c'}{\frac{Q'_p Q}{Q_p} + Q_p + Q'} \\ &= \frac{c'}{2 - \psi(p) + \frac{1}{\varepsilon_Q(p)} \left[ \alpha_Q(p) \frac{1-\psi(p)}{\psi(p)} + \varepsilon_\psi(p) \right]} \\ &< 0,\end{aligned}$$

where

$$\begin{aligned}\varepsilon_\psi(p) &:= \frac{\psi'(p)p}{\psi(p)}, \\ \alpha_Q(p) &:= \frac{Q''(p)p}{-Q'(p)}.\end{aligned}$$

Hence,  $\tilde{p}(h)$  is decreasing in  $h$ .

The firm profit plus wage bill is defined by

$$\bar{\Pi}(h) := Q(\tilde{p}(h))[\tilde{p}(h) - c(h)] - \tau(h).$$

Then, at  $h^*$ , we have  $\tilde{p}(h^*) = p^*$ , and

$$\begin{aligned}\bar{\Pi}'(h^*) &= \tilde{p}'(p^*) [Q'(p^*)[p^* - c(h^*)] + Q(p^*)] - c'(h^*)Q(p^*) - \tau'(h^*) \\ &= Q(p^*) \{ \tilde{p}'(h^*)\psi(p^*) - c'(h^*)[1 - L_I] \},\end{aligned}$$

where we used the first-order conditions and the relation

$$Q_p(p) = \frac{Q'(p)}{1 - \psi(p)}.$$

Using the expression for  $\bar{p}'(h)$ , we have  $\bar{\Pi}'(h^*) \geq 0$  if and only if

$$L_I \leq \bar{L}_I = 1 - \Psi(p^*).$$

where

$$\Psi(p) = \frac{\psi(p)}{2 - \psi(p) + \frac{1}{\varepsilon_Q(p)} \left[ \alpha_Q(p) \frac{1-\psi(p)}{\psi(p)} + \varepsilon_\psi(p) \right]}.$$

Based on this preliminary analysis, we prove Proposition 2.

1. First, suppose that  $\psi(p) = 0$  for all  $p$ . Then, we have  $\Psi(p) = 0$  and that  $\bar{L}_I(p) = 1$  for all  $p$ , implying that the equilibrium training is always underinvestment.
2. Next, suppose that  $\psi(p) = 1$  for all  $p$ . In this case, we have  $\Psi(p) = 1$  for all  $p$  and that  $\bar{L}_I(p) = 0$  for all  $p$ . Therefore, we have  $\bar{L}_I = 0$ , implying that  $\bar{\Pi}'(h^*) < 0$  always holds.

3. Under the assumption that  $\rho(h)\psi(\tilde{p}(h))$  is increasing in  $h$ , consider the condition for the overinvestment:

$$\rho(h^*)\psi(\tilde{p}(h^*)) \geq 1 - L_I^*.$$

In Lemma 2, we show that  $L^*$  is increasing in  $\sigma$ , so the right-hand side of this condition is decreasing in  $\sigma$ . Furthermore, because we show in Lemma 2 that  $h^*$  is increasing in  $\sigma$ , the left-hand side is increasing in  $\sigma$ , which guarantee the existence of the cutoff  $\sigma^\dagger$  such that  $\Psi(\tilde{p}(h^*)) = 1 - L_I^*$ , and  $h^*$  is overinvestment if and only if  $\sigma > \sigma^\dagger$ . Similarly,  $h^*$  is underinvestment if and only if  $\sigma < \sigma^\dagger$ . □

#### A.1.4 Proof of Proposition 3

Under the Shubik-Levitan demand specification, we have

$$Q(p) = \frac{v-p}{n}, \quad Q'(p) = -\frac{1}{n},$$

$$Q_p(p) = -\frac{1}{n} \left[ 1 + \frac{\gamma}{1-\gamma} \frac{n-1}{n} \right], \quad \psi(p) = \frac{\frac{\gamma}{1-\gamma} \frac{n-1}{n}}{1 + \frac{\gamma}{1-\gamma} \frac{n-1}{n}}.$$

Using this, the pricing formula for  $\tilde{p}(h)$  is written as

$$v - p - \left[ 1 + \frac{\gamma}{1-\gamma} \frac{n-1}{n} \right] (p - c) = 0,$$

and thus we obtain

$$\tilde{p}(h) = \frac{\tilde{p}'(h)}{c'(h)} = \frac{v + \left[ 1 + \frac{\gamma}{1-\gamma} \frac{n-1}{n} \right] c(h)}{2 + \frac{\gamma}{1-\gamma} \frac{n-1}{n}}.$$

Hence, the pass-through rate is given by

$$\rho(h) = \frac{1 + \frac{\gamma}{1-\gamma} \frac{n-1}{n}}{2 + \frac{\gamma}{1-\gamma} \frac{n-1}{n}},$$

implying that we have

$$\rho(h)\psi(p) = \frac{\frac{\gamma}{1-\gamma} \frac{n-1}{n}}{1 + \frac{\gamma}{1-\gamma} \frac{n-1}{n}}.$$

As shown in Proposition A.1, in Chen-Riordan example of Example 1, the equilibrium retention rate  $L_I^*$  is given by

$$L_I^* = \frac{1}{n} + \frac{(n-1)\sigma}{n(2n-1)\theta}.$$

Hence, putting these together, the equilibrium training level is overinvestment if and only if

$$\frac{\frac{\gamma}{1-\gamma} \frac{n-1}{n}}{1 + \frac{\gamma}{1-\gamma} \frac{n-1}{n}} \geq 1 - \frac{1}{n} - \frac{(n-1)\sigma}{n(2n-1)\theta}.$$

Because the left-hand side is increasing in  $\gamma$  and the right-hand side is decreasing in  $\sigma$ , we obtain Proposition 3. □

### A.1.5 Proof of Lemma 2

Here, we first show part (ii) of Lemma 2 and then part (i). Afterward, we show part (iii), followed by the total surplus result.

First, we show that when the wages are given by  $\tilde{w}_I$  and  $\tilde{w}_O$ , the worker surplus  $\widetilde{\text{WS}}$  is decreasing in  $\sigma$ . Note that  $\Delta\tilde{w} := \tilde{w}_O - \tilde{w}_I$  is given by the same condition as that of  $\Delta w^*$  and that  $\Delta X' > 0$  by Assumption 1. Applying the implicit function theorem to equation (7), we have

$$\begin{aligned}\frac{\partial \Delta w^*}{\partial \sigma} &= \frac{-\frac{\partial}{\partial \sigma} [\Delta w - \Delta X(\sigma - \Delta w)]}{\frac{\partial}{\partial \Delta w} [\Delta w - \Delta X(\sigma - \Delta w)]} \\ &= \frac{\Delta X'(\sigma - \Delta w)}{1 + \Delta X'(\sigma - \Delta w)} \\ &\in (0, 1),\end{aligned}$$

which shows that  $\sigma - \Delta w^*$  is increasing in  $\sigma$ , implying part (ii) of the lemma.

Next, we show the comparative statics results for  $\tilde{w}_I$  and  $\tilde{w}_O$ . By equation (A.1),  $\tilde{w}_I$  is given by

$$\tilde{w}_I = -c'(h)hQ(\tilde{p}(h)) - X_I(\sigma - \Delta\tilde{w}).$$

Therefore, we have

$$\begin{aligned}\frac{\partial \tilde{w}_I}{\partial \sigma} &= -X'_I(\sigma - \Delta w) \times \left(1 - \frac{\partial \Delta\tilde{w}}{\partial \sigma}\right) \\ &= -\underbrace{X'_I(\sigma - \Delta w)}_{(+)} \frac{1}{1 + \Delta X'(\sigma - \Delta w)} \\ &< 0,\end{aligned}$$

and

$$\begin{aligned}\frac{\partial \tilde{w}_O}{\partial \sigma} &= -X'_O(\sigma - \Delta w) \times \left(1 - \frac{\partial \Delta\tilde{w}}{\partial \sigma}\right) \\ &= -\underbrace{X'_O(\sigma - \Delta w)}_{(-)} \frac{1}{1 + \Delta X'(\sigma - \Delta w)'} \\ &> 0.\end{aligned}$$

By noting that  $X_I(0) = X_O(0)$ , we have  $\tilde{w}_I(0, h) = \tilde{w}_O(0, h)$ . Together with Lemma (ii), we have  $\tilde{w}_I(\sigma, h) < \tilde{w}_O(\sigma, h)$  for all  $\sigma > 0$ , which shows (i) of the lemma.

Finally, we show the welfare results of Lemma 2

1. We first show that  $\widetilde{\text{WS}}(\sigma, h)$  is decreasing in  $\sigma$ . The effect of  $\sigma$  on  $\widetilde{\text{WS}}(\sigma, h)$  is given by

$$\begin{aligned}
\frac{\partial \widetilde{\text{WS}}}{\partial \sigma} &= \underbrace{-(1 - L_I)}_{\text{direct effect} < 0} + \underbrace{\frac{\partial \tilde{w}_I}{\partial \sigma} L_I + \frac{\partial \tilde{w}_O}{\partial \sigma} (1 - L_I)}_{\text{indirect effects of wage changes}} \\
&= -(1 - L_I) + (1 - L_I) \frac{\partial \Delta w}{\partial \sigma} + \frac{\partial \tilde{w}_I}{\partial \sigma} \\
&= -(1 - L_I) \underbrace{\left(1 - \frac{\partial \Delta w}{\partial \sigma}\right)}_{(+)} + \underbrace{\frac{\partial w_I}{\partial \sigma}}_{(-)} < 0,
\end{aligned}$$

which shows that  $\widetilde{\text{WS}}(\sigma, h)$  is decreasing in  $\sigma$ .

Because  $\widetilde{\text{CS}}(\sigma, h)$  depends on  $\sigma$  only through  $\bar{p}(h)$ , it is constant when  $h$  is fixed.

2. We have

$$\begin{aligned}
\frac{\partial \widetilde{\text{IP}}}{\partial \sigma} &= -\frac{\partial \widetilde{\text{WB}}}{\partial \sigma} \\
&= \left(1 - \frac{\partial \Delta w^*}{\partial \sigma}\right) L'_I(\sigma - \Delta w) \Delta w - \frac{\partial \tilde{w}_I}{\partial \sigma} L_I + \frac{\partial \tilde{w}_O}{\partial \sigma} (1 - L_I) \\
&= \frac{L_I}{1 + \Delta X'} \left[ \frac{X_I - X_O}{X_I} + X'_I + X'_O \frac{1 - L_I}{L_I} \right],
\end{aligned}$$

which may be either positive or negative, depending on the value of  $X'_I$  and  $X'_O$ . In the Hotelling setting of Example A.1, this is always positive. In Example A.5, we have

$$\frac{\partial \widetilde{\text{IP}}}{\partial \sigma} \propto -(2L_I^* - 1) < 0$$

around  $\sigma \simeq 0$ .

3. Finally, we show that  $\widetilde{\text{TS}}(\sigma, h)$  can either decrease or increase with  $\sigma$ .

Note that  $\widetilde{\text{TS}}$  is given by

$$\widetilde{\text{TS}} = \widetilde{\text{CS}}(\sigma, h) + \widetilde{\text{WS}}(\sigma, h) + \bar{\Pi}(h) - \widetilde{\text{WB}}(\sigma, h)$$

where

$$\widetilde{\text{WB}}(\sigma, h) = \tilde{w}_I(\sigma, h) L_I^* + \tilde{w}_O(1 - L_I^*)$$

Hence, defining

$$\text{NMU}(\sigma, h) := \widetilde{\text{WS}}(\sigma, h) - \widetilde{\text{WB}}(\sigma, h),$$

the effect of  $\sigma$  on  $\widetilde{\text{TS}}$  can be simplified to

$$\begin{aligned}
\frac{\partial \widetilde{\text{TS}}}{\partial \sigma} &= \frac{\partial \widetilde{\text{NMU}}}{\partial \sigma} \\
&= -(1 - L_I) + \left(1 - \frac{\partial \Delta w^*}{\partial \sigma}\right) L'_I(\sigma - \Delta w) \Delta w \\
&= -(1 - L_I) + \frac{1}{1 + \Delta X'} L'_I(\sigma - \Delta w) \Delta w \\
&= -(1 - L_I) + \frac{1}{1 + \Delta X'} L'_I(\sigma - \Delta w) [X_I - X_O]
\end{aligned}$$

At  $\sigma = 0$ , we have  $X_I = X_O$  and  $L_I = 1/n$ , so we have

$$\frac{\partial \widetilde{\text{TS}}}{\partial \sigma} = -\frac{n-1}{n} < 0.$$

Therefore, by the continuity of the model, there exists  $\tilde{\sigma} > 0$  such that  $\partial \widetilde{\text{TS}}/\partial \sigma < 0$  for all  $\sigma < \tilde{\sigma}$ .

The existence of the case where  $\partial \widetilde{\text{TS}}/\partial \sigma > 0$  is shown in Online Appendix A.2.4.

To show  $\widetilde{\text{TS}}(\sigma, h) < \widetilde{\text{TS}}(0, h)$  for all  $\sigma > 0$ , it suffices to show that  $\text{NMU}(\sigma, h) := \widetilde{\text{WS}}(\sigma, h) - \widetilde{\text{WB}}(\sigma, h)$  satisfies  $\text{NMU}(\sigma, h) < \text{NMU}(0, h)$  for all  $\sigma > 0$ . To this end, we introduce several additional notations.

First, define  $\text{ws}(w_I, w_O, \sigma) := \text{WS}(w_I, w_O, \dots, w_O; \sigma)$  and wage bill  $\text{wb}(w_I, w_O; \sigma) := L_I(\sigma - w_O + w_I)w_I + [1 - L_I(\sigma - w_O + w_I)]w_O$  as functions of  $(w_I, w_O, \sigma)$  and let non-monetary worker utility be denoted as

$$\text{nu}(w_I, w_O; \sigma) := \text{ws}(w_I, w_O; \sigma) - \text{wb}(w_I, w_O; \sigma).$$

Because we have

$$\begin{aligned}
\frac{\partial \text{nu}}{\partial w_I} &= -L'_I(\sigma - \Delta w)(w_I - w_O) \\
&= L'_I(\sigma - \Delta w) \Delta w, \\
\frac{\partial \text{nu}}{\partial w_O} &= L'_I(\sigma - \Delta w)(w_I - w_O) \\
&= -L'_I(\sigma - \Delta w) \Delta w,
\end{aligned}$$

we have

$$\begin{aligned}
\text{nu}(w_I, w_O; \sigma) &= \text{nu}(w_I, \Delta w + w_I; \sigma) \\
&= \text{nu}(0, \Delta w; \sigma) + \int_0^{w_I} \left( \frac{\partial \text{nu}(x, \Delta w + x; \sigma)}{\partial w_I} + \frac{\partial \text{nu}(x, \Delta w + x; \sigma)}{\partial w_O} \right) dx \\
&= \text{nu}(0, \Delta w; \sigma)
\end{aligned}$$

Letting  $\text{nu}_0(\Delta w; \sigma) := \text{nu}(0, \Delta w; \sigma)$ , we have

$$\begin{aligned}\frac{\partial \text{nu}_0}{\partial \sigma} &= -L_I + L'_I \Delta w, \\ \frac{\partial \text{nu}_0}{\partial \Delta w} &= -L'_I(\sigma - \Delta w) \Delta w.\end{aligned}$$

Hence,  $\text{nu}_0(0, \sigma)$  is decreasing in  $\sigma$ , and  $\text{nu}_0(\Delta, \sigma)$  is decreasing in  $\Delta w$  for all  $\Delta w > 0$ . Finally, for any  $\sigma > 0$  and  $\Delta w > 0$ , we have

$$\begin{aligned}\text{nu}_0(\Delta w; \sigma) &= \text{nu}_0(0; \sigma) + \int_0^{\Delta w} \frac{\partial \text{nu}_0(x; \sigma)}{\partial \Delta w} dx \\ &< \text{nu}_0(0; \sigma) \\ &= \text{nu}_0(0; 0) + \int_0^{\sigma} \frac{\partial \text{nu}_0(0; \sigma)}{\partial \sigma} dx \\ &< \text{nu}_0(0; 0)\end{aligned}$$

which establishes that  $\text{NMU}(\sigma, h) = \text{nu}_0(\Delta w^*; \sigma) < \text{nu}_0(0; 0) = \text{NMU}(0, h)$ , implying  $\widetilde{\text{TS}}(\sigma, h) < \widetilde{\text{TS}}(0, h)$ .  $\square$

### A.1.6 Proof of Lemma 3

Applying the implicit function theorem to equation (7), we have

$$\frac{d\Delta w^*}{d\sigma} = \frac{\Delta X'}{1 + \Delta X'} \in (0, 1).$$

Therefore, we have

$$\frac{dL_I^*}{d\sigma} = L'_I(\sigma - \Delta w^*) \frac{d}{d\sigma} (\sigma - \Delta w^*) > 0.$$

Hence,  $L_I^*$  is increasing in  $\sigma$ , establishing the first statement of part (i) of the lemma. Because  $L_O^* = (1 - L_I^*)/(n - 1)$ , it is decreasing in  $\sigma$ .

Next, we show that  $p^*$  is decreasing in  $\sigma$  and that  $h^*$  is increasing in  $\sigma$ . Given the equilibrium value of  $L_I^*$ , the equilibrium condition for  $p^*$  and  $h^*$  is written as

$$\begin{aligned}f_p &= Q_p(p)[p - c(h)] + Q(p) = 0, \\ f_h &= -c'(h)Q(p)L_I^* - \tau'(h) = 0,\end{aligned}$$

Thus, we have

$$J \begin{pmatrix} \frac{dp^*}{d\sigma} \\ \frac{dh^*}{d\sigma} \end{pmatrix} = - \begin{pmatrix} 0 \\ -c'(h)Q(p) \frac{dL_I^*}{d\sigma} \end{pmatrix},$$

where

$$J := \begin{pmatrix} f_{pp} & f_{ph} \\ f_{hp} & f_{hh} \end{pmatrix}$$

and  $f_{xy}$  is  $\partial f_x / \partial y$ , for  $x, y \in \{p, h\}$ . Using Cramer's Rule, we have

$$\begin{aligned}\frac{dp^*}{d\sigma} &= \frac{-f_{ph}c'(h)Q(p) \frac{dL_I^*}{d\sigma}}{\det J} < 0, \\ \frac{dh^*}{d\sigma} &= \frac{f_{pp}c'(h)Q(p) \frac{dL_I^*}{d\sigma}}{\det J} > 0,\end{aligned}$$

where the first inequality follows from the fact that  $f_{ph} < 0$  and  $\det J > 0$  by the regularity of the equilibrium, and the second inequality follows from the fact that the regularity,  $\det J > 0$ , also implies that  $f_{pp} < 0$ . Hence,  $h^*$  increases with  $\sigma$  and  $p^*$  decreases with  $\sigma$ , where the former establishes the second statement of part (i) of the lemma. Because  $p^*$  decreases with  $\sigma$ ,  $Q^*$  increase with  $\sigma$ .

The second part of Lemma 3 that  $CS^*$  increases with  $\sigma$  simply follows from the fact that  $p^*$  decreases with  $\sigma$ , establishing part (ii) of the lemma.

Finally, we show part (iii) of the lemma that  $WS^*$  and  $IP^*$  may increase or decrease depending on parameterizations. To this end, we consider a duopolistic setting listed in Table 3. In all of the following parameterizations, the equilibrium retention rate  $L_I^*$  is given by  $L_I^* = 1/2 + \sigma/(6\theta)$ .

| $D_i$             | $L_{ji}$   | $c$                              | $\tau$                           |
|-------------------|--|----------------------------------|----------------------------------|
| $A \exp(1 - p_i)$ | $\frac{1}{2} + \frac{v_{ji} - v_{j-i}}{2\theta}$ | $\log \bar{C} - \alpha \log H_i$ | $\frac{\kappa}{\delta} h^\delta$ |

Table 3: A duopolistic specification that leads to underinvestments.

|                            | $h^*$  | $p^*$        | $w_I^*$   | $w_O^*$   |
|----------------------------|--|--------------|---|---|
| $\alpha = 1/2, \delta = 1$ | $\left(\frac{AL_I}{2\kappa\bar{C}}\right)^2$ | $c(h^*) + 1$ | $\frac{Ah^{1/2}}{2\bar{C}} - \theta - \frac{\sigma}{3}$ | $\frac{Ah^{1/2}}{2\bar{C}} - \theta + \frac{\sigma}{3}$ |
| $\alpha = 1, \delta = 2$   | $\frac{AL_I}{\kappa\bar{C}}$                 | $c(h^*) + 1$ | $\frac{Ah}{\bar{C}} - \theta - \frac{\sigma}{3}$        | $\frac{Ah}{\bar{C}} - \theta + \frac{\sigma}{3}$        |

Table 4: Symmetric equilibrium outcomes in the examples.

Now, consider the case where  $\alpha = 1/2$ ,  $\delta = 1$ . In this case, the equilibrium variables are given by the first row of Table 4. Therefore, the equilibrium profit of each firm is

$$\Pi^* = \frac{A^2}{2\kappa\bar{C}^2} L_I^* - L_I^* w_I^* - L_O^* w_O^* - \kappa h^*,$$

which has the following derivative with respect to  $\sigma$ :

$$\frac{d\Pi^*}{d\sigma} = -\frac{\sigma}{6\theta} \frac{A^2}{12\theta\kappa\bar{C}^2} + \frac{2\sigma}{9\theta},$$

which is positive if and only if

$$16\theta\kappa > \frac{A^2}{\bar{C}^2}.$$

The derivative of worker surplus with respect to  $\sigma$  is given by

$$\frac{dWS^*}{d\sigma} = \frac{A^2}{4\kappa\bar{C}^2} \frac{1}{6\theta} - \frac{1}{2} - \frac{\sigma}{18\theta}.$$

If

$$12\theta\kappa < \frac{A^2}{\bar{C}^2}.$$

hold, then there exists  $\sigma_L > 0$  such that if  $\sigma < \sigma_L$ , then worker surplus increases with  $\sigma$ . Therefore, if there exists  $\nu \in (12, 16)$  such that

$$\frac{A^2}{\bar{C}^2} = \nu\theta\kappa,$$

and  $\sigma$  is sufficiently small, both worker surplus and industry profit increase with  $\sigma$ . When  $\sigma \simeq 0$  and  $\nu < 12$ , then industry profit increases with  $\sigma$  and workers surplus decreases with  $\sigma$ . When  $\nu > 16$ , industry profit decreases with  $\sigma$  and worker surplus increases with  $\sigma$ . These results show that each of  $WS^*$  and  $IP^*$  can increase or decrease with  $\sigma$  depending on parameterizations.

Lastly, we check that the second-order condition for the optimality is satisfied in our analysis (see Online Appendix A.2.2 for further details). To satisfy the second-order condition around  $L_I = 1/2$ , we must have

$$\frac{A^2}{\kappa \bar{C}^2} + \frac{A^4}{4\kappa^2 \bar{C}^4} - 8\theta\kappa < 0$$

Suppose that  $A^2/\bar{C}^2 = \nu\theta\kappa$  for  $\nu \in (12, 16)$ , the above condition can be rewritten as

$$\theta\kappa \left[ \frac{\nu}{\kappa} + \frac{\nu^2\theta}{4\kappa} - 8 \right] < 0.$$

Therefore, as long as  $\kappa$  is sufficiently large, this condition is guaranteed.

To show that  $TS^*$  can be maximized at  $\sigma \neq 0$ , we use the setting with  $\alpha = 1$ ,  $\delta = 2$  of the model in Table 3 to show the existence of both possibilities. In this case, the equilibrium variables are given by the second row of Table 4. Then, The equilibrium total surplus is given by

$$TS^* = n \left[ \frac{2A^2}{\kappa \bar{C}^2} L_I + WS - L_I w_I - (1 - L_O) w_O \right]$$

At  $\sigma = 0$ , we have

$$\frac{dTS^*}{d\sigma} = \frac{A^2}{3\theta\kappa C^2} - \frac{1}{2},$$

which is positive if and only if

$$\frac{A^2}{C^2} > \frac{3\theta\kappa}{2},$$

establishing the last point of the lemma.  $\square$

### A.1.7 Proof of Proposition 4

In the proof of Lemma 3, we already showed the example in which equilibrium worker surplus and industry profit increase with  $\sigma$  simultaneously.

Next, we show that when the equilibrium training level is overinvestment, an increase in switching cost  $\sigma$  does never increase both the equilibrium industry profit and the equilibrium worker surplus when  $\sigma < \hat{\sigma}$ . To see this, note that the sum of the industry profit and the worker surplus is given by

$$n \left[ \bar{\Pi}(h^*) + NMU(\sigma) \right].$$

Note that because we have the relation  $NMU = WS - WB$ , we have  $NMU'(\sigma) = d(WS - WB)/d\sigma := \partial \widetilde{TS}/\partial \sigma$ . Then, as we showed in the proof of Lemma 2,  $\partial \widetilde{TS}/\partial \sigma$  when  $\sigma < \tilde{\sigma}$ , implying that  $NMU'(\sigma) < 0$  when  $\sigma < \tilde{\sigma}$ . Furthermore,  $\bar{\Pi}'(h^*) < 0$  when the equilibrium training level is overinvestment. Because the sum of worker surplus and industry profits must decrease with  $\sigma$ , it is never the case that both of them increase with  $\sigma$  when  $\sigma < \tilde{\sigma}$ .  $\square$

### A.1.8 Proof of Proposition 5

For a given symmetric training level  $h$ , let  $\tilde{p}(h, \zeta)$  be the solution to the equation

$$f_p = \tilde{Q}_p(p; \zeta)[p - c(h)] + \tilde{Q}(p; \zeta) = 0.$$

Then, the equilibrium training level  $h^*$  is given by the solution to the equation

$$f_h = -c'(h)\tilde{Q}(\tilde{p}(h, \zeta); \zeta)L_I^* - \tau'(h) = 0,$$

Therefore, we have

$$\frac{dh^*}{d\zeta} = \frac{c'L_I^* \left[ \frac{\partial \tilde{Q}}{\partial \zeta} + \frac{\partial \tilde{p}}{\partial \zeta} \frac{\partial \tilde{Q}}{\partial p} \right]}{-c''\tilde{Q}L_I^* - \tau'' - c' \frac{\partial \tilde{Q}}{\partial p} \frac{\partial \tilde{p}}{\partial h}},$$

which has the same sign with

$$\frac{\partial \tilde{Q}}{\partial \zeta} + \frac{\partial \tilde{p}}{\partial \zeta} \frac{\partial \tilde{Q}}{\partial p}.$$

The effect of  $\zeta$  on the equilibrium output is given by

$$\frac{dQ^*}{d\zeta} = \frac{\partial \tilde{Q}}{\partial \zeta} + \frac{\partial \tilde{p}}{\partial \zeta} \frac{\partial \tilde{Q}}{\partial p} + \frac{dh^*}{d\zeta} \frac{\partial \tilde{p}}{\partial h} \frac{\partial \tilde{Q}}{\partial p},$$

with again has the same with with

$$\frac{\partial \tilde{Q}}{\partial \zeta} + \frac{\partial \tilde{p}}{\partial \zeta} \frac{\partial \tilde{Q}}{\partial p}.$$

Finally, we have

$$\frac{dw_I^*}{d\zeta} = \frac{dw_O^*}{d\zeta} = \frac{dh^*}{d\zeta} [(-c' - c''h)Q^*] - \frac{dQ^*}{d\zeta} c'h,$$

which is positive if and only if  $dQ^*/d\zeta$  is positive. □

### A.1.9 Proof of Example 2

1. When the demand function is given by Shubik-Levitan specification, we have  $Q(p) = (v - p)/n$ , and

$$Q_p(p) = \frac{-1}{n} \left[ 1 + \frac{\gamma}{1 - \gamma} \frac{n - 1}{n} \right].$$

Then,  $\tilde{p}$  is given by the solution to the system of equation

$$- \left[ 1 + \frac{\gamma}{1 - \gamma} \frac{n - 1}{n} \right] (p - c) + v - p = 0.$$

By applying the implicit function theorem, we obtain

$$\frac{\partial \tilde{p}}{\partial \gamma} = -\frac{\frac{1}{(1-\gamma)^2} \frac{n-1}{n} (p-c)}{1 + \frac{\gamma}{1-\gamma} \frac{n-1}{n} + 1} < 0.$$

Hence,

$$\frac{\partial \tilde{Q}(\tilde{p}; \gamma)}{\partial \gamma} = \frac{\partial \tilde{Q}}{\partial p} \frac{\partial \tilde{p}}{\partial \gamma} > 0.$$

2. From Proposition 4 of [Nocke and Schutz \(2018\)](#), we know that the product-market profit  $\pi := [\tilde{p}(h) - c(h)]\tilde{Q}(\tilde{p}(h), V_0)$  decreases with  $V_0$ . Furthermore, from the optimality condition, under logit specification, we have

$$\pi = \frac{\mu \tilde{Q}(\tilde{p}(h), V_0)}{1 - \tilde{Q}(\tilde{p}(h), V_0)},$$

which is increasing in  $\tilde{Q}$ . This establishes that  $\tilde{Q}(\tilde{p}, V_0)$  is decreasing in  $V_0$ .  $\square$

#### A.1.10 Proof of Proposition 6

We first show that in the case of overinvestment, an increase in  $\zeta$  always lowers industry profit when  $dQ^*/d\zeta > 0$ . The industry profit can be written as

$$\text{IP}^* = \bar{\Pi}(h^*) - \text{WB}^*.$$

Noting that

$$\bar{\Pi}(h^*) = \tilde{Q}(p^*; \zeta)(p^* - c(h^*)) - \tau(h^*),$$

we have

$$\begin{aligned} \frac{d\bar{\Pi}}{d\zeta} &= \frac{dh^*}{d\zeta} \bar{\Pi}'(h^*) + \frac{\partial \tilde{Q}}{\partial \zeta}(p^* - c(h^*)) + \frac{\partial \tilde{p}}{\partial \zeta} \psi(p)Q(p) \\ &< \frac{dh^*}{d\zeta} \bar{\Pi}'(h^*). \end{aligned}$$

From Proposition 5,  $dh^*/d\zeta > 0$  when  $dQ^*/d\zeta > 0$ . Hence,  $d\bar{\Pi}(h^*)/d\zeta < 0$  in the case of overinvestment. Also, we know from the proof of Proposition 5 that  $dw_I^*/d\zeta = dw_O^*/d\zeta > 0$  when  $dQ^*/d\zeta > 0$ . Hence,  $d\text{WB}^*/d\zeta > 0$ . Put together, we have

$$\begin{aligned} \frac{d\text{IP}^*}{d\zeta} &= \frac{d\bar{\Pi}(h^*)}{d\zeta} - \frac{d\text{WB}^*}{d\zeta} \\ &< 0. \end{aligned}$$

We next show that in the case of underinvestment, an increase in  $\zeta$  may increase the industry profit.

We consider the Shubik-Levitan demand function with  $c(h) = \bar{C} - \eta h^{\alpha_c}/\alpha_c$  and  $\tau(h) = \kappa h^{\alpha_\tau}/\alpha_\tau$ , where  $\alpha_c \in (0, 1]$  and  $\alpha_\tau > \alpha_c$ . Then, we have  $Q(p) = (v - p)/n$ , and

$$Q_p(p) = \frac{-1}{n} \left[ 1 + \frac{\gamma}{1 - \gamma} \frac{n - 1}{n} \right].$$

We consider the effect of the increase in  $\gamma$  at the neighborhood of  $\gamma = 0$ , where  $Q_p(p) = Q'(p)$  holds in such a case.

The solution to  $Q_p(p)[p - c(h)] + Q(p) = 0$ ,  $\bar{p}(h; \gamma)$ , is given by the solution to

$$v - p - \left( 1 + \frac{\gamma}{1 - \gamma} \frac{n - 1}{n} \right) [p - c(h)] = 0.$$

The equilibrium training level  $h^*$  is given by the condition

$$\begin{aligned} & -c'(h^*)L_I^*Q(\bar{p}(h^*; \gamma)) - \tau'(h^*) \\ & = \eta(h^*)^{\alpha_c - 1}L_I^*Q^* - \kappa(h^*)^{\alpha_\tau - 1}, \end{aligned}$$

which implies that

$$Q^* = \frac{\kappa}{\eta L_I^*} h^{\alpha_\tau - \alpha_c}$$

and thus

$$\frac{dQ^*}{d\gamma} = \frac{\kappa}{\eta L_I^*} \frac{dh^*}{d\gamma} (\alpha_\tau - \alpha_c) h^{\alpha_\tau - \alpha_c - 1}.$$

We use this relation to establish the effects of  $\gamma$  on the profit. Because we have

$$\Pi^* = Q(p^*)[p^* - c(h^*)] - \tau(h^*) - L_I^*w_I^* - (1 - L_I^*)w_O^*,$$

the effect of  $\gamma$  on  $\Pi^*$  can be written as

$$\begin{aligned} \left. \frac{d\Pi^*}{d\gamma} \right|_{\gamma=0} &= \frac{dp^*}{d\gamma} \underbrace{[Q' \cdot (p - c) + Q]}_{=0} + \frac{dh^*}{d\gamma} [-c'(h^*)Q^* - \tau'(h^*)] - \frac{dw_I^*}{d\gamma} \\ &= \frac{dh^*}{d\gamma} \eta(h^*)^{\alpha_c - 1} Q^* (1 - L_I^*) - \frac{d}{d\gamma} (\eta h^{\alpha_c} Q^*) \\ &= \frac{dh^*}{d\gamma} \eta(h^*)^{\alpha_c - 1} Q^* (1 - L_I^* - \alpha_c) - \frac{dQ^*}{d\gamma} \eta(h^*)^{\alpha_c} \\ &= \frac{dh^*}{d\gamma} \eta \left[ (h^*)^{\alpha_c - 1} Q^* (1 - L_I^* - \alpha_c) - \frac{\kappa}{\eta L_I^*} (\alpha_\tau - \alpha_c) (h^*)^{\alpha_\tau - 1} \right] \\ &= \frac{dh^*}{d\gamma} \eta Q^* (h^*)^{\alpha_c - 1} \left[ 1 - L_I^* - \alpha_c - (\alpha_\tau - \alpha_c) \frac{\kappa}{\eta L_I^*} \frac{(h^*)^{\alpha_\tau - \alpha_c}}{Q^*} \right] \\ &= \frac{dh^*}{d\gamma} \eta Q^* (h^*)^{\alpha_c - 1} [1 - L_I^* - \alpha_\tau] \end{aligned}$$

which is positive if  $L_I^* + \alpha_\tau < 1$ . □

## A.2 Further details on the model and analysis

### A.2.1 Examples that satisfy Assumption 1

Here, we provide a list of examples that satisfy Assumption 1.

**Example A.1.** *Hotelling (1929).* Suppose that  $n = 2$ ,

$$\varepsilon_{jil} = \begin{cases} \bar{V} - \theta x & \text{if } i = 1, \text{ and} \\ \bar{V} - \theta(1 - x) & \text{if } i = 2, \end{cases}$$

where  $\bar{V} > 0$  and  $\theta > 0$  are constants, and  $x \sim U[0, 1]$ . Then, assuming that  $\bar{V}$  is sufficiently large,  $L(v_{ji}, v_{j-i})$  is given by

$$L(v_{ji}, v_{j-i}) = \frac{1}{2} + \frac{v_{ji} - v_{jk}}{2\theta},$$

where  $k \neq i$ .

In Example A.1, we have

$$L(0, -y, -y, \dots, -y) = \frac{1}{2} + \frac{y}{2\theta}$$

and

$$\frac{\partial L_{ii}}{\partial w_{ii}} = \frac{\partial L_{ji}}{\partial w_{ij}} = \frac{1}{2\theta}$$

for all  $i$  and  $j \neq i$ . Therefore,

$$X_I(y) = \theta + y$$

is increasing in  $y$ , and

$$X_O(y) = \theta - y$$

is decreasing in  $y$ .

The wage difference is thus given by

$$\Delta w - \Delta X(\sigma - \Delta w) = \Delta w - 2(\sigma - \Delta w) = 0,$$

implying that

$$\Delta w^* = \frac{2\sigma}{3}$$

and

$$L_I^* = \frac{1}{2} + \frac{\sigma}{6\theta}.$$

□

**Example A.2.** Suppose that consumer  $l$  are uniformly located on one of  $n(n - 1)/2$  Hotelling lines connecting two firms. Let  $\mathcal{N}_l = \{i_{l1}, i_{l2}\}$  be the firms at the edge of the Hotelling line the consumer  $l$  is located. The value of  $\varepsilon_{jil}$  is given by

$$\varepsilon_{jil} = \begin{cases} \bar{V} - \theta x & \text{if } i = i_{l1}, \\ \bar{V} - \theta(1 - x) & \text{if } i = i_{l2}, \\ 0 & \text{otherwise,} \end{cases}$$

where  $\bar{V} > 0$  and  $\theta > 0$  are constants, and  $x \sim U[0, 1]$ . Then, assuming that  $\bar{V}$  is sufficiently large,  $L(v_i, v_j)$  is given by

$$\begin{aligned} L(v_{ji}, v_{j-i}) &= \frac{1}{n} + \frac{\sum_{k \neq i} v_{ji} - v_{jk}}{n(n-1)\theta} \\ &= \frac{1}{n} + \frac{\sum_{k \neq i} \Delta v_{jik}}{n(n-1)\theta}. \end{aligned}$$

In Example A.2, we have

$$L(0, -y, -y, \dots, -y) = \frac{1}{n} + \frac{y}{n\theta}$$

and

$$\frac{\partial L_{ii}}{\partial w_{ii}} = \frac{\partial L_{ji}}{\partial w_{ij}} = \frac{1}{n\theta}$$

for all  $i$  and  $j \neq i$ . Therefore,

$$X_I(y) = \theta + y$$

is increasing in  $y$ , and

$$X_O(y) = \theta - \frac{y}{n-1}$$

is decreasing in  $y$ .

The wage difference is thus given by

$$\Delta w - \Delta X(\sigma - \Delta w) = \Delta w - (\sigma - \Delta w) \frac{n}{n-1} = 0,$$

implying that

$$\Delta w^* = \frac{n\sigma}{2n-1}$$

and

$$L_I^* = \frac{1}{n} + \frac{(n-1)\sigma}{n(2n-1)\theta}$$

Given  $(w_I, w_O)$  and  $\Delta w = w_O - w_I$ , worker surplus is given by

$$\begin{aligned} \text{WS} &= L_I w_I + (1 - L_I) w_O + \theta - \frac{2\theta}{n} \left( \int_0^{\frac{1}{2} + \frac{\sigma - \Delta w}{\theta}} x dx + \int_0^{\frac{1}{2} - \frac{\sigma - \Delta w}{\theta}} x dx \right) - \frac{(n-2)\theta}{4n} \\ &= L_I w_I + (1 - L_I)(w_O - \sigma) + \theta - \frac{\theta}{n} \left[ \left( \frac{1}{2} + \frac{\sigma - \Delta w}{\theta} \right)^2 + \left( \frac{1}{2} - \frac{\sigma - \Delta w}{\theta} \right)^2 + \frac{n-2}{4} \right] \end{aligned}$$

□

**Example A.3.** *Perloff and Salop (1985).* Suppose that  $\varepsilon_{jil}$  follows an i.i.d. distribution function  $F$  that is continuously differentiable and has log-concave density function  $f$  in the support  $\mathbb{R}$ . Then,  $L(v_{ji}, v_{j-i})$  is given by

$$L(v_{ji}, v_{j-i}) = \int_{-\infty}^{\infty} \prod_{k \in \mathcal{N} \setminus \{i\}} F(\epsilon + v_{ji} - v_{jk}) f(\epsilon) d\epsilon.$$

In Example A.3, we have

$$\begin{aligned} X_I(y) &= \frac{L_I}{\partial L_{ii}/\partial w_{ii}} \\ &= \frac{\int_{-\infty}^{\infty} (n-1)f(\epsilon)F^{n-2}(\epsilon)[1-F(\Delta w - \sigma + \epsilon)]d\epsilon}{\int_{-\infty+\sigma-\Delta w}^{\infty} (n-1)f(\epsilon)F^{n-2}(\epsilon)f(\Delta w - \sigma + \epsilon)d\epsilon} \\ &= \frac{\int_{-\infty}^{\infty} (n-1)f(\epsilon)F^{n-2}(\epsilon)[1-F(\Delta w - \sigma + \epsilon)]d\epsilon}{\int_{-\infty+\sigma-\Delta w}^{\infty} (n-1)f(\epsilon)F^{n-2}(\epsilon)f(\Delta w - \sigma + \epsilon)d\epsilon}. \end{aligned}$$

We first show that  $X_I(y)$  is increasing in  $y$ . To this end, we show that

$$\frac{1}{X_I(y)} = \int_{-\infty}^{\infty} \frac{f(\epsilon - y)}{1 - F(\epsilon - y)} \left[ \frac{(n-1)f(\epsilon)F^{n-2}(\epsilon)[1 - F(\epsilon - y)]}{\int_{-\infty}^{\infty} (n-1)f(\epsilon)F^{n-2}(\epsilon)[1 - F(\epsilon - y)]d\epsilon} \right] d\epsilon$$

is decreasing in  $y$ . Letting  $\tilde{\epsilon} := \epsilon_{(n-1)} - y$ , we can rewrite  $1/X_I$  as

$$\frac{1}{X_I(y)} = \int_{-\infty}^{\infty} \frac{f(\tilde{\epsilon})}{1 - F(\tilde{\epsilon})} \left[ \frac{(n-1)f(\tilde{\epsilon} + y)F^{n-2}(\tilde{\epsilon} + y)[1 - F(\tilde{\epsilon})]}{\int_{-\infty}^{\infty} (n-1)f(\tilde{\epsilon} + y)F^{n-2}(\tilde{\epsilon} + Y)[1 - F(\tilde{\epsilon})]d\epsilon} \right] d\tilde{\epsilon}$$

Let

$$\begin{aligned} H_I(x; y) &:= \Pr(\tilde{\epsilon} < x | \tilde{\epsilon} < \epsilon_I) \\ &= \frac{\int_{-\infty}^x (n-1)f(\tilde{\epsilon} + y)F^{n-2}(\tilde{\epsilon} + y)[1 - F(\tilde{\epsilon})]d\tilde{\epsilon}}{\int_{-\infty}^{\infty} (n-1)f(\tilde{\epsilon} + y)F^{n-2}(\tilde{\epsilon} + y)[1 - F(\tilde{\epsilon})]d\tilde{\epsilon}}. \end{aligned}$$

Then, we can further rewrite  $1/X_I$  as

$$\frac{1}{X_I(y)} = \int_{-\infty}^{\infty} \frac{f(\tilde{\epsilon})}{1 - F(\tilde{\epsilon})} dH_I(\tilde{\epsilon}; y).$$

The log-concavity of  $f(\cdot)$  implies the log-concavity of  $1 - F(\cdot)$ , and  $f(\cdot)/[1 - F(\cdot)]$  is an increasing function. Thus, if  $H_I$  is increasing in  $y$  for all  $x \in [-\infty, \infty]$ ,  $1/X_I(y)$  is decreasing in  $y$  because an increase in  $y$  decreases  $\tilde{\epsilon}$  in the sense of first-order stochastic dominance. We have  $\partial H_I/\partial y > 0$  if and only if

$$\frac{\int_{-\infty}^x f'_{(n-1)}(\tilde{\epsilon} + y)[1 - F(\tilde{\epsilon})]d\tilde{\epsilon}}{\int_{-\infty}^x f_{(n-1)}(\tilde{\epsilon} + y)[1 - F(\tilde{\epsilon})]d\tilde{\epsilon}} > \frac{\int_{-\infty}^{\infty} f'_{(n-1)}(\tilde{\epsilon} + y)[1 - F(\tilde{\epsilon})]d\tilde{\epsilon}}{\int_{-\infty}^{\infty} f_{(n-1)}(\tilde{\epsilon} + y)[1 - F(\tilde{\epsilon})]d\tilde{\epsilon}}.$$

We therefore have  $\partial H_I/\partial y > 0$  if

$$h_I(x) := \frac{\int_{-\infty}^x f'_{(n-1)}(\tilde{\epsilon} + y)[1 - F(\tilde{\epsilon})]d\tilde{\epsilon}}{\int_{-\infty}^x f_{(n-1)}(\tilde{\epsilon} + y)[1 - F(\tilde{\epsilon})]d\tilde{\epsilon}}$$

is decreasing in  $x$ . By taking a derivative, we have

$$\text{sign}[h'_I(x)] = \text{sign} \left\{ \int_{-\infty}^x [1 - F(\tilde{\epsilon})] \left[ \frac{f'_{(n-1)}(x + y)}{f_{(n-1)}(x + y)} - \frac{f'_{(n-1)}(\tilde{\epsilon} + y)}{f_{(n-1)}(\tilde{\epsilon} + y)} \right] d\tilde{\epsilon} \right\},$$

which is negative because  $f'_{(n-1)}/f_{(n-1)}$  is a decreasing function, we follows from the log-concavity of  $f$ .<sup>22</sup>

Putting these together, we have that  $1/X_I(y)$  is decreasing in  $y$ , implying that  $X_I(y)$  is increasing in  $y$ .

Next, we establish that  $X_O(y)$  is decreasing in  $y$ . To this end, let  $L_O$  be written in an alternative form:

$$L_O = \int_{-\infty}^{\infty} \frac{1 - F^{n-1}(\epsilon_I + y)}{n - 1} f(\epsilon_I) d\epsilon_I.$$

Then, we can write  $1/X_O$  as

$$\begin{aligned} & \frac{1}{X_O(y)} \\ &= \int_{-\infty}^{\infty} \left[ \frac{\int_{\epsilon_I+y}^{\infty} f(\epsilon) dF^{n-2}(\epsilon) + f(\epsilon_I + y)F^{n-2}(\epsilon_I + y)}{1 - F^{n-1}(\epsilon_I + y)} \right] \left[ \frac{1 - F^{n-1}(\epsilon_I + y)}{\int_{-\infty}^{\infty} [1 - F^{n-1}(\epsilon_I + y)] f(\epsilon_I) d\epsilon_I} \right] f(\epsilon_I) d\epsilon_I \end{aligned}$$

Teh, Liu, Wright and Zhou (2023) show that  $1/X_O(y)$  is increasing in  $y$ , implying that  $X_O(y)$  is decreasing in  $y$ .  $\square$

**Example A.4.** *Logit.* In the random utility specification of Example A.3 with  $F(v) = \exp[-\exp(-x/\theta)]$ ,  $L(v_{ji}, v_{j-i})$  is given by

$$L(v_{ji}, v_{j-i}) = \frac{\exp\left(\frac{v_{ji}}{\theta}\right)}{\sum_{k=1}^n \exp\left(\frac{v_{jk}}{\theta}\right)}.$$

In Example A.4,  $L_{ij}$  is given by (see Anderson et al. (1992) for the detail)

$$L_{ij} = \frac{\exp\left(\frac{w_{ji}}{\theta}\right)}{\sum_{k=1}^n \exp\left(\frac{w_{ki}}{\theta}\right)}.$$

Then,

$$\frac{\partial L_{ij}}{\partial w_{ji}} = \frac{1}{\theta} L_{ij} (1 - L_{ij}),$$

so

$$\begin{aligned} L_I(y) &= \frac{\exp\left(\frac{y}{\theta}\right)}{\exp\left(\frac{y}{\theta}\right) + n - 1}, & L_O(y) &= \frac{1}{n - 1} [1 - L_I(y)], \\ X_I(y) &= \frac{\theta}{1 - L_I(y)}, & X_O(y) &= \frac{\theta}{1 - L_O(y)}, \\ X'_I(y) &= \frac{L_I(y)}{1 - L_I(y)} > 0, & X'_O(y) &= L'_O(y) \frac{1}{[1 - L_O(y)]^2} < 0. \end{aligned}$$

---

<sup>22</sup>If  $f$  is log-concave, then  $F$  is also log-concave. Then, because

$$\log f_{(n-1)}(\epsilon) = \log(n - 1) + \log f(\epsilon) + (n - 2) \log F(\epsilon)$$

is also concave,  $f_{(n-1)}$  is also log-concave.

**Example A.5** (Linear city model with nonlinear transport cost). Consider an example where  $n = 2$ , and

$$\varepsilon_{ji} = \begin{cases} \bar{V} - d(x) & \text{if } i = 1, \\ \bar{V} - d(1-x) & \text{if } i = 2 \end{cases}$$

where  $\bar{V} > 0$  is sufficiently large constant, and  $x \sim U[0, 1]$ , where

$$d(y) = \begin{cases} g(y) & \text{if } y \in [0, \frac{1}{2} - e] \cup [\frac{1}{2} + e, 1], \\ m(y) = y + 2y^3 - 2y^4 & \text{if } y \in [\frac{1}{2} - e, \frac{1}{2} + e], \end{cases}$$

where  $g(y)$  is an increasing function such that  $g(1/2 - e) = m(1/2 - e)$ ,  $g'(1/2 - e) = m'(1/2 - e)$ ,  $g(1/2 + e) = m(1/2 + e)$ ,  $g'(1/2 + e) = m'(1/2 + e)$  and sufficiently convex to satisfy Assumption 1.

In the region where  $x \in [1/2 - e, 1/2 + e]$ , we have

$$\begin{aligned} d(x) - d(1-x) &= (2x-1) \{1 + 2[x^2 - x + 1 - (x^2 + (1-x)^2)]\} \\ &=: \Delta d(x). \end{aligned}$$

We have

$$\begin{aligned} \Delta d'(x) &= 2 \{1 + 2[x^2 - x + 1 - (x^2 + (1-x)^2)] - (2x-1)\}, \\ \Delta d''(x) &= -8(2x-1). \end{aligned}$$

In this setting, the worker chooses firm 1 if and only if  $x \leq \hat{x}$  such that

$$y = \Delta d(\hat{x}).$$

Hence, we have  $L_I = \hat{x}$ ,  $L_O = 1 - L_I = 1 - \hat{x}$ ,

$$\begin{aligned} \frac{\partial L_{ji}}{\partial w_{ji}} &= \frac{1}{\Delta d'(\hat{x})}, \\ X_I(y) &= \Delta d'(\hat{x})\hat{x}, \quad X_O(y) = \Delta d'(\hat{x})(1 - \hat{x}), \\ \frac{dX'_I(y)}{d\hat{x}} &= 1 + \hat{x}\Delta d''(\hat{x}), \quad \frac{dX'_O(y)}{d\hat{x}} = -1 + (1 - \hat{x})\Delta d''(\hat{x}). \end{aligned}$$

Concerning the proof of Lemma 2-(ii), note that the expression in equation 2 is proportional to

$$\begin{aligned} &2x - 1 + (1-x)X'_O + xX'_I \\ &= 2x - 1 + (1-x)(-1 + (1-\hat{x})\Delta d''(\hat{x})) + x(1 + \hat{x}\Delta d''(\hat{x})) \\ &= -6(2x-1). \end{aligned}$$

Hence, as long as  $L_I^* \simeq 1/2$  so that the above analysis applies,  $\widetilde{\text{IP}}$  can decrease with  $\sigma$  at  $\sigma \simeq 0$ .  $\square$

### A.2.2 Full characterizations of models used for proofs

**Equilibrium characterization of Example 1** Here, we provide the characterization of equilibrium in Example 1.

**Proposition A.1** (Equilibrium in Example 1). *In Example 1, the symmetric equilibrium satisfies*

$$\begin{aligned}\Delta w^* &= \frac{n\sigma}{2n-1}, \\ L_I^* &= \frac{1}{n} + \frac{(n-1)\sigma}{n\theta(2n-1)}.\end{aligned}$$

Furthermore, if  $\alpha_c = 1$  and  $\alpha_\tau = 2$ , we have the closed-form expressions for  $(p^*, h^*)$ :

$$\begin{aligned}p^* &= \frac{v + \left[1 + \frac{\gamma}{1-\gamma} \frac{n-1}{n}\right] \left[\bar{C} - \frac{\eta^2}{n\kappa} L_I^* v\right]}{1 + \left[1 + \frac{\gamma}{1-\gamma} \frac{n-1}{n}\right] \left[1 - \frac{\eta^2}{n\kappa} L_I^*\right]}, \\ h^* &= \frac{\eta L_I^* \left[1 + \frac{\gamma}{1-\gamma} \frac{n-1}{n}\right] (v - \bar{C})}{n\kappa \left[1 + \left[1 + \frac{\gamma}{1-\gamma} \frac{n-1}{n}\right] \left[1 - \frac{\eta^2}{n\kappa} L_I^*\right]\right]}.\end{aligned}$$

Equilibrium wage levels are given by

$$\begin{aligned}w_I^* &= \eta h^* Q^* - (\theta + \sigma - \Delta w^*), \\ w_O^* &= \eta h^* Q^* - \left(\theta - \frac{\sigma - \Delta w^*}{n-1}\right).\end{aligned}$$

Proposition A.5 in Online Appendix A.4 shows the same result in the extended version of the model with overlapping ownership. Proposition A.1 follows from Proposition A.5 by assuming overlapping ownership away (i.e., setting  $\lambda = 0$  in the notation of Proposition A.5). Hence, we relegate the proof of Proposition A.1 to Online Appendix A.4.

**Models in Table 3:** Here, we present the full characterizations of equilibrium in the models shown in Table 3, which are used for the proofs of Lemma 3. Specifically, we consider two cases where  $\alpha = 1/2$  and  $\delta = 1$  and where  $\alpha = 1$  and  $\delta = 2$  as shown in Table 4.

*Preliminary:* When

$$D_i(\mathbf{p}) = A \exp(1 - p_i),$$

the profit-maximizing price given marginal cost  $c_i = c(H_i)$  is

$$p(c_i) = c_i + 1.$$

The product-market profit given the marginal cost  $c_i$  is then

$$A \exp(-c_i).$$

If the marginal cost is given by

$$c(H) = \log \bar{C} - \alpha \log H,$$

where  $\alpha \leq 1$ , the product-market profit is written as

$$\pi(H_i) = \frac{AH_i^\alpha}{\bar{C}}.$$

Therefore, firm  $i$ 's reduced profit function is given by

$$\pi_i^L(w_{ij}, w_{ji}, h_i) = \pi(H_i) - L_{ii}w_{ii} - L_{ji}w_{ij} - \tau(h_i).$$

In the following, we derive the condition under which the characterization of symmetric equilibrium given by Proposition 1 is sufficient. To this end, we examine the condition under which the second-order condition for the best-response of each firm is indeed satisfied at the symmetric equilibrium. Particularly, we consider the following three problems sequentially:

1. Fixed  $w_{ii}$  and  $h_i$ , choose  $w_{ij}$  to maximize the profit  $\pi_i^L$  of firm  $i$ .
2. Fixed  $w_{ii}$ , choose  $h_i$  to maximize the profit  $\pi_i^L$  of firm  $i$ , taking into account that  $w_{ij}$  is optimally chosen later.
3. Choose  $w_{ii}$  to maximize the profit  $\pi_i^L$  of firm  $i$ , taking into account that  $h_i$  and  $w_{ij}$  will be optimally chosen later.

*Case where  $\alpha = 1/2$  and  $\delta = 1$ :* First, consider the choice of  $w_{ij}$  given  $h_i$  and  $w_{ii}$ . The first-order condition is given by

$$\frac{\partial \pi_i^L}{\partial w_{ij}} = \frac{1}{2\theta} \left( \alpha \frac{Ah_j}{\bar{C}} H_i^{\alpha-1} - w_{ij} \right) - L_{ji} = 0.$$

The second-order condition

$$\frac{\partial^2 \pi_i^L}{\partial w_{ij}^2} = (\alpha - 1)\alpha \frac{Ah_j^2 H_i^{\alpha-2}}{4\theta^2 \bar{C}} - \frac{1}{2\theta} - \frac{1}{2\theta} < 0,$$

is always satisfied. Furthermore, note that  $L_{ij}$  is decreasing in  $h_i L_{ii}$  but  $H_i$  is increasing in  $h_i L_{ii}$

Next, consider the choice of  $h_i$  given  $w_{ii}$ . Solving for the first-order condition, we have

$$\frac{\partial}{\partial h_i} \left( \pi_i^L \Big|_{\frac{\partial \pi_i^L}{\partial w_{ij}}=0} \right) = \frac{A}{\bar{C}} \alpha L_{ii} H_i^{\alpha-1} - \kappa = 0.$$

The second-order condition

$$\frac{\partial^2}{\partial h_i^2} \left( \pi_i^L \Big|_{\frac{\partial \pi_i^L}{\partial w_{ij}}=0} \right) = (\alpha - 1)\alpha \frac{AL_{ii}}{\bar{C}} \frac{dH_i}{dh_i} H_i^{\alpha-2} < 0,$$

is always satisfied.

To derive the optimal choice of  $h_i$ , we reformulate the solutions  $L_{ji}$  and  $h_i$  in the following way. We solve for  $H_i$ ,  $h_i$ , and  $L_{ji}$  given  $L_{ii}$ , which is given by the system of equations

$$\begin{aligned} H_i &= \left( \frac{A}{\kappa \bar{C}} \alpha L_{ii} \right)^{\frac{1}{1-\alpha}} \\ L_{ji} &= \frac{1}{4} - \frac{w_{jj} + \sigma}{4\theta} + \frac{h_j \kappa}{2L_{ii}} \\ h_i &= \frac{H_i - L_{ji} h_j}{L_{ii}} \\ &= \left( \frac{A}{\kappa \bar{C}} \alpha \right)^{\frac{1}{1-\alpha}} L_{ii}^{\frac{\alpha}{1-\alpha}} - \frac{h_j}{L_{ii}} \left( \frac{1}{4} - \frac{w_{jj} + \sigma}{4\theta} + \frac{h_j \kappa}{2L_{ii}} \right) \end{aligned}$$

Finally, we derive the optimal choice of  $w_{ii}$ . The first-order condition is given by

$$\frac{\partial}{\partial w_{ii}} \left( \pi_i^L \Big|_{\frac{\partial \pi_i^L}{\partial w_{ij}}=0, \frac{\partial \pi_i^L}{\partial h_i}=0} \right) = \frac{1}{2\theta} \left( \frac{A h_i}{\bar{C}} \alpha H_i^{\alpha-1} - w_{ij} \right) - L_{ii} = 0$$

In terms of  $L_{ii}$ , this can be written as

$$\Upsilon(L_{ii}) := \frac{1}{2\theta} \left[ \left( \frac{A}{\kappa \bar{C}} \alpha \right)^{\frac{1}{1-\alpha}} L_{ii}^{\frac{-1+2\alpha}{1-\alpha}} - \frac{h_j}{L_{ii}^2} \left( \frac{1}{4} - \frac{w_{jj} + \sigma}{4\theta} + \frac{h_j \kappa}{2L_{ii}} \right) \right] - 2L_{ii} + \text{constant}$$

Note that this function is concave in  $L_{ii}$ , so it equals 0 at most twice, and the largest solution satisfies  $\Upsilon'(L_{ii}) < 0$  and is the profit-maximizing solution.

When  $\alpha = 1/2$ , we have

$$\Upsilon(L_{ii}) = \frac{1}{2\theta} \left[ \left( \frac{A}{2\kappa \bar{C}} \right)^2 - \frac{h_j}{L_{ii}^2} \left( \frac{1}{4} - \frac{w_{jj} + \sigma}{4\theta} + \frac{h_j \kappa}{2L_{ii}} \right) \right] - 2L_{ii} + \text{constant}$$

At the symmetric equilibrium  $L_{ii} = L_I$ , we have

$$\begin{aligned} \Upsilon'(L_I) &= \frac{h(1-L_I)}{\theta L_I^3} + \frac{h^2 \kappa}{4\theta L_I^4} - 2 \\ &= \left( \frac{A}{2\kappa \bar{C}} \right)^2 \frac{1-L_I}{\theta L_I} + \left( \frac{A}{2\kappa \bar{C}} \right)^4 \frac{\kappa}{\theta} - 2 \end{aligned}$$

When  $\sigma = 0$  so that  $L_I = 1/2$ , we have

$$\Upsilon'(L_I) = \frac{A^2}{\theta \kappa^2 \bar{C}^2} \left( \frac{1}{4} + \frac{A^2}{16\kappa \bar{C}^2} \right) - 2,$$

which is negative if and only if

$$8\theta \kappa > \frac{A^2}{\kappa \bar{C}^2} + \frac{A^4}{4\bar{C}^4 \kappa^2},$$

which is the case whenever  $\kappa$  is sufficiently large.

Case where  $\alpha = 1$  and  $\delta = 2$ : In this case, firm  $i$ 's optimal choice of  $w_{ij}$  given  $h_i$  and  $w_{ii}$  is

$$\frac{\partial \pi_i^L}{\partial w_{ij}} = \frac{1}{2\theta} \left( \frac{Ah_j}{\bar{C}} - w_{ij} \right) - L_{ji} = 0.$$

The second-order condition

$$\frac{\partial^2 \pi_i^L}{\partial w_{ij}^2} = -\frac{1}{2\theta} - \frac{1}{2\theta} < 0,$$

is always satisfied. Next, consider the choice of training level given  $L_{ii}$ , which is given by the first-order condition

$$\frac{\partial}{\partial h_i} \left( \pi_i^L \Big|_{\frac{\partial \pi_i^L}{\partial w_{ij}}=0} \right) = \frac{A}{\bar{C}} L_{ii} - \kappa h_i = 0. \Rightarrow h_i = \frac{A}{\kappa \bar{C}} L_{ii}$$

The second-order condition

$$\frac{\partial^2}{\partial h_i^2} \left( \pi_i^L \Big|_{\frac{\partial \pi_i^L}{\partial w_{ij}}=0} \right) = -\kappa < 0$$

is always satisfied. Finally, consider the choice of  $w_{ii}$ , which is given by the first-order condition

$$\frac{\partial}{\partial w_{ii}} \left( \pi_i^L \Big|_{\frac{\partial \pi_i^L}{\partial w_{ij}}=0, \frac{\partial \pi_i^L}{\partial h_i}=0} \right) = \frac{1}{2\theta} \left( \frac{Ah_i}{\bar{C}} - w_{ij} \right) - L_{ii} = 0.$$

The second-order condition

$$\frac{\partial^2}{\partial w_{ii}^2} \left( \pi_i^L \Big|_{\frac{\partial \pi_i^L}{\partial w_{ij}}=0, \frac{\partial \pi_i^L}{\partial h_i}=0} \right) = \frac{1}{4\theta^2} \frac{A^2}{\kappa \bar{C}^2} - \frac{1}{2\theta} - \frac{1}{2\theta} < 0$$

is satisfied if and only if

$$4\theta\kappa > \frac{A^2}{\bar{C}^2},$$

which holds when  $\kappa$  is sufficiently large.

### A.2.3 Under- versus overinvestment

In this section, we propose the following assumption and show that under this assumption, the local derivative  $\bar{\Pi}'(h^*)$  is sufficient for determining whether training level is overinvestment or underinvestment in the strict sense that  $h^*$  is above or below the level that maximizes  $\bar{\Pi}(h)$ .

**Assumption A.1.** *The pass-through rate times aggregate diversion ratio,  $\rho(h)\psi(\tilde{p}(h))$ , is increasing in  $h$ .*

**Proposition A.2.** *Let  $h^\dagger = \arg \max_h \bar{\Pi}(h)$ . Under Assumption A.1,  $h^\dagger < h^*$  if and only if  $\bar{\Pi}'(h^*) > 0$ .*

*Proof.* Suppose  $\bar{\Pi}'(h^*) > 0$  holds. We show that  $h^\dagger < h^*$ .

First, note that for any  $h$ , we have

$$\begin{aligned}\bar{\Pi}(h) &= \tilde{p}'(h)\psi(\tilde{p}(h))Q(\tilde{p}(h)) - c'(h)Q(\tilde{p}(h)) - \tau'(h) \\ &= -c'(h) \left\{ Q(\tilde{p}(h)) [1 - \rho(h)\psi(\tilde{p}(h))] - \frac{\tau'(h)}{-c'(h)} \right\},\end{aligned}$$

which is positive if and only if

$$\Delta\bar{\Pi}(h) := Q(\tilde{p}(h)) [1 - \rho(h)\psi(\tilde{p}(h))] - \frac{\tau'(h)}{-c'(h)} > 0.$$

Now consider the value of  $\Delta\bar{\Pi}(h)$  for  $h \leq h^*$ . Noting that  $h \leq h^*$  implies

$$Q(\tilde{p}(h))L_I^* - \frac{\tau'(h)}{-c'(h)} \geq 0,$$

we have

$$\begin{aligned}\Delta\bar{\Pi}(h) &\geq Q(\tilde{p}(h)) [1 - L_I^* - \rho(h)\psi(\tilde{p})] \\ &> Q(\tilde{p}(h)) [\rho(h^*)\psi(p^*) - \rho(h)\psi(\tilde{h})] \\ &\geq 0,\end{aligned}$$

where the first inequality holds from  $h \leq h^*$ , the second holds from  $\bar{\Pi}'(h^*) > 0$ , and the third from Assumption A.1. Hence, when  $\bar{\Pi}'(h^*) > 0$ ,  $\Delta\bar{\Pi}(h) > 0$  for all  $h \leq h^*$ , implying that  $h^\dagger > h^*$ .

Similarly, we can show that when  $\bar{\Pi}'(h^*) < 0$ ,  $h^\dagger < h^*$ . □

The next question is whether Assumption A.1 is restrictive or not. It turns out that Assumption A.1 can be characterized only in terms of the demand characteristics.

**Proposition A.3.** *Assumption A.1 holds if and only if the function*

$$\Psi(p) := \frac{\psi(p)}{2 - \psi(p) + \frac{1}{\varepsilon_Q(p)} \left[ \alpha_Q(p) \frac{1-\psi(p)}{\psi(p)} + \varepsilon_\psi(p) \right]}$$

is increasing in  $p$ ,

*Proof.* Remember that we have

$$\begin{aligned}\tilde{p}'(h) &= \frac{Q_p c'}{\frac{Q'_p Q}{Q_p} + Q_p + Q'} \\ &= \frac{c'}{2 - \psi(p) + \frac{1}{\varepsilon_Q(p)} \left[ \alpha_Q(p) \frac{1-\psi(p)}{\psi(p)} + \varepsilon_\psi(p) \right]} \\ &< 0,\end{aligned}$$

where

$$\begin{aligned}\varepsilon_\psi(p) &:= \frac{\psi'(p)p}{\psi(p)}, \\ \alpha_Q(p) &:= \frac{Q''(p)p}{-Q'(p)}.\end{aligned}$$

Hence, we can write  $\rho(h)$  as

$$\rho(h) = \frac{1}{2 - \psi(p) + \frac{1}{\varepsilon_Q(p)} \left[ \alpha_Q(p) \frac{1-\psi(p)}{\psi(p)} + \varepsilon_\psi(p) \right]},$$

and thus

$$\rho(h)\psi(\tilde{p}(h)) = \Psi(\tilde{p}(h)),$$

Noting that  $\tilde{p}(h)$  is a decreasing function,  $\rho(h)\psi(\tilde{p}(h))$  is decreasing in  $h$  if and only if  $\Psi(p)$  is increasing in  $p$ , which completes the proof.  $\square$

We note that  $\Psi(p)$  is increasing in  $p$  under various specifications.

- Full coverage: when  $\psi(p) = 1$  for all  $p$ , we have  $\Psi(p) = 1$  for all  $p$ , so Assumption A.1 always hold.
- Independent goods: when  $\psi(p) = 0$  for all  $p$ , we have  $\Psi(p) = 0$  for all  $p$ , so Assumption A.1 always hold.
- Shubik-Levitan: under Shubik-Levitan specification,  $\Psi(p)$  is constant, as shown in Proposition 3. Hence, Assumption A.1 always hold.
- Logit: Under logit demand specification, one can show that  $\Psi(p)$  is increasing in  $p$ .

#### A.2.4 Condition for switching cost to lower equilibrium NMU

In the proof of Lemma 2, we showed that

$$\frac{\partial \widetilde{\text{TS}}}{\partial \sigma} = -(1 - L_I) + \frac{1}{1 + \Delta X'} L'_I (\sigma - \Delta w) [X_I - X_O].$$

Here, we apply this characterization to [Hotelling \(1929\)](#) and Logit examples to show the specific values of  $\tilde{\sigma}$ .

**Hotelling specification of Example A.1** In Hotelling setting of Example A.1, we have

$$\begin{aligned}\Delta X' &= 2, \\ L'_I &= \frac{1}{2\theta}, \\ X_I(y) - X_O(y) &= 2y = 2\theta(2L_I - 1)\end{aligned}$$

Hence,

$$\begin{aligned}
\frac{\partial \widetilde{\text{TS}}}{\partial \sigma} &= -(1 - L_I) + \frac{1}{1 + \Delta X'} L'_I(\sigma - \Delta w) [X_I - X_O] \\
&= -(1 - L_I) + \frac{1}{3}(2L_I - 1) \\
&= \frac{5}{3}L_I - \frac{4}{3},
\end{aligned}$$

which is positive if and only if  $L_I^* > 4/5$ .

**Logit specification of Example A.4** Under the logit specification of Example A.4, we have

$$\begin{aligned}
L'_I(y) &= \frac{1}{\theta} L_I(y) [1 - L_I(y)], \\
X_I(y) - X_O(y) &= \frac{\theta}{1 - L_I(y)} - \frac{\theta}{1 - L_O(y)}, \\
X'_I(y) &= \frac{L_I(y)}{1 - L_I(y)}, \\
X'_O(y) &= -\frac{1}{n - 1} \frac{L_I(y) [1 - L_I(y)]}{[1 - L_O(y)]^2}.
\end{aligned}$$

Hence, we obtain

$$\begin{aligned}
\frac{\partial \widetilde{\text{TS}}}{\partial \sigma} &= -(1 - L_I) + \frac{1}{1 + \Delta X'} L'_I(\sigma - \Delta w) [X_I - X_O] \\
&= -(1 - L_I) + \frac{1}{1 + \Delta X'} L_I(1 - L_I) \left( \frac{1}{1 - L_I} - \frac{1}{1 - L_O} \right) \\
&= \frac{1 - L_I}{1 + \Delta X'} \left( -1 - \Delta X' + \frac{L_I}{1 - L_I} - \frac{L_I}{1 - L_O} \right) \\
&= \frac{1 - L_I}{1 + \Delta X'} \left( -1 + X'_O - \frac{L_I}{1 - L_O} \right) \\
&< 0,
\end{aligned}$$

establishing that  $\widetilde{\text{TS}}$  is always decreasing in  $\sigma$  under logit specification and thus  $\tilde{\sigma} = \infty$ .

### A.3 Segmented labor markets

In this section, we consider the case where firms are segmented into different labor markets so that they face the product market competition with all rivals but labor market competition with less rivals.

**Setting** We consider the setting where there are  $M$  labor markets served by  $N$  firms, so the total number of firms is  $n = MN$ . Let  $m = 1, 2, \dots, M$  be a market and  $\mathcal{N}_m$  be the set of firms in labor market  $m$ .

In a market  $m$ , A worker initially hired by firm  $i \in \mathcal{N}_m$  can choose whether to stay at firm  $i$  or to move to firm  $j \in \mathcal{N}_m \setminus \{i\}$ . Hence, the worker choice probability function is given by

$$L_{ij}(v_i) = \Pr(v_{ij} + \varepsilon_{ij} \geq v_{ik} + \varepsilon_{ik} \text{ for all } k \in \mathcal{N}_m), \quad (\text{A.3})$$

where  $v_i = (v_{ik})_{k \in \mathcal{N}_m}$ . Each firm  $i \in \mathcal{N}_m$  sets the retention wage  $w_{ii}$  and poaching wages  $(w_{ij})_{j \in \mathcal{N}_m \setminus \{i\}}$ . Otherwise, all the setting is the same as the baseline model.

This extended model nests the original setting when  $M = 1$ . For a fixed  $n$ , as  $M$  increases, labor market becomes more segmented, while keeping the total number of firms fixed.

In this extended model, only the definition of  $X_I(y)$  and  $X_O(y)$  change from the baseline model of Section 3. In particular, the definitions of  $X_I(y)$  and  $X_O(y)$  are given by

$$X_I(y) := \frac{L_{ii}}{\partial L_{ii} / \partial w_{ii}} \Big|_{v_{ii} - v_{ik} = y \text{ for all } k \in \mathcal{N}_m \setminus \{i\}}, \quad (\text{A.4})$$

$$X_O(y) := \frac{L_{ji}}{\partial L_{ji} / \partial w_{ij}} \Big|_{v_{ij} - v_{ii} = -y, \text{ and } v_{ij} - v_{ik} = 0 \text{ for all } k \in \mathcal{N}_m \setminus \{i\}}, \quad (\text{A.5})$$

for any  $i, j \in \mathcal{N}_m$ . These definitions coincides with that of the baseline model when  $\mathcal{N}_m = \mathcal{N}$ .

**Analysis** The equilibrium conditions are the same as the baseline model except for the definitions of  $L_{ij}(v_i)$  in equation (A.3) and  $X_I(y)$  and  $X_O(y)$  in equations (A.4) and (A.5). Below, we discuss how the change in  $M$  (holding  $n = NM$  fixed) affects the equilibrium values of  $X_I(y^*)$ ,  $X_O(y^*)$ ,  $L_I^*$ , and other equilibrium variables.

Holding the total number of firms  $n = NM$  fixed, an increase in the number of labor markets effectively reduce the number of labor-market competitors in each labor market. To yield clear-cut results, we use the parametric specification of Example 1, where  $L_{ij}$  is given by the spokes form of [Chen and Riordan \(2007\)](#). Using Proposition A.1, we can show that

$$\begin{aligned} \Delta w^* &= \frac{N\sigma}{2N-1}, \\ L_I^* &= \frac{1}{N} + \frac{(N-1)\sigma}{N\theta(2N-1)}, \\ X_I(y^*) &= \theta + \sigma - \Delta w^*, \\ X_O(y^*) &= \theta - \frac{\sigma - \Delta w^*}{N-1} \end{aligned}$$

Then, we have

$$\begin{aligned}\frac{d \Delta w^*}{d N} &= \frac{-\sigma}{(2N-1)^2} < 0, \\ \frac{d L_I^*}{d N} &= -\frac{1}{N^2} + \frac{N(2N-1) - (N-1)(4N-1)}{\theta N^2 (N-1)^2} \sigma \\ &= -\frac{1}{N^2} - \frac{(2N^2 - 4N + 1)\sigma}{\theta N^2 (2N-1)^2} \\ &< 0, \text{ for any } N \geq 2.\end{aligned}$$

Hence,  $\Delta w^*$  and  $L_I^*$  are decreasing in  $N$ . Furthermore, under [Chen and Riordan \(2007\)](#) specification, we have  $d X_I(y^*)/d N = -d \Delta w^*/d N > 0$  and  $d X_O(y^*)/d N = d X_I(y^*)/d N - d \Delta w^*/d N > 0$ , so the wage markdown increases with the number of local labor market competitors  $N$ .

Putting the above discussion together, we obtain the following proposition

**Proposition A.4.** *Under the parametric specification of Example 1, suppose that the value of  $M$  increases while holding  $n = NM$  fixed (so that the value of  $N$  decreases). Then:*

1. *increases  $L_I^*$ ,  $h^*$ , and  $Q^*$ ;*
2. *increases  $w_I^*$  and  $w_O^*$ ; and*
3. *increases  $CS^*$  and  $WS^*$  and has ambiguous effects on  $IP^*$ .*

In light of Proposition A.4, labor market segregation (i.e., an increase in  $M$ ) has features similar to labor switching costs  $\sigma$ . Specifically, for a worker initially hired at firm  $j \in \mathcal{N}_m$ , the cost of switching to firm  $j \notin \mathcal{N}_m$  can be viewed as prohibitively large. From this perspective, labor market segmentation effectively serves as one way of increasing switching costs. Thus, it is natural that the comparative statics with respect to  $\sigma$  and  $M$  have some similarities.

Proposition A.4 shows that labor market segregation lowers labor mobility ( $1 - L_I^*$ ). Hence, together with Proposition 3, we find that, under the parametric specification of Example 1, the equilibrium training level tends to be overinvestment when labor markets are segmented. Then, in view of Proposition 4, the clear-cut welfare improvement arising from switching costs is less likely when labor markets are segmented.

## A.4 Model with overlapping ownership

### A.4.1 Setting

Consider an extension of the model in which the objective function of each firm  $i \in \mathcal{N}$  is given by the weighted sum of firms' profits:

$$\Omega_i = \pi_i + \lambda \sum_{j \neq i} \pi_j,$$

where the weight parameter  $\lambda \in [0, 1)$  is often called the Edgeworth coefficient of effective sympathy among firms that arises from overlapping ownership (see [López and Vives, 2019](#);

Vives and Vravosinos, forthcoming). We consider the model that is identical to the one presented in Section 3 except that each firm  $i$ 's objective function is  $\Omega_i$  instead of  $\pi_i$ , and that we let  $\sigma = 0$  in order to focus on the effects of overall market power associated with  $\lambda$ .

#### A.4.2 Equilibrium analysis

The first-order conditions for the equilibrium prices, wages, and training are given by

$$\frac{\partial \Omega_i}{\partial p_i} = D_i + \frac{\partial D_i}{\partial p_i} [p_i - c(H_i)] + \lambda \sum_{j \in \mathcal{N} \setminus \{i\}} \frac{\partial D_j}{\partial p_i} [p_j - c(H_j)] = 0,$$

$$\frac{\partial \Omega_i}{\partial w_{ik}} = L_{ki} + \left[ \frac{\partial L_{ki}}{\partial w_{ik}} [-c'(H_i) D_i h_k - w_{ki}] + \lambda \sum_{j \in \mathcal{N} \setminus \{i\}} \frac{\partial L_{kj}}{\partial w_{ik}} [-c'(H_j) D_j h_k - w_{kj}] \right] = 0, \quad (\text{A.6})$$

and

$$\frac{\partial \Omega_i}{\partial h_i} = -c'(H_i) D_i L_{ii} - \lambda \sum_{j \in \mathcal{N} \setminus \{i\}} c'(H_j) D_j L_{ij} - \tau'(h_i) = 0$$

Using the same notation as that in Section 4, at the symmetric equilibrium, we have  $p_i = p$ ,  $D_i = Q(p)$ ,  $L_{ii} = L_I$ ,  $L_{ij} = L_O$ ,  $h_i = h$  for all  $i$  and  $j \neq i$ . Furthermore,  $w_{ii} = w_I$  and  $w_{ji} = w_O$  for all  $i$  and  $j \neq i$ . Let  $\Delta w := w_O - w_I$ .

Here, we use the Shubik-Levitan and Chen-Riordan specification of Example 1 to provide a sharper equilibrium characterization.

**Proposition A.5** (Equilibrium in Example 1 with overlapping ownership). *In Example 1 with overlapping ownership, the symmetric equilibrium satisfies*

$$\Delta w^*(\sigma, \lambda) = \frac{n\sigma}{2n - 1 + \lambda},$$

$$L_I^*(\sigma, \lambda) = \frac{1}{n} + \frac{(n - 1 + \lambda)\sigma}{n\theta(2n - 1 + \lambda)}.$$

Furthermore, if  $\alpha_c = 1$  and  $\alpha_\tau = 2$ , we have the closed-form expressions for  $(p^*, h^*)$ :

$$p^* = \frac{v + \left[ 1 + (1 - \lambda) \frac{\gamma}{1 - \gamma} \frac{n-1}{n} \right] \left\{ \bar{C} - \frac{\eta^2}{n\kappa} [L_I^* + \lambda(1 - L_I^*)] v \right\}}{1 + \left[ 1 + (1 - \lambda) \frac{\gamma}{1 - \gamma} \frac{n-1}{n} \right] \left\{ 1 - \frac{\eta^2}{n\kappa} [L_I^* + \lambda(1 - L_I^*)] \right\}},$$

$$h^* = \frac{\eta [L_I^* + \lambda(1 - L_I^*)]}{n\kappa} \frac{\left[ 1 + (1 - \lambda) \frac{\gamma}{1 - \gamma} \frac{n-1}{n} \right] (v - \bar{C})}{1 + \left[ 1 + (1 - \lambda) \frac{\gamma}{1 - \gamma} \frac{n-1}{n} \right] \left\{ 1 - \frac{\eta^2}{n\kappa} [L_I^* + \lambda(1 - L_I^*)] \right\}}$$

Equilibrium wage levels are given by

$$w_I^* = \eta h^* Q^* - \frac{\theta + \sigma - (1 + \lambda) \Delta w^*}{1 - \lambda},$$

$$w_O^* = \eta h^* Q^* - \frac{\theta - \frac{\sigma - (1 + \lambda) \Delta w^*}{n-1}}{1 - \lambda}.$$

*Proof of Proposition A.5.* As in the baseline model, we first consider from the equilibrium retention rate. Under [Chen and Riordan \(2007\)](#) specification, the conditions for the equilibrium wages can be written as

$$\begin{aligned} -L_I + \frac{-(1-\lambda)c'(h)Q(p)h - w_I + \lambda w_O}{n\theta} &= 0, \\ -L_O + \frac{-(1-\lambda)c'(h)Q(p)h - w_O + \frac{\lambda}{n-1}[(n-2)w_O + w_I]}{n\theta} &= 0, \end{aligned}$$

which combine to

$$n(\sigma - \Delta w) - (n-1 + \lambda)\Delta w = 0,$$

so we obtain the equilibrium wage difference  $\Delta w^*$  and retention rate  $L_I^*$ :

$$\begin{aligned} \Delta w^*(\sigma, \lambda) &= \frac{n\sigma}{2n-1 + \lambda}, \\ L_I^*(\sigma, \lambda) &= \frac{1}{n} + \frac{(n-1 + \lambda)\sigma}{n\theta(2n-1 + \lambda)}. \end{aligned}$$

Next, given the equilibrium retention rate  $L_I^*$ , we consider the equilibrium prices and training. The symmetric equilibrium condition under overlapping ownership is given by

$$\begin{aligned} f_p(p, h) &= Q(p) + [(1-\lambda)Q_p(p) + \lambda Q'(p)] [p - c(h)], \\ f_h(p, h) &= -c'(h)[L_I^* + \lambda(1 - L_I^*)]Q(p) - \tau'(h) = 0. \end{aligned}$$

Under [Shubik and Levitan \(2013\)](#) specification. We have

$$\begin{aligned} Q(p) &= \frac{v-p}{n}, \\ Q'(p) &= -\frac{1}{n}, \\ Q_p(p) &= -\frac{1}{n} \left[ 1 + \frac{\gamma}{1-\gamma} \frac{n-1}{n} \right]. \end{aligned}$$

We can rewrite  $f_p$  and  $f_h$  as

$$\begin{aligned} f_p(p, h, L_I^*, \lambda) &= v - p - \left\{ (1-\lambda) \left[ 1 + \frac{\gamma}{1-\gamma} \frac{n-1}{n} \right] + \lambda \right\} (p - \bar{C} + \eta h), \\ f_h(p, h, L_I^*, \lambda) &= \eta[L_I^* + \lambda(1 - L_I^*)](v - p) - \kappa h. \end{aligned}$$

Let  $\tilde{h}(p, L_I, \lambda)$  be the value of  $h$  such that  $f_h(p, h, L_I, \lambda) = 0$ , that is,

$$\tilde{h}(p, L_I^*, \lambda) = \frac{\eta[L_I^* + \lambda(1 - L_I^*)]}{\kappa} \frac{v - p}{n}.$$

Plugging this into  $f_p = 0$ , we have the equilibrium condition for  $p$ :

$$\begin{aligned} \tilde{h}_p(p, L_I^*, \lambda) &= v - p - \left\{ (1-\lambda) \left[ 1 + \frac{\gamma}{1-\gamma} \frac{n-1}{n} \right] + \lambda \right\} \left\{ p - \bar{C} + \frac{\eta^2[L_I^* + \lambda(1 - L_I^*)]}{\kappa} \frac{v - p}{n} \right\} \\ &= 0. \end{aligned}$$

Solving this, we obtain the equilibrium value of  $(p^*, Q^*, h^*)$ :

$$p^*(\sigma, \lambda) = \frac{v + \left[1 + (1 - \lambda) \frac{\gamma}{1 - \gamma} \frac{n-1}{n}\right] \left\{ \bar{C} - \frac{\eta^2}{n\kappa} [L_I^* + \lambda(1 - L_I^*)] v \right\}}{1 + \left[1 + (1 - \lambda) \frac{\gamma}{1 - \gamma} \frac{n-1}{n}\right] \left\{ 1 - \frac{\eta^2}{n\kappa} [L_I^* + \lambda(1 - L_I^*)] \right\}},$$

$$Q^*(\sigma, \lambda) = \frac{v - p^*(\sigma, \lambda)}{n} = \frac{1}{n} \frac{\left[1 + (1 - \lambda) \frac{\gamma}{1 - \gamma} \frac{n-1}{n}\right] (v - \bar{C})}{1 + \left[1 + (1 - \lambda) \frac{\gamma}{1 - \gamma} \frac{n-1}{n}\right] \left\{ 1 - \frac{\eta^2}{n\kappa} [L_I^* + \lambda(1 - L_I^*)] \right\}},$$

$$h^*(\sigma, \lambda) = \frac{\eta[L_I^* + \lambda(1 - L_I^*)]}{n\kappa} \frac{\left[1 + (1 - \lambda) \frac{\gamma}{1 - \gamma} \frac{n-1}{n}\right] (v - \bar{C})}{1 + \left[1 + (1 - \lambda) \frac{\gamma}{1 - \gamma} \frac{n-1}{n}\right] \left\{ 1 - \frac{\eta^2}{n\kappa} [L_I^* + \lambda(1 - L_I^*)] \right\}}$$

Finally, consider the equilibrium wages. Rearranging equation (A.6), we obtain the equilibrium wages

$$\begin{aligned} w_I^* &= \eta h^* Q^* - \frac{n\theta L_I^* - \lambda \Delta w^*}{1 - \lambda} \\ &= \eta h^* Q^* - \frac{\theta + \sigma - (1 + \lambda) \Delta w^*}{1 - \lambda}, \\ w_O^* &= \eta h^* Q^* - \frac{n\theta L_O^* + \frac{\lambda}{n-1} \Delta w^*}{1 - \lambda} \\ &= \eta h^* Q^* - \frac{\theta - \frac{\sigma - (1 + \lambda) \Delta w^*}{n-1}}{1 - \lambda} \end{aligned}$$

□

#### A.4.3 Effects of overlapping ownership

**No switching costs** To understand the effects of overlapping ownership  $\lambda$  in the simplest manner, we start from the case where  $\sigma = 0$ . In this case, we have  $\Delta w^* = 0$  and  $L_I^* = 1/n$ , implying that

$$\begin{aligned} w_I^* = w_O^* = w^* &= \eta h^* Q^* - \frac{\theta}{1 - \lambda} \\ &= \frac{\eta^2}{\kappa} \frac{1 + \lambda(n-1)}{n} [Q^*(0, \lambda)]^2 - \frac{\theta}{1 - \lambda} \end{aligned}$$

We have

$$\frac{dQ^*(0, \lambda)}{d\lambda} = \frac{(n-1)(v - \bar{C})}{n^2} \frac{-\frac{\gamma}{1-\gamma} + \frac{\eta^2}{n\kappa} \left[1 + (1 - \lambda) \frac{\gamma}{1 - \gamma} \frac{n-1}{n}\right]^2}{\left\{ 1 + \left[1 + (1 - \lambda) \frac{\gamma}{1 - \gamma} \frac{n-1}{n}\right] \left[1 - \frac{\eta^2}{n\kappa} \frac{1 + \lambda(n-1)}{n}\right] \right\}^2}$$

which is positive if and only if

$$-\frac{\gamma}{1 - \gamma} + \frac{\eta^2}{n\kappa} \left[1 + (1 - \lambda) \frac{\gamma}{1 - \gamma} \frac{n-1}{n}\right]^2 > 0.$$

Because the above expression is monotonically decreasing in  $\lambda$ , there exists a critical value  $\hat{\lambda}$  such that  $dQ(0, \lambda)/d\lambda > 0$  if and only if  $\lambda < \hat{\lambda}$ .

Now consider the effect of  $\lambda$  on  $w^*$ , which is given by

$$\frac{dw^*(\lambda)}{d\lambda} = \frac{\eta^2}{\kappa} \left[ \frac{n-1}{n} [Q^*(0, \lambda)]^2 + 2 \frac{1 + \lambda(n-1)}{n} \frac{dQ^*(0, \lambda)}{d\lambda} Q^*(0, \lambda) \right] - \frac{\theta}{(1-\lambda)^2}$$

Noting that  $Q^*(\sigma, \hat{\lambda})$  is independent of  $\theta$ , we have  $dw^*/d\lambda > 0$  if and only if

$$\theta < \hat{\theta} := (1-\lambda)^2 \frac{\eta^2}{\kappa} \left[ \frac{n-1}{n} [Q^*(0, \lambda)]^2 + 2 \frac{1 + \lambda(n-1)}{n} \frac{dQ^*(0, \lambda)}{d\lambda} Q^*(0, \lambda) \right]$$

To summarize, we have

**Proposition A.6.** *Suppose that  $\sigma = 0$  and consider an increase in  $\lambda$ .*

1. *There exists a critical value  $\hat{\lambda}$  such that consumer surplus increases with the increase in  $\lambda$  if and only if  $\lambda < \hat{\lambda}$ .*
2. *From any starting value  $\lambda$ , there exists a critical value  $\hat{\theta}(\lambda)$  such that an increase in  $\lambda$  increases worker surplus if and only if  $\theta < \hat{\theta}(\lambda)$ .*

Hence, there are four possibilities regarding the effects of overlapping ownership  $\lambda$  on consumers and workers:

1.  $\lambda < \hat{\lambda}$  and  $\theta < \hat{\theta}(\lambda)$ : both consumers and workers benefit from the increase in  $\lambda$ .
2.  $\lambda < \hat{\lambda}$  and  $\theta > \hat{\theta}(\lambda)$ : consumers benefit from the increase in  $\lambda$  but workers suffer from it.
3.  $\lambda > \hat{\lambda}$  and  $\theta < \hat{\theta}(\lambda)$ : consumers suffer from the increase in  $\lambda$  but workers benefit from it.
4.  $\lambda > \hat{\lambda}$  and  $\theta > \hat{\theta}(\lambda)$ : both consumers and workers suffer from the increase in  $\lambda$ .

An increase in overlapping ownership benefits consumers when the level of overlapping ownership is low, but it hurts them when the level is high. Workers may also benefit from an increase in the overlapping ownership, which is the case when job differentiation is sufficiently small. Note that such a situation arises only when the overlapping ownership increases training levels—without an increase in training, consumers and workers always suffer from overlapping ownership that simply reduces competition.

**Positive switching costs** How do switching costs factor into this? When  $\sigma > 0$ , using the characterization of Proposition A.5 we have

$$\frac{dL_I^*}{d\lambda} = \frac{\sigma}{\theta(2n-1+\lambda)^2} > 0$$

which implies that the equilibrium retention rate increases with  $\lambda$ . This follows from the fact that, as firms earn more margins from retained workers than poached workers,

overlapping ownership induces firms to lower the poaching wages more than retention wages—to induce workers to stay at their initial employers.

Now, consider the effect of overlapping ownership on equilibrium quantities and training. To this end, define

$$\begin{aligned} A(\sigma, \lambda) &:= (1 - \lambda) \frac{\gamma}{1 - \gamma} \frac{n - 1}{n}, \\ B(\sigma, \lambda) &:= L_I^* + \lambda(1 - L_I^*), \\ a &:= \frac{\eta^2}{n\kappa}. \end{aligned}$$

Then, we can write

$$\begin{aligned} Q^*(\sigma, \lambda) &= \frac{1}{n} \frac{A(\sigma, \lambda)}{1 + A(\sigma, \lambda)[1 - aB(\sigma, \lambda)]}, \\ h^*(\sigma, \lambda) &= \frac{a}{\eta} \frac{A(\sigma, \lambda)B(\sigma, \lambda)}{1 + A(\sigma, \lambda)[1 - aB(\sigma, \lambda)]}. \end{aligned}$$

A calculation shows that  $dQ^*(\sigma, \lambda)/d\lambda \geq 0$  holds if and only if

$$\frac{\partial A}{\partial \lambda} + a \times [A(\sigma, \lambda)]^2 \frac{\partial B}{\partial \lambda} \geq 0, \quad (\text{A.7})$$

and  $dh^*/d\lambda \geq 0$  holds if and only if

$$\frac{\partial A}{\partial \lambda} B + \frac{\partial B}{\partial \lambda} A(1 + A) \geq 0. \quad (\text{A.8})$$

From these two characterizations, we obtain the following propositions:

**Proposition A.7.** *There exists  $(\sigma_Q, \sigma_h)$  such that  $\sigma_Q < \sigma_h$ , and*

1.  $dQ^*/d\lambda \geq 0$  if and only if  $\sigma \leq \sigma_Q$ ;
2.  $dh^*/d\lambda \geq 0$  if and only if  $\sigma \leq \sigma_h$ .

*Consumer surplus increases with  $\lambda$  if and only if  $\sigma \leq \sigma_Q$ . When  $\sigma < \sigma_h$ , equilibrium worker surplus may increase or decrease with  $\lambda$ . When  $\sigma \geq \sigma_h$ , both equilibrium retention and poaching wages decrease with  $\lambda$ , so worker surplus decreases with it.*

We note that  $Q^*$  may either increase or decrease with  $\lambda$  for all the relevant range of  $\sigma$ . In this case, we set either  $\sigma_Q = 0$  or  $\sigma_Q = \infty$ . Similar notation applies to  $\sigma_h$ .

*Proof.* First, we show the threshold property of  $\sigma_Q$  and  $\sigma_h$ . To see this, note that we have

$$\begin{aligned}\frac{\partial A}{\partial \sigma} &= 0, \\ \frac{\partial A}{\partial \lambda} &= -\frac{\gamma}{1-\gamma} \frac{n-1}{n} < 0, \\ \frac{\partial B}{\partial \sigma} &= \frac{\partial L_I^*}{\partial \sigma} (1-\lambda) > 0, \\ \frac{\partial B}{\partial \lambda} &= \frac{\partial L_I^*}{\partial \lambda} (1-\lambda) + 1 - L_I^* \\ &= (1-\lambda) \frac{\sigma}{\theta(2n-1+\lambda)^2} + \frac{n-1}{n} - \frac{\sigma(n-1+\lambda)}{n\theta(2n-1+\lambda)} > 0, \\ \frac{\partial^2 A}{\partial \sigma \partial \lambda} &= 0, \\ \frac{\partial^2 B}{\partial \sigma \partial \lambda} &= -\frac{1}{\theta} \frac{(2n-1+\lambda)(n-1+\lambda) - n(1-\lambda)}{n(2n-1+\lambda)} < 0.\end{aligned}$$

Hence, the left-hand side of the equation (A.7) can be written as

$$-\frac{\gamma}{1-\gamma} \frac{n-1}{n} + aA^2 \frac{\partial B}{\partial \lambda},$$

which has the derivative with respect to  $\sigma$  as follows:

$$aA^2 \frac{\partial^2 B}{\partial \sigma \partial \lambda} < 0.$$

Hence, there exists a critical value  $\sigma_Q$  such that  $dQ^*/d\lambda \geq 0$  if and only if  $\sigma \leq \sigma_Q$ . Similarly, the left-hand side of equation (A.8) can be written as

$$-\frac{\gamma}{1-\gamma} B + \frac{\partial B}{\partial \lambda} A(1+A),$$

which has the derivative with respect to  $\sigma$  as follows:

$$-\frac{\gamma}{1-\gamma} \frac{\partial B}{\partial \sigma} + \frac{\partial^2 B}{\partial \sigma \partial \lambda} A(1+A) < 0.$$

Hence, there exists a critical value  $\sigma_h$  such that  $dh^*/d\lambda \geq 0$  if and only if  $\sigma \leq \sigma_h$ .

Next, we show  $\sigma_Q < \sigma_h$ . To see this, suppose  $\sigma = \sigma_Q$  and consider whether the condition (A.8) holds or not. When the condition (A.7) holds with equality, the left-hand side of equation (A.8) can be rewritten as

$$\frac{\partial B}{\partial \lambda} A [1 - A(1 - aB)] > 0,$$

which holds because  $1 - A(1 - aB) > 0$  must hold for the regularity condition (Assumption 3) to hold. Thus, we have  $\sigma_Q < \sigma_h$ .

Finally, we turn to the welfare results. Because  $CS^*$  increases with  $\lambda$  if and only if  $Q^*$  increases. The comparative statics result on  $WS^*$  for  $\sigma < \sigma_h$  comes from Proposition A.6 for the case where  $\sigma = 0$ . When  $\sigma \geq \sigma_h$ , the derivative of equilibrium wages are given by

$$\begin{aligned}\left. \frac{dw_I^*}{d\lambda} \right|_{\sigma=\bar{\sigma}} &= \eta \left( \frac{dh^*}{d\lambda} Q^* + h^* \frac{dQ^*}{d\lambda} \right) - \frac{d}{d\lambda} \left[ \frac{\theta + \sigma - (1+\lambda)\Delta w^*}{1-\lambda} \right], \\ \left. \frac{dw_O^*}{d\lambda} \right|_{\sigma=\bar{\sigma}} &= \eta \left( \frac{dh^*}{d\lambda} Q^* + h^* \frac{dQ^*}{d\lambda} \right) - \frac{d}{d\lambda} \left[ \frac{\theta - \frac{\sigma - (1+\lambda)\Delta w^*}{n-1}}{1-\lambda} \right],\end{aligned}$$

where the first bracketed terms are negative. We show that the second bracketed terms are negative as well.

Noting that the condition  $L_I^* \leq 1$  is equivalent to

$$\begin{aligned}\theta + \frac{(n-1+\lambda)\sigma}{2n-1+\lambda} &\leq n\theta \\ \implies \theta &\geq \frac{1}{n-1} \frac{(n-1+\lambda)\sigma}{2n-1+\lambda} \geq \frac{\sigma}{2n-1+\lambda},\end{aligned}$$

we have

$$\begin{aligned}\frac{d}{d\lambda} \left[ \frac{\theta + \sigma - (1+\lambda)\Delta w^*}{1-\lambda} \right] &= \frac{\theta + \sigma - (1+\lambda)\Delta w^*}{(1-\lambda)^2} - \frac{\Delta w + (1+\lambda) \frac{d\Delta w^*}{d\lambda}}{1-\lambda} \\ &= \frac{\theta + \sigma - 2\Delta w^* - (1-\lambda^2) \frac{d\Delta w^*}{d\lambda}}{(1-\lambda)^2} \\ &= \frac{1}{(1-\lambda)^2} \left[ \theta + \sigma - \frac{2n\sigma}{2n-1+\lambda} + (1-\lambda)^2 \frac{n\sigma}{(2n-1+\lambda)^2} \right] \\ &= \frac{1}{(1-\lambda^2)} \left[ \theta - \frac{(1-\lambda)\sigma}{2n-1+\lambda} + (1-\lambda^2) \frac{n\sigma}{(2n-1+\lambda)^2} \right] \\ &> 0,\end{aligned}$$

which implies that  $dw_I^*/d\lambda < 0$ . Similarly, we can show that  $dw_O^*/d\lambda < 0$ . This establishes that  $WS^*$  decreases with  $\lambda$  when  $\sigma \geq \sigma_h$ .  $\square$

Proposition A.7 demonstrates that, when switching costs are sufficiently high, in the sense that  $\sigma \geq \sigma_h$ , then an increase in  $\lambda$  never increases the equilibrium training level. In such a case, an increase in  $\lambda$  hurts both consumers and workers, implying that the potential for pro-competitive effects of overlapping ownership disappears when labor mobility is low.

## A.5 Model with initial hiring

Here, we extend the main model by introducing an additional stage in which firms choose how many workers to hire. We detail the model below.

### A.5.1 Setting

In the extended model, there is a continuum of workers  $x \in [0, \bar{X}]$  with unitary density. Each worker  $x \in [0, \bar{X}]$  has an opportunity cost  $\phi(x)$  of working for any firm in the

industry. We assume that  $\phi(x)$  is increasing in  $x$ , and in the limit,  $\lim_{x \rightarrow \bar{X}} \phi(x) = \infty$  holds. Firms choose the mass of workers  $z_i$  it hires, which is assumed to be unobservable to workers. Workers initially hired at firm  $i$  get wage  $w_i^0$ , which is determined to clear the market.

When choosing which firm to work for (if any), each worker observes only the initial wage offers  $\mathbf{w}^0 = (w_i^0)_{i=1}^n$ . Hence, each worker form an expectation over the vector of initial hiring  $\mathbf{z}^e = (z_i^e)_{i=1}^n$  and future wages  $\mathbf{w}^{1e}$  after observing initial wages  $\mathbf{w}^0$ .

**Workers** The net present value of the worker  $x \in [0, \bar{X}]$  from working for firm  $i$  is given by

$$V_i^0(w_i^0, \mathbf{w}^1, x) = w_i^0 - \phi(x) + \delta \text{WS}_i^1(\mathbf{w}^1),$$

where  $\delta \in [0, 1]$  is a common discount factor, and  $\text{WS}_i^1(\mathbf{w}^1)$  is the post-employment worker surplus, which is equivalent to the worker surplus WS in the main model.

**Wage determination** We assume that given the expected future wages  $\mathbf{w}^{1e}$  and a vector of initial hiring levels  $\mathbf{z}$  set by firms, the initial employment wages  $\mathbf{w}^0$  are determined to clear the market; that is, exactly  $z_i$  workers will be employed by firm  $i$  as a consequence of workers' utility-maximizing choices. This assumption is often made in the literature of network goods and platforms (e.g., [Katz and Shapiro, 1985](#)) and multi-product quality competition ([Johnson and Myatt, 2003](#)).

**Timing and equilibrium** The game proceeds as follows:

1. Firms simultaneously and independently choose  $(z_i)_{i=1}^n$
2. Initial wages  $(w_i^0)_{i=1}^n$  are determined in the market, and workers choose which firm to work for, if any.
3. Each firm observes its own hiring level  $z_i$  alone and chooses its training level  $h_i$ .
4. Without observing  $(z_j, h_j)_{j \neq i}$ , each firm  $i$  chooses  $p_i$  and  $(w_{ji})_{j=1}^n$ .

We focus on symmetric perfect Bayesian equilibrium with passive belief, which could be viewed as fulfilled-expectation equilibrium in the spirit of [Katz and Shapiro \(1985\)](#). Formally, a symmetric perfect Bayesian equilibrium is defined by a tuple  $(z^{**}, w^{0**}, w_I^{**}, w_O^{**}, h^{**}, p^{**})$  such that

1. Each firm chooses  $z_i = z^{**}$ ;
2. Initial wages are set at  $w_i^0 = w^{0**}$  and exactly  $z^{**}$  workers are hired by each firm;
3. Each firm sets  $h^{**}$ ;
4. Each firm sets  $w_{ii} = w_I^{**}$  and  $w_{ji} = w_O^{**}$  for  $j \neq i$ , and  $p_i = p^{**}$ .

With passive belief, after observing off-the-equilibrium initial wages  $\mathbf{w}^0 \neq (w^{0**}, \dots, w^{0**})$ , workers continue to believe that all firms have chosen  $z^{**}$  and will therefore choose the on-path equilibrium strategies  $(h^{**}, p^{**}, w_I^{**}, w_O^{**})$ .

### A.5.2 Equilibrium analysis

We derive and characterize the symmetric perfect Bayesian equilibrium with passive beliefs. To this end, we solve the model backwardly, along with the assumption of passive belief. We first solve for the equilibrium after any history where initial hiring has settled. Then, we derive the “inverse labor supply,” which is derived from the equilibrium choice of workers and market-clearing wage determination. Using this inverse supply, we derive the equilibrium initial hiring to complete the equilibrium characterization.

Let  $ca_i = \{h_i, p_i, (w_{ij})_{j=1}^n\}$  be the *continuation action* of firm  $i$ , and  $ca_{-i} = (ca_j)_{j \neq i}$  be the continuation action profile other than firm  $i$ . Also let  $ca^*$  be the symmetric-equilibrium continuation action on the on-equilibrium path, and  $ca_{-i}^*$  a continuation action profile other than firm  $i$  with  $ca_j = ca^*$  for all  $j \neq i$ . Similarly, let  $z_{-i} = (z_j)_{j \neq i}$  be the vector of initial hiring by firms other than  $i$ , and  $z_{-i}^*$  be  $z_{-i}$  such that  $z_j = z^*$  for all  $j \neq i$ .

**Equilibrium after initial hiring** Consider the stage where initial hiring:  $(z_i)_{i=1}^n$  has been made. For any given price vector  $\mathbf{p} = (p_i)_{i=1}^n$ , the product-market demand for firm  $i$  is given by  $D_i(\mathbf{p})$ ; for any given retention and poaching wages  $(w_{ij})$  offered to workers initially hired by firm  $j$ , the worker’s probability to stay at or switch to firm  $i$  is given by  $L_{ji}$ , as shown in the main text.

Generally, firms’ best-responses in this subgame depend on their belief over hiring and training choice made by other firms. Nonetheless, with our assumption of passive beliefs, the relation between the history and beliefs will be greatly simplified, as we show below.

Consider the situation where firm  $i$  have observed  $w_i$  but not  $z_{-i}$ . Under the passive belief assumption, firm  $i$  continues to believe that other firms have set  $z^{**}$ . Because it is common knowledge in equilibrium that all firms believe that all other firms have chosen  $z_j = z^{**}$ , firm  $i$  believes that other firms will choose symmetric equilibrium continuation actions  $ca_j = ca^*$ . Hence, each firm chooses its continuation action  $ca_i$  under such a belief.

Firm  $i$ ’s problem after its choice of  $z_i$ , is then reduced to the following maximization problem,

$$\max_{ca_i} \Pi_i^1(ca_i, ca_{-i}^{**}; z_i, z_{-i}^{**}) = D_i(\mathbf{p})[p_i - c(H_i)] - \sum_{j=1}^n z_j L_{ji} w_{ji} - z_i \tau(h_i),$$

where

$$H_i = \sum_{j=1}^n z_j L_{ji} h_j,$$

and  $z_j = z^{**}$ ,  $p_j = p^{**}$ ,  $h_j = h^{**}$ , for all  $j \neq i$ , and  $w_{jk} = w_O^{**}$  for all  $j \neq i$ ,  $k \neq j$ , and  $w_{jj} = w_I^{**}$  for all  $j \neq i$ .

Let  $ca_i^*(z_i)$  be the profit-maximizing  $ca_i$  given  $z_i$ . Before characterizing firms’ profit-maximizing choices in this stage, we offer a characterization of wage determination at the initial hiring stage. Afterward, we revisit firms’ profit-maximizing choices.

**Worker choice and inverse demand** Consider worker’s choice and the wage determination in the initial hiring stage. Let  $\mathbf{z}^{**} = (z^{**}, \dots, z^{**})$  be the vector of symmetric equilibrium initial hiring levels, and  $\mathbf{z} = (z_i)_{i=1}^n$  be the vector of actual initial hiring levels.

Given the initial offerings  $\mathbf{w}^0 = (w_i^0)_{i=1}^n$ , workers still believe that firms have chosen  $\mathbf{z}^{**}$  and thus continue to expect that future wages will be given by  $\mathbf{w}^{1**}$ . Hence, a worker with opportunity cost  $\phi(x)$  enjoys the following net expected utility from being hired by firm  $i$ :

$$V_i^0(w_i^0, \mathbf{w}^{1**}) = w_i^0 + \delta \text{WS}_i^1(\mathbf{w}^{1**}) - \phi(x).$$

Now we characterize how the equilibrium initial wages are determined to clear the market. Under the assumption of symmetric equilibrium, all  $\text{WS}_i^1(\mathbf{w}^{1**})$  is the same across  $i = 1, \dots, n$ . Let  $\text{WS}^{1**}$  be the common value of  $\text{WS}_i^1(\mathbf{w}^{1**})$ . Then, the worker's net utility from being employed by firm  $i$  is reduced to

$$w_i^0 + \delta \text{WS}^{1**} - \phi(x).$$

To have possible employments  $z_i > 0$  for  $i = 1, \dots, n$ ,  $w_i^0$  must be the same across  $i = 1, \dots, n$ . Let  $w^0$  be the common value. Also let  $Z := \sum_{i=1}^n z_i$ . Given  $w^0$ , the worker  $x$ 's utility from being employed by any employer is given by

$$w^0 + \delta \text{WS}^{1**} - \phi(x).$$

For  $Z$  worker to be initially employed,  $w^0$  must satisfy

$$w^0 + \delta \text{WS}^{1**} - \phi(Z) = 0.$$

Then, as a function of total employment  $Z$ , the initial wage  $w^0$  can be written as

$$w^0(Z) = \phi(Z) - \delta \text{WS}^{1**}.$$

Note that, the worker surplus in this setting is

$$\begin{aligned} \text{WS}^0(Z) &= \int_0^Z [w^0(Z) + \delta \text{WS}^{1**} - \phi(x)] dx \\ &= \int_0^Z [\phi(Z) - \phi(x)] dx, \end{aligned}$$

which depends only on the value of  $Z$ .

**Initial hiring choice** Now consider the initial stage in which firms choose the hiring levels  $(z_i)_{i=1}^n$ . As we discussed above, under the passive belief assumption, each firm  $i$ 's choice of hiring level  $z_i$  does not affect the future actions of other firms. This implies that each firm  $i$  takes the future moves of other firms as given.

Hence, at this stage, each firm  $i$  chooses  $z_i$  to maximize its own profit  $\Pi_i^0$ , which is given by

$$\Pi_i^0(z_i; z_{-i}^{**}, \text{ca}_{-i}^{**}) = -[\phi(Z) - \text{WS}^{1**}]z_i + \delta \Pi_i^1(\text{ca}_i^*(z_i), \text{ca}_{-i}^{**}; z_i, z_{-i}^{**}).$$

Because the sequential choice of  $z_i$  and  $\text{ca}_i$  is equivalent to simultaneous choice of  $s_i := (z_i, \text{ca}_i)$ , the equilibrium condition can be reduced to the solution to the following problem.

**Lemma A.1.** *A symmetric equilibrium with passive belief is given by a tuple  $(z^{**}, p^{**}, h^{**}, w_I^{**}, w_O^{**})$  that is the solution to the problem*

$$\max_{s_i} \Pi_i^0(s_i, s_{-i}^{**}).$$

where  $s_{-i}^{**}$  is the vector  $(s_j)_{j \neq i}$  such that  $z_j = z_j^{**}$  and  $\text{ca}_j = \text{ca}_j^{**}$  for all  $j \neq i$ .

**Equilibrium characterization** Using Lemma A.1, we now characterize equilibrium. From the optimality requirements, the equilibrium values of  $(z^{**}, p^{**}, h^{**}, w_I^{**}, w_O^{**})$  must satisfy the first-order conditions:

$$\begin{aligned} \frac{\partial \Pi_i^0}{\partial z_i} &= \underbrace{-\phi'(Z)z_i - [\phi(Z) - \delta \text{WS}^{1**}]}_{=-\frac{dw^0(Z)}{dZ}z_i - w^0(Z)} \\ &\quad + \delta \left[ \underbrace{-c'(H_i)D_i(p)L_{ii}h_i - L_{ii}w_{ii} - \tau(h_i)}_{\partial \Pi_i^1 / \partial z_i} \right] = 0, \end{aligned} \quad (\text{FOC-z})$$

$$\frac{\partial \Pi_i}{\partial w_{ij}} = \delta \left[ z_j \frac{\partial L_{ji}}{\partial w_{ij}} [-c'(H_i)D_i h_j - w_{ij}] - z_j L_{ji} \right] = 0,$$

$$\frac{\partial \Pi_i}{\partial h_i} = \delta [-z_i c'(H_i)L_{ii}D_i - z_i \tau'(h_i)] = 0, \quad (\text{FOC-h})$$

$$\frac{\partial \Pi_i}{\partial p_i} = \delta \left[ \frac{\partial D_i}{\partial p_i} [p_i - c(H_i)] + D_i \right] = 0. \quad (\text{FOC-p})$$

Now we further simplify the equilibrium conditions. Noting that we have

$$H_i = \sum_{j=1}^n z_j h_j L_{ji} = z^{**} h^{**}$$

in a symmetric equilibrium, we obtain the symmetric equilibrium conditions for  $(w_I^{**}, w_O^{**})$ :

$$\begin{aligned} w_I^{**} &= -c'(z^{**} h^{**}) h^{**} Q(p^{**}) - X_I^{**}, \\ w_O^{**} &= -c'(z^{**} h^{**}) h^{**} Q(p^{**}) - X_O^{**}. \end{aligned}$$

These two equations lead to the following lemma.

**Lemma A.2.** *The equilibrium retention rate and poaching rate satisfy  $L_I^{**} = L_I^*$  and  $L_O^{**} = L_O^*$ . The equilibrium worker non-monetary utility also satisfies  $\text{NMU}^{**} = \text{NMU}^*$ .*

Hence, for any given  $(p, h, z)$ , the equilibrium condition is rewritten as  $(w_I(p, h, z), w_O(p, h, z))$ :

$$w_I(p, h, z) = -c'(zh)hQ(p) - X_I^*, \quad (\text{A.10})$$

$$w_O(p, h, z) = -c'(zh)hQ(p) - X_O^*. \quad (\text{A.11})$$

Next, consider the equilibrium conditions for  $(p^{**}, h^{**}, z^{**})$ . As is the case in the main model, for a given  $z$ , the first-order conditions (FOC-h) and (FOC-p) can be written as

$$f_p(p, h, z; \sigma) = \frac{p - c(zh)}{p} - \frac{1}{\varepsilon_Q(p)} = 0, \quad (\text{A.12})$$

$$f_h(p, h, z; \sigma) = -c'(zh)L_I^*Q(p) - \tau'(h) = 0. \quad (\text{A.13})$$

Now consider the equilibrium condition for  $z$ . For a given  $(p, h, z)$ , the equilibrium condition (FOC-z) can be rewritten as

$$-\phi'(nz)z - \phi(nz) + \delta [\text{WS}^{1**} - c'(zh)Q(p)L_I h - L_I w_I - \tau(h)] = 0$$

Furthermore, with symmetric wages  $\{w_I(p, h, z), w_O(p, h, z)\}$ , the equilibrium post-employment worker surplus is

$$\begin{aligned} \text{WS}^1(p, h, z) &= L_I^* w_I(p, h, z) + (1 - L_I^*) w_O(p, h, z) + \text{NMU}^* \\ &= -c'(zh)hQ(p) - L_I^* X_I^* - (1 - L_I^*) X_O^* + \text{NMU}^* \end{aligned}$$

Combined, the equilibrium condition (FOC-z) is further rewritten as

$$\begin{aligned} f_z(p, h, z; \sigma) &= -\phi'(nz)z - \phi(nz) + \delta \left\{ -c'(zh)hQ(p) - (1 - L_I^*) X_O^* + \text{NMU}^* - \tau(h) \right\} \\ &= 0 \end{aligned} \tag{A.14}$$

Summarizing the above discussion, we obtain the following lemma.

**Lemma A.3.** *Consider a model with initial hiring. In the symmetric equilibrium with passive belief,  $(z^{**}, p^{**}, h^{**}, w_I^{**}, w_O^{**})$  must satisfy the system of equations (A.14), (A.10), (A.11) (A.13), and (A.12).*

Let

$$\begin{aligned} \text{WS}^{0**} &= \text{WS}^0(nz^{**}), \\ \text{CS}^{**} &= \text{CS}(\mathbf{p}^{**}), \\ \Pi^{1**} &= Q(p^{**})[p^{**} - c(z^{**}h^{**})] - z^{**} \{L_I^* w_I^{**} + (1 - L_I^*) w_O^{**} + \tau(h^{**})\}, \\ \Pi^{0**} &= \delta \Pi^{1**} - [\phi(nz^{**}) - \delta \text{WS}^{1**}] z^{**}, \end{aligned}$$

be the equilibrium total worker surplus, consumer surplus, post-employment firm profit, and initial-period firm profit.

### A.5.3 Effects of labor market power

We can imagine two types of comparative statics with respect to  $\sigma$ . The first one is the effect of an *unexpected* change in  $\sigma$ , which occurs after the realization of  $z^{**}$ . Another one is an *expected* change in  $\sigma$ , which occurs before the realization of  $z^{**}$ . In the former case, the comparative statics analysis in the main model applies, while it does not in the latter case. Hereafter, we focus on the latter case.

The central question we ask with the expected change in  $\sigma$  is to what extent initial hiring affects the welfare implications of labor market power  $\sigma$ . For example:

1. If an increase in  $\sigma$  hurts future worker surplus, does the initial wage offer compensate it?
2. If an increase in  $\sigma$  increases future retention rate, does it increase the firms' incentive to hire workers and increase initial wage offers?
3. Does the presence of  $z$  change the nature of the effect of  $\sigma$  on workers, firms, and consumers?

In the following, we first review comparative statics with respect to  $\sigma$  for a given  $z$ , which amounts to considering unexpected changes in  $\sigma$ . Afterward, we consider the effects of expected changes in  $\sigma$ .

**Comparative statics after initial hiring** Suppose for a moment that all firms have set the same value of  $z$  and each firm has hired  $z$  workers.

Now, let a tuple  $\{h^*(z, \sigma), p^*(z, \sigma), w_I^*(z, \sigma), w_O^*(z, \sigma)\}$  be the solution to the system of equations (A.10), (A.11) (A.13), and (A.12) for any given value of  $z$ , and similarly define  $Q^*(z, \sigma)$ ,  $CS^*(z, \sigma)$ ,  $WS^{1*}(z, \sigma)$ , and  $IP^{1*}(z, \sigma)$ . Note that after redefining  $\bar{c}(h) := c(zh)$ , the conditions for the tuple  $\{h^*(z, \sigma), p^*(z, \sigma), w_I^*(z, \sigma), w_O^*(z, \sigma)\}$  is exactly the same as what we obtained in the main analysis—that is, the equilibrium condition for  $(h^*, p^*, w_I^*, w_O^*)$ . Hence, the comparative statics result in this subgame is equivalent to what we obtained—that is, Lemma 3.

**Corollary 1.** *For any given  $z$ , as  $\sigma$  increases,*

- (i) (a)  $h^*(z, \sigma)$  increases, and (b)  $Q^*(z, \sigma)$  increases and  $p^*(z, \sigma)$  decreases,
- (ii)  $CS^*(z, \sigma)$  increases, and
- (iii)  $WS^{1*}(z, \sigma)$  and  $IP^{1*}(z, \sigma)$  may increase or decrease depending on parameterizations.

**Initial hiring and worker surplus** Now, we examine the effect of  $\sigma$  on the initial hiring  $z^{**}$ , which determines the equilibrium worker surplus  $WS^{0**} = WS^0(nz^{**})$ .

We rewrite the equilibrium condition as

$$\begin{aligned} F_z(z; \sigma) &:= f_z(p^*(z, \sigma), h^*(z, \sigma), z; \sigma) \\ &= -\phi'(nz)z - \phi(nz) \\ &\quad + \delta \left\{ -c'(zh^*(z, \sigma))h^*(z, \sigma)Q^*(z, \sigma) - (1 - L_I^*)X_O^* + NMU^* - \tau(h^*(z)) \right\} \\ &= 0. \end{aligned}$$

Under the regularity assumption that  $\partial F_z / \partial z < 0$ ,  $z^{**}$  increases with  $\sigma$  if and only if  $\partial F_z / \partial \sigma > 0$ , which is equivalent to the condition that

$$\Lambda(z, \sigma) := -c'(zh^*(z, \sigma))h^*(z, \sigma)Q^*(z, \sigma) - (1 - L_I^*)X_O^* + NMU^* - \tau(h^*(z, \sigma))$$

increases with  $\sigma$  at  $z = z^{**}$ .

**Proposition A.8.** *The equilibrium initial hiring  $z^{**}$  can increase or decrease with  $\sigma$  depending on parameterizations, so does the equilibrium total worker surplus  $WS^{0**} = WS^0(nz^{**})$ .*

Note that the second-period worker surplus is given by

$$WS^{1*}(z, \sigma) = -c'(zh^*(z, \sigma))h^*(z, \sigma)Q^*(z, \sigma) - L_I^*X_I^* - (1 - L_I^*)X_O^* + NMU^*$$

Then, we have

$$\Lambda(z, \sigma) - WS^{1*}(z, \sigma) = L_I^*X_I^* - \tau(h^*(z, \sigma))$$

Hence, two differences are present between  $\Lambda(z, \sigma)$  and  $WS^{1*}(z, \sigma)$ . One is that firms in the initial stage internalize the gains from market power  $L_I^*X_I^*$  they exert to the retained

workers, which increases firms' incentive to hire workers. Another is that because initial employment accompanies the training cost  $\tau(h^*)$ , firms are less willing to hire workers in the initial period if the training level is high. The first effect makes the labor market power  $\sigma$  more beneficial to workers, whereas the second effect makes it more harmful.

Specifically, the derivative of this difference is given by

$$\begin{aligned} \frac{\partial}{\partial \sigma} \left[ \Lambda(z, \sigma) - \text{WS}^{1*}(z, \sigma) \right] &= \frac{\partial(L_I^* X_I^*)}{\partial \sigma} - \frac{\partial h^*}{\partial \sigma} \tau'(h^{**}) \\ &= \frac{\partial(L_I^* X_I^*)}{\partial \sigma} - \frac{\partial h^*}{\partial \sigma} c'(zh^*) h^* Q(p^*) \end{aligned}$$

which can be either positive or negative.

One implication of this result is that whenever  $\partial\Lambda/\partial\sigma < \partial\text{WS}^{1*}/\partial\sigma$  holds, workers in the extended model are more likely to be harmed by an increase in  $\sigma$  than in the baseline model. Therefore, our baseline result—an increase in labor market power  $\sigma$  never benefits workers when the training level is overinvestment and firms benefit from it—remains to hold in the extended model when  $\partial\Lambda/\partial\sigma < \partial\text{WS}^{1*}/\partial\sigma$  holds.

To see how our welfare implication in the baseline model carries over in this case, consider the situation where firm make a joint decision to increase the value of  $\sigma$  from  $\sigma$  to  $\sigma + \Delta\sigma$  by small amount  $\Delta\sigma > 0$ , taking  $z$  as given. In this case, firms increase  $\sigma$  if and only if it increases  $\text{IP}^{1*}(z, \sigma)$ . Then, such an increase in  $\sigma$  always decreases  $\text{WS}^{1*}(z, \sigma)$  when the equilibrium training level  $h^*(z, \sigma)$  is overinvestment. Furthermore, as we have discussed above, if  $\partial\Lambda/\partial\sigma < \partial\text{WS}^{1*}/\partial\sigma$ , the increase in  $\sigma$  also decreases  $z^{**}$  and thus  $\text{WS}^{0**}$ .

To make progress, we use the Shubik-Levitan demand, linear-quadratic training specification, and the Chen-Riordan spokes model of post-employment wage competition. Below are specific calculations.

**Proposition A.9.** *Consider the case where the demand function  $D_i(\cdot)$  is given by Shubik-Levitan specification, marginal cost function  $c(\cdot)$  and training cost function  $\tau(\cdot)$  are given by linear and quadratic specifications, and labor supply function  $L_{ji}(\cdot)$  is given by Chen-Riordan specification:*

$$\begin{aligned} D_i(\mathbf{p}) &= \frac{1}{n} \left[ v - p_i + \frac{\gamma}{1 - \gamma} (\bar{p} - p_i) \right], \\ c(H) &= \bar{C} - \eta H, \\ \tau(h) &= \frac{\kappa}{2} h^2 \\ L_{ji} &= \frac{1}{n} + \frac{\sum_{k \neq i} (v_{ji} - v_{jk})}{n(n-1)\theta}. \end{aligned}$$

*There exists a critical value  $\bar{v}$  such that  $\partial\Lambda/\partial\sigma - \partial\text{WS}^{1*}/\partial\sigma < 0$  if and only if  $v > \bar{v}$ .*

*Proof.* Under Shubik-Levitan demand specification, linear marginal cost, and quadratic

training cost, we have

$$Q^*(z, \sigma) = \frac{\left(1 + \frac{\gamma}{1-\gamma} \frac{n-1}{n}\right) (v - \bar{C})}{n \left(2 + \frac{\gamma}{1-\gamma} \frac{n-1}{n}\right) - \left(1 + \frac{\gamma}{1-\gamma} \frac{n-1}{n}\right) \frac{z\eta^2 L_I^*}{\kappa}},$$

$$h^*(z, \sigma) = \frac{\eta}{\kappa} \frac{\left(1 + \frac{\gamma}{1-\gamma} \frac{n-1}{n}\right) (v - \bar{C})}{n \left(2 + \frac{\gamma}{1-\gamma} \frac{n-1}{n}\right) - \left(1 + \frac{\gamma}{1-\gamma} \frac{n-1}{n}\right) \frac{z\eta^2 L_I^*}{\kappa}} L_I^*$$

When, we have

$$\begin{aligned} -c'(zh^*)h^*Q^* &= \eta h^* Q^* \\ &= \frac{\eta^2}{\kappa} L_I^* \left[ \frac{\left(1 + \frac{\gamma}{1-\gamma} \frac{n-1}{n}\right) (v - \bar{C})}{n \left(2 + \frac{\gamma}{1-\gamma} \frac{n-1}{n}\right) - \left(1 + \frac{\gamma}{1-\gamma} \frac{n-1}{n}\right) \frac{z\eta^2 L_I^*}{\kappa}} \right]^2 \end{aligned}$$

$$\frac{\partial [-c'(zh^*)h^*Q^*]}{\partial L_I^*} = \frac{\eta^2 \left[ \left(1 + \frac{\gamma}{1-\gamma} \frac{n-1}{n}\right) (v - \bar{C}) \right]^2 \left[ n \left(2 + \frac{\gamma}{1-\gamma} \frac{n-1}{n}\right) + \left(1 + \frac{\gamma}{1-\gamma} \frac{n-1}{n}\right) \frac{z\eta^2 L_I^*}{\kappa} \right]}{\kappa \left[ n \left(2 + \frac{\gamma}{1-\gamma} \frac{n-1}{n}\right) - \left(1 + \frac{\gamma}{1-\gamma} \frac{n-1}{n}\right) \frac{z\eta^2 L_I^*}{\kappa} \right]^3}$$

Furthermore, under Chen-Riordan labor market specification, we have

$$\begin{aligned} L_I &= \frac{1}{n} + \frac{\sigma - \Delta w}{n\theta}, \\ L_O &= \frac{1}{n} - \frac{\sigma - \Delta w}{n(n-1)\theta}, \\ X_I &= \theta + \sigma + -\Delta w \\ X_O &= \theta - \frac{\sigma - \Delta w}{n-1}. \end{aligned}$$

The equilibrium wage difference is given by

$$\Delta w^* = \frac{n\sigma}{2n-1},$$

which implies that

$$\begin{aligned} L_I^* &= \frac{1}{n} + \frac{(n-1)\sigma}{n(2n-1)\theta}, \\ X_I^* &= \theta + \frac{n-1}{2n-1}. \end{aligned}$$

Hence,

$$L_I^* X_I^* = n\theta \left( \frac{1}{n} + \frac{(n-1)\sigma}{n(2n-1)\theta} \right)^2$$

$$\frac{\partial (L_I^* X_I^*)}{\partial \sigma} = \frac{\partial L_I^*}{\partial \sigma} 2n\theta L_I^*$$

Using this specification, we obtain

$$\frac{\partial (\Lambda - \text{WS}^{1*})}{\partial \sigma} = \frac{\partial L_I^*}{\partial \sigma} \left[ 2n\theta L_I^* - \frac{\partial [-c'(zh^*)h^*Q^*]}{\partial L_I^*} \right],$$

which is negative if and only if

$$2n\theta L_I^* - \frac{\eta^2 \left[ \left(1 + \frac{\gamma}{1-\gamma} \frac{n-1}{n}\right) (v - \bar{C}) \right]^2 \left[ n \left(2 + \frac{\gamma}{1-\gamma} \frac{n-1}{n}\right) + \left(1 + \frac{\gamma}{1-\gamma} \frac{n-1}{n}\right) \frac{z\eta^2 L_I^*}{\kappa} \right]}{\left[ n \left(2 + \frac{\gamma}{1-\gamma} \frac{n-1}{n}\right) - \left(1 + \frac{\gamma}{1-\gamma} \frac{n-1}{n}\right) \frac{z\eta^2 L_I^*}{\kappa} \right]^3} \quad (\text{A.15})$$

is negative. Because the second negative term in the expression (A.15) grows as  $v$  increases, we have the critical value  $\bar{v}$  such that expression (A.15) is negative if and only if  $v > \bar{v}$ . □

Hence, if the product market demand is high, then the baseline result remains to hold in the extended model. On the contrary, when  $\partial(\Lambda - \text{WS}^{1*})/\partial\sigma > 0$  holds, then even if training level is overinvestment and post-employment industry profit goes up, workers may benefit from an increase in  $\sigma$ , which is not possible in the baseline model.

**Firm profit** The profit implication of initial hiring is ambiguous for multiple reasons. First, as we discussed above,  $z^{**}$  may increase or decrease with  $\sigma$ , and the impact of  $z^{**}$  on industry profit is also ambiguous, whether its direct effect or indirect effects through the changes in price, training levels, and wages.

To see the effect of an exogenous increase in  $z$ , first consider how  $(p^*, h^*)$  changes with  $z$ .

**Proposition A.10.** *An exogenous increase in  $z$  decreases  $p^*(z, \sigma)$ , but can either increase or decrease  $h^*(z, \sigma)$ .*

*Proof.* Remember that the equilibrium conditions for  $(p, h)$  are given by

$$\begin{aligned} f_p(p, h, z; \sigma) &= \frac{1}{\varepsilon_Q(p^*)} - \frac{p - c(zh)}{p} = 0, \\ f_h(p, h, z; \sigma) &= -c'(zh)L_I^*Q(p) - \tau'(h) = 0. \end{aligned}$$

Here, note that  $f_{pp} < 0$  and  $f_{hh} < 0$ , and  $f_{ph} < 0$  and  $f_{hp} < 0$  as shown below:

$$\begin{aligned} f_{pp} &= \frac{-\varepsilon'_Q(p)}{[\varepsilon_Q(p)]^2} - \frac{c'(zh)}{p^2} < 0, \\ f_{ph} &= \frac{zc'(zh)}{p} < 0, \\ f_{hp} &= -c'(zh)L_I^*Q'(p) < 0, \\ f_{hh} &= -zc''(zh)L_I^*Q(p^*) - \tau''(h) < 0. \end{aligned}$$

Furthermore, the derivatives of  $f_p$  and  $f_h$  with respect to  $z$  are given by

$$f_{pz} = \frac{h^* c'(zh^*)}{p^*} < 0,$$

$$f_{hz} = -c''(zh^*) h^* L_I^* Q(p^*) \leq 0.$$

Hence, by using the implicit function theorem, we have

$$\begin{aligned} \text{Sign} \left( \frac{\partial p^*}{\partial z} \right) &= \text{Sign} \left( f_{pz} - \frac{f_{hz}}{f_{hh}} f_{ph} \right) \\ &= \text{Sign} \left( \frac{h^* c'(zh^*)}{p^*} - h^* \frac{c''(zh^*) L_I^* Q(p^*)}{z c''(zh^*) L_I^* Q(p^*) + \tau'''(h^*)} \frac{z c'(zh^*)}{p} \right) \\ &= -, \\ \text{Sign} \left( \frac{\partial h^*}{\partial z} \right) &= \text{Sign} \left( f_{hz} - \frac{f_{pz}}{f_{pp}} f_{hp} \right) \\ &= \text{Sign} \left( -c'' h L_I Q + \frac{\frac{h c'}{p}}{\frac{-\varepsilon'_Q(p)}{[\varepsilon_Q(p)]^2} - \frac{c}{p^2}} c' L_I Q' \right) \\ &= \text{Sign} \left( -c'' Q + \frac{[c']^2 [-Q']}{\frac{\varepsilon'_Q(p)p}{[\varepsilon_Q(p)]^2} + \frac{c}{p}} \right) \end{aligned}$$

When  $c(H) = \bar{C} - \eta H$ ,  $h^*$  increases with  $z$ . When  $c(H) = \bar{C} - \eta \log H$  and  $Q(p) = \exp(-p)$ , then  $\varepsilon_Q(p) = p$ ,  $p = c + 1$ , and  $c'' = \eta/H^2$ ,  $c' = \eta/H$ . Then, we have

$$-c'' Q + \frac{[c']^2 [-Q']}{\frac{\varepsilon'_Q(p)p}{[\varepsilon_Q(p)]^2} + \frac{c}{p}} = \frac{\eta Q}{H^2} (-1 + \eta),$$

which is positive if and only if  $\eta > 1$ . □

We turn to industry profit. There are various ways through which  $z$  affects second-period industry profit. Suppose that an increase in  $\sigma$  increases  $z^{**}$ . How does it affect industry profits through the change in  $z^{**}$ ? To consider this, let  $\Pi^0(z, p, h)$  be the firm profit under symmetric strategies:

$$\begin{aligned} \Pi^1(z, p, h) &= Q(p)[p - c(zh)] - z \{ -c'(zh) h Q(p) - L_I^* X_I^* - (1 - L_I^*) X_O^* + \tau(h) \}, \\ \Pi^0(z, p, h) &= \delta \Pi^1(z, p, h, w_I, w_O) - \{ \phi(nz) - \delta [ -c'(zh) h Q(p) - L_I^* X_I^* - (1 - L_I^*) X_I^* + \text{NMU}^* ] \} z \\ &= \delta \{ Q(p)[p - c(zh)] - z \tau(h) + z \text{NMU}^* \} - \phi(nz) z \\ &= \delta [ \bar{\Pi}(z, p, h) + z \text{NMU}^* ] - \phi(nz) z, \end{aligned}$$

where

$$\bar{\Pi}(z, p, h) := Q(p)[p - c(zh)] - z \tau(h)$$

is the post-employment industry profit plus wage bill. Note that here the industry profit plus wage bill again serves as a useful welfare measure: the equilibrium firm profit depends on  $(p, h)$  only through the industry profit plus wage bill.

We have three effects to consider:

- Direct effect: Holding  $(p, h)$  fixed, an increase in  $z$  affects  $\Pi^0$  by

$$\frac{\partial \Pi^0}{\partial z} = \delta \left[ \frac{\partial \bar{\Pi}}{\partial z} + \text{NMU}^* \right] - [\phi(nz) + n\phi'(nz)].$$

- Price effect: Holding  $(z, h)$  fixed, an increase in  $p$  affects  $\Pi^0$  by

$$\frac{\partial \Pi^0}{\partial p} = \delta \frac{\partial \bar{\Pi}}{\partial p} > 0.$$

Hence, a price decrease led by an increase in  $z$  lowers industry profits.

- Training effect: Holding  $(z, p)$  fixed, an increase in  $h$  affects  $\Pi^0$  by

$$\frac{\partial \Pi^0}{\partial h} = \delta \frac{\partial \bar{\Pi}}{\partial h} > 0$$

Hence, industry profit tends to increase with  $h$  (holding  $p$  fixed). Because training may increase or decrease with  $z$ , the equilibrium industry profit may also increase or decrease with  $z$ .

**Consumer surplus** Lastly, we discuss whether consumer surplus increases or decreases with  $\sigma$ . In the baseline model, an increase in  $\sigma$  always increases consumer surplus by increasing the equilibrium training level  $h^*$ . In the current extension, however, this is no longer true. This is because the equilibrium price depends on the total human capital  $z^{**}h^{**}$  retained by each firm in equilibrium, and an increase in  $\sigma$  may lower the equilibrium hiring level  $z^{**}$ .

Because Proposition A.10 shows that an increase in  $z$  lowers prices, the equilibrium price decreases with  $\sigma$  whenever it increases  $z^{**}$ . Otherwise, the equilibrium price may increase because of the decrease in  $z^{**}$ . This implies that whenever an increase in  $\sigma$  increases worker surplus  $WS^{0**}$ , it also increases the equilibrium consumer surplus  $CS^{**}$ . Even if it decreases worker surplus, consumer surplus may still increase, but it may also decrease.

The fact that an increase in  $\sigma$  increases consumer surplus whenever it increases worker surplus implies that, if it increases the worker surplus and industry profit at the same time, it also increases the consumer surplus, leading to a clear-cut welfare improvement. As we discussed after Proposition A.9,  $z$  is more likely to increase than industry profit plus wage bill when  $v < \bar{v}$ . In such a case, a clear-cut welfare improvement may occur even when the equilibrium training level exhibits overinvestment.

**Summary** To summarize, incorporating an explicit initial hiring stage allows us to endogenize overall employment in the industry, which was treated as exogenous in Sections 3-5. An unanticipated increase in  $\sigma$  affects the equilibrium outcomes in the same way as in the main analysis. An anticipated increase in  $\sigma$  can reduce or raise employment. When employment increases, both consumers and workers are better off. However, when employment decreases, unlike the main model, consumer surplus can decrease in response to an increase in  $\sigma$ . Despite the possibility that both consumer and the worker surplus can decrease with an increase in  $\sigma$ , the possibility of a clear-cut improvement in welfare remains. Importantly, it can arise even when there is overinvestment in training.

## A.6 Model with firm-specific training

Here, we present the details of the extension with firm-specific training discussed in Section 6.2.

The wage first-order conditions are now given by

$$\begin{aligned}\frac{\partial \Pi_i}{\partial w_{ii}} &= \frac{\partial L_{ii}}{\partial w_{ii}} [-h_i c'(H_i) D_i - w_{ii} + s_i \beta] - L_{ii} = 0, \\ \frac{\partial \Pi_i}{\partial w_{ij}} &= \frac{\partial L_{ji}}{\partial w_{ij}} [-h_j c'(H_i) D_i - w_{ij}] - L_{ji} = 0.\end{aligned}$$

The training first-order condition is given by

$$\frac{\partial \Pi_i}{\partial s_i} = -\beta L_{ii} - \Psi'(s_i) = 0.$$

In the symmetric equilibrium, these first-order conditions can be written as

$$\begin{aligned}\Delta w &= \Delta X(\sigma - \Delta w) + \beta s, \\ \beta L_I &= \Psi'(s).\end{aligned}$$

This system of equations gives the pair of solutions  $(s^*, L_I^*)$  that is independent of  $(p, h)$ . By applying the implicit function theorem, it is immediate that the equilibrium retention rate  $L_I^*$  is increasing in the importance of firm-specific training  $\beta$  (provided that the equilibrium satisfies the stability condition).

**Proposition A.11.** *Suppose that  $\Psi(\cdot)$  is convex enough to satisfy*

$$1 + \Delta X'(\sigma - \Delta w) - \frac{\beta^2 L_I'(\sigma - \Delta w)}{\Psi''(s)} > 0.$$

Then, As  $\beta$  increases,

(i)  $L_I^*$ ,  $h^*$ ,  $Q^*$ ,  $w_I^*$ , and  $w_O^*$  increase.

*Proof.* Let  $\tilde{s}(L_I, \beta) = (\Psi')^{-1}(\beta L_I)$  be the solution to the equation  $\beta L_I = \Psi'(s)$ . Then, the equilibrium condition for  $\Delta w$  is given by

$$G(\Delta w, \beta) := \Delta w - \Delta X(\sigma - \Delta w) + \beta \tilde{s}(L_I(\sigma - \Delta w), \beta).$$

Under the presumption, we have a unique solution  $\Delta w^*$  to  $G(\Delta w, \beta) = 0$ . Using the implicit function theorem, we have

$$\frac{d \Delta w^*}{d \beta} = -\frac{\partial G / \partial \beta}{\partial G / \partial \Delta w} = -\frac{s + \beta \frac{L_I}{\Psi''(s)}}{1 + \Delta X'(\sigma - \Delta w) - \frac{\beta^2 L_I'(\sigma - \Delta w)}{\Psi''(s)}} < 0.$$

Hence, we immediately have  $d L_I^* / d \beta > 0$ ,  $d h^* / d \beta > 0$ , and  $d Q^* / d \beta > 0$ .

As we assume that  $X_O(y)$  is decreasing in  $y$ ,  $X_O(\sigma - \Delta w^*)$  is decreasing in  $\beta$ , which implies that  $w_O^*$  is increasing in  $\beta$ . Finally, noting that  $w_I^* = w_O^* - \Delta w^*$ ,  $w_I^*$  is increasing in  $\beta$ , which completes the proof.  $\square$

Because of the independence of  $s^*$  on  $h$ , we can define the industry profit gross of wage bill,  $\bar{\Pi}(h)$  analogously,

$$\bar{\Pi}(h, L_I) = Q(\tilde{p}(h)) [\tilde{p}(h) - c(h)] - \tau(h) - K + \beta s L_I - \Psi(s).$$

Hence, at the equilibrium training level  $h^*$ , the derivative of  $\bar{\Pi}$  with respect to  $h$  is given by the same expression:

$$\frac{\partial \bar{\Pi}(h^*, L_I^*)}{\partial h} = \tilde{p}'(h^*) Q(p^*) \psi(p^*) - c'(h^*) (1 - L_I^*) Q(p^*).$$

Hence, Proposition 3 applies in this setting as well. In this case, firm-specific training affects the likelihood of overinvestment only through the equilibrium value of  $L_I^*$ . For example, when  $\beta$  increases, firm-specific training becomes more important, and  $s^*$  increases and so does  $L_I^*$ . Therefore, overinvestment is more likely when  $\beta$  is high. In other words, the more important firm-specific training is, the more likely industry-specific training is overinvestment.

However, in the presence of industry-specific training, an increase in  $L_I$  has an additional channels through which an increase in  $\sigma$  affects industry profit plus wage bill. To see this, suppose that  $L_I$  increases while  $h$  is held fixed. Then, its effect on industry profit plus wage bill is

$$\frac{\partial \bar{\Pi}}{\partial L_I} = \beta s > 0.$$

Hence, an increase in  $L_I$  (which is led by an increase in  $\sigma$  in our comparative statics exercise) increases industry profit plus wage bill by increasing the benefit of firm-specific training.

To summarize, as firm-specific training becomes more important, the retention rate increases, and thereby industry-specific training becomes more likely to be overinvestment. This makes the clear-cut welfare improvement of labor market power (measured by switching costs) less likely. At the same time, as firm-specific training becomes more important, an increase in retention rate driven by an increase in switching costs increases the effectiveness of firm-specific training, which makes the clear-cut welfare improvement more likely. The more important firm-specific training is, the more important the latter effect is. Hence, we expect that when  $\beta$  is low, an increase in  $\beta$  makes a clear-cut welfare improvement less likely, whereas when  $\beta$  is high, an increase in  $\beta$  may make a clear-cut welfare improvement more likely.

## References

- Anderson, Simon P, Andre De Palma, and Jacques-Francois Thisse**, *Discrete Choice Theory of Product Differentiation*, MIT Press, 1992.
- Chen, Yongmin and Michael H Riordan**, “Price and Variety in the Spokes Model,” *Economic Journal*, 2007, 117 (522), 897–921.
- Hotelling, Harold**, “Stability in Competition,” *Economic Journal*, 1929, 39 (153), 41–57.
- Johnson, Justin P and David P Myatt**, “Multiproduct quality competition: Fighting brands and product line pruning,” *American Economic Review*, 2003, 93 (3), 748–774.
- Katz, Michael L and Carl Shapiro**, “Network externalities, competition, and compatibility,” *American Economic Review*, 1985, 75 (3), 424–440.
- López, Ángel L and Xavier Vives**, “Overlapping Ownership, R&D Spillovers, and Antitrust Policy,” *Journal of Political Economy*, 2019, 127 (5), 2394–2437.
- Nocke, Volker and Nicolas Schutz**, “Multiproduct-firm Oligopoly: An Aggregative Games Approach,” *Econometrica*, 2018, 86 (2), 523–557.
- Perloff, Jeffrey M and Steven C Salop**, “Equilibrium with Product Differentiation,” *Review of Economic Studies*, 1985, 52 (1), 107–120.
- Shubik, Martin and Richard Levitan**, *Market Structure and Behavior*, Harvard University Press, 2013.
- Teh, Tat-How, Chunchun Liu, Julian Wright, and Junjie Zhou**, “Multihoming and Oligopolistic Platform Competition,” *American Economic Journal: Microeconomics*, 2023, 15 (4), 68–113.
- Vives, Xavier and Orestis Vravosinos**, “Free entry in a Cournot market with overlapping ownership,” *American Economic Journal: Microeconomics*, forthcoming.