Optimal Taxation and Market Power*

Jan Eeckhout UPF Barcelona†
Chunyang Fu UCASS‡
Wenjian Li Zhejiang University§
Xi Weng Peking University¶

February 29, 2024

Abstract

Should optimal income taxation change when firms have market power? The recent rise of market power has led to an increase in income inequality and a deterioration in efficiency and welfare. We analyze how the planner can optimally set taxes on the labor income of workers and on the profits of entrepreneurs to induce a constrained efficient allocation. Our theory obtains optimal tax rates that depend on markups and identifies four different channels of taxes on welfare: 1. the Mirrleesian incentive effect; 2. the Pigouvian tax correction of the negative externality of market power; 3. redistribution through altered factor prices; 4. reallocation of output towards the most productive firms. Our quantitative analysis of the US economy in 1980 and 2019 shows that the average optimal labor income tax rate in 2019 falls by 10 percentage points compared to 1980. Instead, the optimal average profit tax rate increases by 3 percentage points, and by 25 percentage points at the top. The optimal profit taxes are positive and regressive at the top, but are less regressive in 2019 than in 1980. Our theory provides concrete proposals how policymakers can optimally use income taxation to redistribute income to the poor while incentivizing production.

Keywords. Optimal Taxation. Optimal profit tax. Market Power. Market Structure. Markups.
JEL. D3; D4; J41.

*We thank Marco Bassetto, Ash Craig, Dirk Krueger, Simon Mongey, Sergio Ocampo, Florian Scheuer, Stefanie Stantcheva, and Aleh Tsyvinski as well as numerous colleagues and seminar participants for useful comments and insightful discussions. Renjie Bao provided superb research assistance. We gratefully acknowledge support from the National Natural Science Foundation of China (Grant Nos. 72373133, 72003198, 72192843 and 72225001), the ERC, Advanced grant 882499, the AEI/FEDER, UE-PGC2018-096370-B-I00, the China Information Economics Association, the Key Laboratory of Mathematical Economics and Quantitative Finance, and the Ministry of Education and Peking University.
†jan.eeckhout@upf.edu – ICREA-BSE-CREI
‡fuchunyang@pku.edu.cn
§wenjian_li@zju.edu.cn
¶wengxi125@gsm.pku.edu.cn

1 Introduction

Market power has an impact on both inequality and efficiency. As market power increases, the share of output accrues disproportionately to owners of monopolistic firms and less to workers. In addition, market power creates inefficiencies in the allocation of resources as prices are too high which leads to deadweight loss and a reduction in welfare. Given that market power changes both efficiency and inequality, understanding the effect of market power on tax design is an important objective, especially in light of the rise of market power and inequality of market power in recent years. Therefore, we ask whether taxes should reflect the extent of market power, and if so, how?

This paper aims to answer these questions by investigating optimal taxation in conjunction with market power. We set up a model that embeds, in an otherwise canonical Mirrlees (1971) taxation framework, endogenous market power as well as a clear distinction between wage-earning workers and profit-earning entrepreneurs who operate in oligopolistic markets. The novelty in the setup is the interaction of the inefficiency of market power in the hands of entrepreneurs with the unobservable effort supply of both workers and entrepreneurs. This model captures a number of empirically relevant features that link inequality to market power, in particular, how market power creates inequality. As market power increases, labor income decreases while there is an increase in the level and inequality of income of entrepreneurs. This is consistent with the decline in the labor share that has been documented and that coincides with the rise of market power.[1] The rise of market power also results in a decrease in output and social welfare.

Our analysis formulates concrete proposals for policymakers how to deal with market power. The most obvious way to address the distortionary effect of market power is to eradicate the root cause of market power itself with antitrust policies. Because the optimal antitrust policy may not be achievable,[2] we design optimal policy when we can rely on income and commodity taxation only. The Mirrleesian tax provides the correct incentives that trade-off efficient effort supply with inequality. In addition, now the optimal tax system simultaneously corrects the externalities that derive from market power in the goods market. The income tax thus also plays the role of a Pigouvian tax: a tax that corrects a market failure, whether it be pollution or in this case, market power. An important insight of this paper is how to optimally trade off different objectives: inequality and efficiency, while simultaneously correcting market power externalities.

We then perform a detailed quantitative analysis that provides precise policy guidance for policymakers on how to use taxes to deal with the rise of market power between 1980 and 2019. The tangible policy rule results in a *concrete formula* that prescribes a marked decline from 2019 compared to 1980 in labor income tax rate by 10 percentage points, and an increase in the profit tax by 3 percentage points on average (25 percentage points at the top). Because the policymaker trades off efficiency and equity considerations, while the profit tax becomes more progressive in 2019 it is still regressive for large firms. Regressive profit taxes ensure the efficient reallocation of production towards more productive firms.

We link this policy prescription to the theory and decompose the tax formula into four components:[3]

[1] See Karabarbounis and Neiman (2014), De Loecker et al. (2020), and Autor et al. (2020).

[2] Antitrust policy faces many challenges because market power has multiple origins: technological, such as entry barriers, returns to scale, and firm heterogeneity; and market structure such as M&A (Sutton (1991, 2001), De Loecker et al. (2019)).

[3] We allow for nonlinear taxes on the income of workers and entrepreneurs, and linear tax on the sales of consumer goods. However, we can focus attention exclusively on the tax on entrepreneurs and workers, and not on the sales tax. It is well-known

(i) the traditional Mirrleesian part; (ii) the Pigouvian part; (iii) an indirect redistribution effect (IRE); and (iv) a reallocation effect (RE). Without inequality, absent any redistribution motives, marginal tax rates on labor income and profits are equal and negative. The externality from the market power leads to too little output (due to high prices), so the tax formula consists exclusively of the Pigouvian correction of the externality which induces the policymaker to subsidize production, where the Pigouvian tax crucially depends on the *average* markup weighted by factor inputs. Instead, when there is inequality and the policymaker has redistributive concerns, the entrepreneurial income tax wedge is a straightforward combination of a standard Mirrleesian formula and a Pigouvian correction subsidy as long as there is monopolistic competition. Oligopolistic competition within markets introduces a Stiglitz-like motivation for further subsidizing entrepreneurial income and achieving redistribution through price effects (IRE), which depends crucially on the cross-inverse demand elasticity.[4] Finally, when firm-level markups are heterogeneous, the optimal tax formula corrects for the misallocation of production between firms (RE), which depends crucially on the gap between the average markup and firm-level markup. We formulate these theoretical results in a sequence of formal theorems and propositions that derive tax wedges explicitly in terms of primitives.

These four channels make clear how market power affects the optimal tax rate. The net effect on labor income combines the Pigouvian correction of the markup externality with the Mirrleesian tradeoff between production and redistribution. The net effect on the profit tax also depends on the indirect redistribution effect[5] as well as the reallocation effect (which incentivizes the production of high-skill entrepreneurs). This is particularly evident for the top entrepreneurial incomes. The skill gap of entrepreneurs increases in the markup, which modifies the Mirrleesian part: higher markups raise top tax rates.

Finally, we also discuss extensively how our results relate to the existing literature and investigate the robustness of our setup. We analyze three alternative specifications of our baseline model: we introduce non-linear sales taxes, we allow the planner to condition taxes on markups, and we introduce capital investment. This analysis shows that our results are robust to these variations of the model setup. An important new insight from the second robustness exercise is the discovery of a new friction. Even if the planner can condition on markups, the solution is still not the first-best. The reason is that entrepreneurs will adjust their decisions – effort as well as the number of workers hired – in response to the planner's optimal tax schedule. In other words, there is an incentive constraint the planner needs to take into account when solving for the optimal tax rate, even when conditioning on markups.

2 The Model Setup

Environment. The economy is static. Agents belong to one of two occupations $o \in \{e, w\}$, entrepreneur or worker. The occupational types are fixed. The measure of workers is N_w; the measure of entrepreneurs N_e is normalized to one. There is a representative firm producing final goods in a competitive market and making zero profits. Production of the final goods needs the composite inputs of firm-level intermediate goods. Each intermediate good is produced by an entrepreneur (idea), and the effort of workers.

(e.g. Chari and Kehoe (1999) and Golosov et al. (2003)) that multiple tax policies can implement the same second-best. In our setting, we can substitute a linear sales tax with a uniform tax on labor income and profits. Therefore, we assume sales taxes are 0.

[4] The cross-inverse demand elasticity is the elasticity of the price with respect to the outputs of competitors in the same market.

[5] The government can reduce the income gap between entrepreneurs by reducing the price of the products of high-income entrepreneurs. To achieve this, the government should decrease the profit tax rate, which raises the output of the firm.

Within each occupation, agents are heterogeneous in their productivity. Denote the ability of an agent by $\theta_o \in \Theta_o \subset \mathbb{R}_+$, distributed according to the cdf $F_o(\theta_o)$ with density $f_o(\theta_o)$. Set $x_o(\theta_o)$ as the efficiency labor provided by θ_o per unit time. It will be convenient to order ability on the unit interval and consider the uniform distribution of ability, so that $\Theta_o = [0,1]$, $F_o(\theta_o) = \theta_o$ and $f_o(\theta_o) = 1$ (see e.g., Tervio (2008)). Since $x_o(\cdot)$ is free, there is no loss to make the above assumptions about the distribution of ability.

Preferences. Workers and entrepreneurs have preferences for consumption and effort. We consider a quasi-linear utility function.[6] $u_o(c_o, l_o) = c_o - \phi_o(l_o)$ is the utility function of an agent of occupation o, where l_o refers to working hours and c_o is the consumption. $\phi_o(\cdot)$ is twice continuously differentiable and strictly convex. To simplify the analysis, we consider the utility function with constant elasticity of labor supply. Set $\varepsilon_o \equiv \frac{\phi_o'(l_o)}{l_o \phi_o''(l_o)}$. Denote by $V_o(\theta_o)$ the indirect utility of a θ_o agent.

Market. One individual firm is indexed by a triple (i, j, θ_e) with $\theta_e \in \Theta_e$ the ability of entrepreneur, $j \in [0, J(\theta_e)]$ the order of markets, and $i = 1, ..., I$ the order of firm in the most granular market. $I \in \mathbb{N}_+$ is the number of firms in one of the most granular markets. $J(\theta_e) \in \mathbb{R}_+$ is the measure of markets of θ_e entrepreneurs. A differentiated input can also be fully identified by the triple (i, j, θ_e). The differentiated inputs produced by I entrepreneurs of θ_e ability in the same market is used to produce the market-level intermediate goods (θ_e, j). The final good is an aggregation of the intermediate goods across θ_e and j.

The labor and final goods markets are perfectly competitive. Instead, the intermediate goods market exhibits market power, modeled as a variation of the structure in Atkeson and Burstein (2008), but with product differentiation in production rather than in preferences. The most granular market is small, where a finite number (I) of entrepreneurs of equal type produce differentiated inputs for a common intermediate good under Cournot competition. The intermediate goods from these granular markets then compete in a unified market under monopolistic competition. Production of a differentiated input needs the composite of entrepreneurial effort (idea), and the effort of workers.

Technology. An agent of ability θ_o who works l_o hours supplies $x_o(\theta_o) l_o$ units of effective labor factors.[7] The pre-tax labor income $y_w(\theta_w) = x_w(\theta_w) l_w(\theta_w) W$ is the effective labor supply multiplied by the wage rate as in Mirrlees (1971). W is the competitive wage any firm pays for one efficient unit of labor.[8] The firm-level output $Q_{ij}(\theta_e) = Q_{ij}(x_e(\theta_e) l_e(\theta_e), L_w(\theta_e))$ is a function of entrepreneurial effort $l_e(\theta_e)$ and labor inputs $L_w(\theta_e)$. The profit of (i, j, θ_e) firm is given by:

$$y_{e,ij}(\theta_e) = (1 - t_s) P_{ij}\left(Q_{ij}(\theta_e), \{Q_{-ij}(\theta_e)\}_{-i \neq i}, \theta_e\right) \cdot Q_{ij}(\theta_e) - W L_w(\theta_e),$$

where $P_{ij}\left(Q_{ij}, \{Q_{-ij}(\theta_e)\}_{-i \neq i}, \theta_e\right)$ is the inverse demand function of (i, j, θ_e) firm. t_s is a sales tax. $Q_{-ij}(\theta_e)$ referers to the output of $(-i, j, \theta_e)$ firm, which is (i, j, θ_e) firm's competitor in the same granular market.

[6] The quasi-linear utility function eliminates the income effect and the complementarity between consumption and labor. This assumption makes the analysis more tractable and is not crucial to the main economic implication of this paper. See Atkinson and Stiglitz (1976), Mirrlees (1976) and Christiansen (1984) how the omitted elements affect the optimal taxation.

[7] The assumption of efficiency units drastically simplifies the solution of the model but it is not innocuous. The efficiency units assumption rules out sorting because firms are indifferent across worker types as long as they provide exactly the same efficiency units. See amongst others Sattinger (1975a), Sattinger (1993) and Eeckhout and Kircher (2018) how the assumption of efficiency implies an absence of sorting. To date, we know of no way how to solve the optimal taxation problem with market power in the presence of sorting.

[8] Throughout this paper, we assume that labor factors supplied by workers of different abilities are perfectly substitutable. For readers who are interested in imperfectly substitutable labor factors, please refer to Sachs et al. (2020) and Cui et al. (2021).

The above profit income function nests many cases in the literature: 1. In Mirrlees (1971), $y_o(\theta_o) = x_o(\theta_o)\, l_o(\theta_o) P$ is determined by the agent's ability, effort, and the competitive price P; 2. In Stiglitz (1982), $y_o(\theta_o) = x_o(\theta_o)\, l_o(\theta_o) P(\theta_o)$. The competitive price $P(\theta_o)$ is heterogeneous due to the imperfect substitutability of labor. In either case, the entrepreneur treats the prices as given.

The profit income reveals what is new in this paper: 1. Entrepreneurs select output strategically to manipulate the prices of their own products; 2. Entrepreneurs competing in the same market influence each other's product prices. The profit tax affects firm-level profit both through its influence on the firm's behavior and the behavior of the firm's competitors. This setup distinguishes our paper from the classic studies by Mirrlees (1971) and Stiglitz (1982). It also distinguishes our paper from the current literature on optimal taxation with monopolistic competition, where firms have monopoly pricing power in the granular markets (see e.g., Gürer (2021) and Boar and Midrigan (2021)).

Atkeson-Burstein Economy. Our most general conclusions do not depend on the specific functional form of the production technology. For some analytical results, we need to specify the production technology and market structure explicitly. For that purpose we study a variation of Atkeson and Burstein (2008). The labor and final goods markets are perfectly competitive. Instead, the intermediate goods market exhibits market power. The firm-level production technology of the intermediate good is as in Lucas (1978), with one heterogeneous entrepreneur hiring an endogenous number of workers to maximize profits. Because the productivities of entrepreneurs and workers are expressed in efficiency units, the technology takes efficiency units as inputs instead of bodies. The quantity of output of a θ_e entrepreneur is therefore:[9]

$$Q_{ij}(\theta_e) = x_e(\theta_e) l_{e,ij}(\theta_e) \cdot L_{w,ij}(\theta_e)^{\xi}, 0 < \xi \leq 1. \tag{1}$$

Note that because of the efficiency units assumption, output $Q_{ij}(\theta_e)$ does not depend on the worker types θ_w that are employed. There is no capital in our model. Therefore we assume that, as in Lucas (1978) or Prescott and Visscher (1980), the entrepreneur is the residual claimant of output, i.e., they "own" the technology θ_e. Therefore, the entrepreneur hires labor to maximize profits.

Given the technology, we can aggregate I close substitutes (say Coke and Pepsi, or Toyota and Ford) in the same market with elasticity $\eta(\theta_e)$ to $Q_j(\theta_e)$, then across all $J(\theta_e)$ markets with elasticity σ to $Q(\theta_e)$, and finally from a continuum of less substitutable input goods $\{Q(\theta_e)\}_{\theta_e \in \Theta_e}$ (say soft drinks and cars) with the same elasticity σ to the final goods Q:[10]

$$Q_j(\theta_e) = \left[I^{-\frac{1}{\eta(\theta_e)}} \sum_{i=1}^{I} Q_{ij}(\theta_e)^{\frac{\eta(\theta_e)-1}{\eta(\theta_e)}} \right]^{\frac{\eta(\theta_e)}{\eta(\theta_e)-1}}, \tag{2}$$

$$Q(\theta_e) = \left[J(\theta_e)^{-\frac{1}{\sigma}} \int_j Q_j(\theta_e)^{\frac{\sigma-1}{\sigma}} dj \right]^{\frac{\sigma}{\sigma-1}}, \tag{3}$$

$$Q = A \left[\int_{\theta_e} \zeta(\theta_e) Q(\theta_e)^{\frac{\sigma-1}{\sigma}} d\theta_e \right]^{\frac{\sigma}{\sigma-1}}. \tag{4}$$

[9]The case where $\xi = 1$, is common in the literature that models imperfect competition through imperfect substitutes (Melitz (2003), Atkeson and Burstein (2008), De Loecker et al. (2019)). The linear technology simplifies the derivations, and in addition, there is no indeterminacy in the firm size because all goods are imperfect substitutes that determine the boundaries of the firm.

[10]With the Atkeson and Burstein (2008) technology, the elasticities of substitution between intermediate goods of different markets are uniform. In settings with more general technology, the elasticities of substitution may be asymmetric. As elasticities of substitution within markets are asymmetric, the assumption of equal elasticities between markets does not matter for the impact of market power. However, it is a matter of how endogenous wages react to tax policy. Readers interested in this can refer to Rothschild and Scheuer (2013) and Sachs et al. (2020).

The elasticity of substitution σ across markets (between soft drinks and cars) is smaller than within markets (between Coke and Pepsi): $\sigma < \eta(\theta_e)$. In order to rule out abnormal markups, throughout this paper we assume that σ is greater than 1.[11] $\zeta(\theta_e)$ is a distribution parameter. As illustrated by Ales et al. (2015), variations in $\zeta(\theta_e)$ capture the technological or preference-based variations in demand for different skills. Last, to abstract from the love-of-variety effect related to I, we normalize the firm-level and market-level output by $I^{-\frac{1}{\eta(\theta_e)}}$ and $J(\theta_e)^{-\frac{1}{\sigma}}$.[12] We then introduce $\chi(\theta_e) = \zeta(\theta_e) f_e(\theta_e)^{-\frac{1}{\sigma}}$ as a modified distribution parameter. Last, set

$$X_e(\theta_e) = A^{\frac{\sigma-1}{\sigma}} x_e(\theta_e)^{\frac{\sigma-1}{\sigma}} \chi(\theta_e)$$

as the composite ability of θ_e entrepreneur.[13]

Information and Policy. In the tradition of Mirrleesian taxation, we assume types θ_o and effort l_o are not observable, while incomes y_w and y_e are observable. As additional actions are introduced, further clarification of information is desirable. In particular, we assume factor inputs L_w are also observable. Then unobservability of θ_e is equivalent to saying that the firm-level output Q_{ij} and price P_{ij} cannot both be observed. Otherwise, the government can back out the type of entrepreneur by $P_{ij} = P_{ij}\left(Q_{ij}, \{Q_{-ij}\}_{-i \neq i}, \theta_e\right)$. There are several justifications for this information constraint. Effective output Q_{ij}, just like the effort l_o, is difficult to measure objectively. Though working hours can be observed, the effort l_o is difficult to observe because the intensity of working cannot be measured objectively. Similarly, although the quantity of products may be counted, quality and therefore effective output is difficult to measure.[14]

We consider that the government can use profit and labor income taxes $T_e : \mathbb{R}_+ \mapsto \mathbb{R}$ and $T_w : \mathbb{R}_+ \mapsto \mathbb{R}$ to be arbitrarily non-linear in the Mirrlees tradition. These direct taxes together with a sales tax $t_s \in \mathbb{R}$ compose the tax policy system $\mathcal{T} \equiv \{T_e, T_w, t_s\}$ that we consider in our benchmark model. Since the labor and profit income taxes are free, we can normalize the sales tax to zero without loss of generality. The after-tax income of workers and entrepreneurs are $c_w = y_w - T_w(y_w)$ and $c_{e,ij} = y_{e,ij} - T_e(y_{e,ij})$.

In our benchmark model, we restrict the government to levying a linear sales tax, which is typically used in the real economy.[15] The government is therefore solving for the third-best, rather than the second-best allocation, under both the information and policy constraints. In section OC.5.1, we extend the model to allow for nonlinear sales tax and compare the third-best with the second-best allocation.

[11] See equation (20) below for details.

[12] Note that the measure of markets, $J(\theta_e) = \frac{N_e f(\theta_e)}{I}$, is also the measure of varieties provided by θ_e entrepreneurs.

[13] In Online Appendix OA.2.3, we show that the equilibrium allocations are determined by the value of composite ability, while prices depend on the structure of the composite ability. The two θ_e-dependent components ($\chi(\theta_e)$ and $x_e(\theta_e)$) of the composite ability are both determinants of the productivity of a firm. Although the equilibrium allocations are determined by the composite productivity, $\chi(\theta_e)$ and $x_e(\theta_e)$ are not perfect substitutes in the sense that the equilibrium prices depend on the specific values of $\chi(\theta_e)$ and $x_e(\theta_e)$. This is because $\chi(\theta_e)$ directly enters the demand function and interacts with the markup while $x_e(\theta_e)$ does not. Interestingly, what's crucial for pinning down the optimal tax formulas is the value of composite ability (see Proposition 4 for example). For general results, see equation (33) and Theorem 1.

[14] In reality, there are many other reasons why prices and yields are difficult to observe efficiently. For example, there may be collusion between companies and between companies and buyers to hide prices or efficiency to reduce tax burdens.

[15] Alternatively, one can consider a linear tax on labor inputs (such as salary tax). Both the linear sales tax and linear tax on the salary pay act as tax wedges between the marginal cost and income of labor inputs $L_{w,ij}$. Since the prediction of optimal taxation is about tax wedges while not about specific tax policies (e.g., see Chari and Kehoe (1999); Golosov et al. (2003); Salanié (2003), pages 64-66), there is no need to introduce both of these indirect taxes. To see this, consider equation (9) below, where if we levy an additional tax t_l on the labor inputs of the firm, the ratio of the marginal income of $L_{w,ij}$ to the marginal cost of $L_{w,ij}$ is $\frac{1+t_l}{1-t_s}$, which means the role of t_l as a tax wedge can be replaced by t_s.

Planner's Objective. The government chooses tax policies to maximize social welfare:

$$\sum_{o\in\{w,e\}} N_o \int_{\theta^o} G\left(V_o(\theta_o)\right) \widetilde{f}_o\left(\theta_o\right) d\theta_o, \tag{5}$$

subject to the budget constraint

$$N_e \int_{\theta_e} T_e\left(y_e\left(\theta_e\right)\right) f_e\left(\theta_e\right) d\theta_e + N_w \int_{\theta_w} T_w\left(y_w\left(\theta_w\right)\right) f_w\left(\theta_w\right) d\theta_w = R$$

and agents' responses to the taxes. $R \in \mathbb{R}_+$ is an exogenous tax revenue. The social welfare function $G : \mathbb{R}_+ \longmapsto \mathbb{R}_+$ is twice differentiable and concave. The PDF $\widetilde{f}_\theta\left(\cdot\right)$ is a Pareto weights schedule, which is assumed to be continuous.[16]

Equilibrium. We formally define equilibrium below once we have solved for the best responses of all agents. We now give an informal definition of equilibrium. Given the tax regime $\mathcal{T}$, a Cournot competitive tax equilibrium allocation and price system are such that the resulting allocation maximizes the final good producer's profit, and maximizes the entrepreneur's and worker's utility subject to the budget constraint. In addition, all markets are clear under the price system, and the government's budget constraint is satisfied, which, given other budget constraints, is equivalent to saying that the social resource constraint is satisfied.

3 Solution

We solve the general technology first (Section 3.1); then for the Atkeson-Burstein technology (Section 3.2).

3.1 The Cournot Competitive Tax Equilibrium

Final Goods Market Solution. We start with the final goods market where we normalize the price of the final good to one. The final good producer chooses the inputs of the intermediate goods to maximize its profit. The demand for the intermediate input solves:

$$\Pi = \max_{Q^D_{ij}(\theta_e)} Q - \int_{\theta_e} \int_j \left[\sum\nolimits_i Q^D_{ij}\left(\theta_e\right) P_{ij}\left(\theta_e\right)\right] dj d\theta_e, \tag{6}$$

where $Q^D_{ij}\left(\theta_e\right)$ is the quantity demanded from firm (i, j, θ_e).

Entrepreneur's Solution. In our benchmark model, we consider the Cournot Competitive Tax Equilibrium in intermediate goods market j between I firms. Because there are a continuum of intermediate good markets j and θ_e, there is only strategic interaction within a market j and all firms treat the output decisions in other intermediate goods markets as given.

All firms treat others' outputs as given. The problem of the entrepreneur in (i, j, θ_e) firm is:

$$V_{e,ij}\left(\theta_e\right) \equiv \max_{l_{e,ij}, L_{w,ij}} c_e - \phi_e\left(l_e\right) \tag{7}$$

$$\text{s.t. } c_{e,ij} = y_{e,ij} - T_e\left(y_{e,ij}\right) \tag{8}$$

$$y_{e,ij} = \left(1-t_s\right) P_{ij}\left(Q_{ij}, \left\{Q_{-ij}\left(\theta_e\right)\right\}_{-i\neq i}, \theta_e\right) Q_{ij} - WL_{w,ij}, \tag{9}$$

where $Q_{ij} = Q_{ij}\left(x_e(\theta_e) l_e, L_w\right)$ is the quantity supplied of the intermediate good. We denote $P_{ij}(\theta_e)$, $L_{w,ij}(\theta_e)$, $c_{e,ij}(\theta_e)$, $y_{e,ij}(\theta_e)$ and $l_{e,ij}(\theta_e)$ as the price of intermediate goods, labor inputs, consumption, profit and effort of the (i, j, θ_e) entrepreneur.

[16] The use of general Pareto weights in the optimal tax literature goes all the way back to Diamond and Mirrlees (1971a) and Diamond and Mirrlees (1971b), and in the context of our model to Scheuer (2014).

Worker's Solution. Type θ_w workers choose labor supply and consumption to maximize their utility, given the wage rate W:

$$V_w(\theta_w) \equiv \max_{l_w} c_w - \phi_w(l_w) \tag{10}$$

$$\text{s.t. } c_w = W x_w(\theta_w) l_w - T_w(W x_w(\theta_w) l_w). \tag{11}$$

For later use, denote $c_w(\theta_w)$, $y_w(\theta_w)$ and $l_w(\theta_w)$ as the consumption, income and effort of a θ_w worker.

Market Clearing. Commodity and labor markets clearing require that the quantity demanded in the output sector $Q_{ij}^D(\theta_e)$ from equation (6) equals the quantity supplied $Q_{ij}^S(\theta_e)$ from equation (7):

$$Q_{ij}^D(\theta_e) = Q_{ij}^S(\theta_e) \tag{12}$$

$$\text{and} \quad Q = \int_{\theta_w} c_w(\theta_w) f_w(\theta_w) d\theta_w + \int_{\theta_e}\int_j \left[\sum\nolimits_i c_{e,ij}(\theta_e)\right] dj d\theta_e + R, \tag{13}$$

$$\text{and} \quad \int_{\theta_w} x_w(\theta_w) l_w(\theta_w) f_w(\theta_w) d\theta_w = \int_{\theta_e}\int_j \left[\sum\nolimits_i L_{w,ij}(\theta_e)\right] dj d\theta_e, \tag{14}$$

where R is the exogenous government revenue.

Solving individuals' and final good producers' problems gives the following equilibrium conditions:

$$P_{ij}(\theta_e) = \frac{\partial Q}{\partial Q_{ij}(\theta_e)}, \tag{15}$$

$$W = (1-t_s)\frac{\partial\left[P_{ij}(\theta_e) Q_{ij}(\theta_e)\right]}{\partial L_{w,ij}(\theta_e)}, \tag{16}$$

$$\phi_w'(l_w(\theta_w)) = W x_w(\theta_w)\left[1 - T_w'(W x_w(\theta_w) l_w(\theta_w))\right], \tag{17}$$

$$\phi_e'\left(l_{e,ij}(\theta_e)\right) = (1-t_s)\frac{\partial\left[P_{ij}(\theta_e) Q_{ij}(\theta_e)\right]}{\partial Q_{ij}(\theta_e)}\frac{\partial Q_{ij}(\theta_e)}{\partial l_{e,ij}(\theta_e)}\left[1 - T_e'\left(y_{e,ij}(\theta_e)\right)\right]. \tag{18}$$

When first-order conditions are both necessary and sufficient to individuals' and final good producer's problems, the equilibrium allocations are determined by (12) to (18) and individuals' budget constraints.

Equilibrium. Throughout this paper, we will consider the following symmetric Cournot competitive tax equilibrium. We refer to the allocation set $\mathcal{A} = \{L_w, l_w, l_e, c_w, c_e\}$ as a combination of consumption schedules $c_o : \Theta_o \longmapsto \mathbb{R}_+$, labor supply schedules $l_o : \Theta_o \longmapsto \mathbb{R}_+$ and labor demand schedule $L_w : \Theta_w \to \mathbb{R}_+$ which are independent on (i, j). Prices $\mathcal{P} = \{P, W\}$ in the equilibrium is a combination of wage rate W and price schedule $P : \Theta_e \longmapsto \mathbb{R}_+$ that independent on (i, j). Formally, we consider the following symmetric Cournot tax equilibrium:

Definition 1 *A Symmetric Cournot Competitive Tax Equilibrium (SCCTE) is a combination of a tax system $\mathcal{T}$, a symmetric allocation $\mathcal{A}$, and a symmetric price system $\mathcal{P}$, such that given the policy and price system, the resulting allocation maximizes the final good producer's profit (6); maximizes the entrepreneurs' and workers' utilities (7) and (10) subject to the budget constraints (8) and (11); the price system satisfies (16) and (15); and labor and commodity markets are cleared, i.e., (12) to (14) are satisfied.*

Note that because of Walras's law, we do not need to impose the government's budget constraint in our definition of SCCTE. Given the agent's budget constraints and market clear conditions, the government's budget constraint must be satisfied.

We now make some common restrictions on the equilibrium that we consider throughout the paper. First, we assume that the mechanisms (tax policies) are sufficiently differentiable and first-order conditions are both necessary and sufficient to the agents' problems. This is a common assumption in the optimal tax literature, which is equivalent to assuming that the optimal tax schedules are twice continuously differentiable and not too regressive (see e.g., Jacquet et al. (2013)). We will show the sufficiency of first-order conditions under our leading production function (see section 4). Second, we assume that:

Assumption 1 *In the symmetric Cournot competitive tax equilibrium:*

(i) $y_w(\theta_w)$ *is differentiable, strictly positive, and strictly increasing in* $\theta_w \in \Theta_w$*;*

(ii) $y_e(\theta_e)$ *is differentiable, strictly positive, and strictly increasing in* $\theta_e \in \Theta_e$*.*

The Spence-Mirrlees condition implies non-decreasing labor income in wages.[17] For simplicity, we assume that $y_w(\theta_w)$ is strictly increasing in θ_w, which in turn implies $x'_w(\cdot) > 0$. With Assumption 1, we can define $F_{y_o}(y_o(\theta_o)) = F_o(\theta_o)$ and $f_{y_o}(y_o(\theta_o)) = F'_{y_o}(y_o(\theta_o))$ as the CDF and PDF of entrepreneurial and labor income. Besides, Assumption 1 excludes cases with mass points.

Notation. In what follows, where there is no confusion, we will drop the subscript ij. For example, in the symmetric equilibrium, the markup is the same for all entrepreneurs with equal types. Therefore, we often denote the markup $\mu_{ij}(\theta_e)$ by $\mu(\theta_e)$, and the labor demand $L_{w,ij}(\theta_e)$ by $L_w(\theta_e)$. The inverse demand function, taking into account strategic interaction between firm i and firms $-i$, simplifies to $P_{ij}\left(Q_{ij}, Q_{-ij}(\theta_e), \theta_e\right)$ as the competitors produce the same amount of products.

Price Elasticity. Define the own and cross-inverse demand elasticity as:

$$\varepsilon^{P,own}_{Q_{ij}}(\theta_e) \equiv \frac{\partial \ln P_{ij}\left(Q_{ij}, Q_{ij}(\theta_e), \theta_e\right)}{\partial \ln Q_{ij}}|_{Q_{ij}=Q_{ij}(\theta_e)} \text{ and } \varepsilon^{P,cross}_{Q_{-ij}}(\theta_e) \equiv \frac{\partial \ln P_{ij}\left(Q_{ij}, Q_{ij}(\theta_e), \theta_e\right)}{\partial \ln Q_{ij}(\theta_e)}|_{Q_{ij}=Q_{ij}(\theta_e)}.$$

The cross-inverse demand elasticity is the elasticity of an entrepreneur's inverse demand function with respect to her competitor's output. The appearance of cross-inverse demand elasticity highlights the difference between Cournot competition and Monopoly. With strategic interaction between competitors in a submarket, the demand of a firm depends on its competitors' outputs.

Markups. Following the literature, we define the markup as the ratio of price to marginal cost

$$\mu_{ij}(\theta_e) \equiv \frac{P_{ij}(\theta_e)}{MC_{ij}(\theta_e)} = \frac{P_{ij}(\theta_e)}{\frac{W}{\frac{\partial Q_{ij}(\theta_e)}{\partial L_{w,ij}(\theta_e)}(1-t_s)}}. \tag{19}$$

The firm's FOC delivers a relationship, known as the Lerner Rule, between the (own) inverse-demand elasticity $\varepsilon^{P,own}_{Q_{ij}}(\theta_e)$ and markups $\mu_{ij}(\theta_e)$.[18] The markup is thus related to the demand elasticity:

$$\mu_{ij}(\theta_e) = \frac{1}{1 + \varepsilon^{P,own}_{Q_{ij}}(\theta_e)}. \tag{20}$$

[17] See e.g., see Salanié (2003), p. 87.

[18] This follows from profit maximization, in equation (16), which implies $W = P_{ij}\left(Q_{ij}(\theta_e), \theta_e\right)[1 + \gamma(\theta_e)] \frac{\partial Q_{ij}(\theta_e)}{\partial L_{w,ij}(\theta_e)}(1 - t_s)$.

The higher the demand elasticity (the lower the inverse demand elasticity), the higher the markup. In our results under the SCCTE, we define the economy-wide aggregate markup as:

$$\mu \equiv \int_{\theta_e} \mu\left(\theta_e\right) \omega\left(\theta_e\right) d\theta_e, \text{ where } \omega\left(\theta_e\right) \equiv \frac{L_w\left(\theta_e\right) f_e\left(\theta_e\right)}{\int_{\theta_e} L_w\left(\theta_e\right) f_e\left(\theta_e\right) d\theta_e}. \tag{21}$$

In the Atkeson-Burstein economy, $\varepsilon_{Q_{-ij}}^{P,cross}\left(\theta_e\right) = -\frac{1}{\sigma} - \varepsilon_{Q_{ij}}^{P,own}\left(\theta_e\right)$ and $\varepsilon_{Q_{ij}}^{P,own}\left(\theta_e\right) = \frac{1}{\mu_{ij}(\theta_e)} - 1$.[19]

3.2 The Atkeson-Burstein Economy

Before solving for the optimal tax policy, we examine the factors that determine the impact of taxation on markup in the Atkeson-Burstein economy. In section C.2, we consider another important special case, i.e., monopolistic competition with non-constant elasticities of substitution (Kimball aggregator).

Markups. As in Atkeson and Burstein (2008), under the above technology with nested CES preferences, the inverse-demand elasticity can be written in weighted form:[20]

$$\varepsilon_{Q_{ij}}^{P,own}\left(\theta_e\right) = -\left[\frac{1}{\eta\left(\theta_e\right)}\left(1 - s_{ij}\right) + \frac{1}{\sigma}s_{ij}\right] \geq -\frac{1}{\sigma}, \tag{22}$$

where s_{ij} is the sales share of firm i in market j. Therefore, the markup depends on the weighted sum of the elasticity of substitution between intermediate goods, and the intensity of competition in the submarket.

$$\mu_{ij}(\theta_e) = \frac{1}{1 + \varepsilon_{Q_{ij}}^{P,own}\left(\theta_e\right)} \leq \frac{\sigma}{\sigma - 1}. \tag{23}$$

The lower the $\eta(\theta_e)$ and σ, the less substitutable the goods are within and between markets, and the higher the markup. Most crucially, the markup increases as the sales share s_{ij}, and hence the number of competitors I, decreases in the market. The smaller the number of competitors I, the smaller the weight on the within market elasticity higher the weight on $\frac{1}{\eta(\theta_e)}$ and the higher the weight on $\frac{1}{\sigma}$. Firms that face little competition face little substitution and hence markups.[21]

Labor Share. In our model, the firm's labor share is simply the ratio of the firm's total wage bill to its revenue. In the absence of capital, the residual therefore is the income of the entrepreneur, i.e., the profit share. Denote by $\nu_{ij}(\theta_e)$ the labor share which can be defined as

$$\nu_{ij}(\theta_e) \equiv \frac{W L_{w,ij}(\theta_e)}{P_{ij}\left(\theta_e\right) Q_{ij}\left(\theta_e\right)\left(1 - t_s\right)}.$$

While superficially this expression hints at an apparent positive relation between the sales tax rate t_s (an increase in t_s increases the labor share), taxes also affect the other variables such $L_{w,ij}, P_{ij}$ and Q_{ij}, all of which are endogenous. When we use the firm's first-order condition, we can rewrite the labor share as

$$\nu_{ij}(\theta_e) = \frac{\xi}{\mu_{ij}(\theta_e)},$$

[19] See Appendix A.2 for details.

[20] See Appendix A.2.2 for details.

[21] We can derive the equivalent inverse demand elasticity under Bertrand competition which is different from the residual demand elasticity under Cournot: $\varepsilon_{Q_{ij}}^{P,own}(\theta_e) = -\left[\left(1 - s_{ij}\right)\eta\left(\theta_e\right) + s_{ij}\sigma\right]^{-1}$. In fact, all our results go through under Bertrand and are similar to Cournot once we adjust equation (22).

which shows a negative relation between the firm's labor share and its markup. Denote the aggregate labor share by

$$\nu \equiv \frac{W \int x_w(\theta_w) \, l_w(\theta_w) \, f_w(\theta_w) \, d\theta_w}{Q}.$$

Although the firm-level labor share is exogenous to the tax rate, the aggregate labor share is does depend on the tax rate. Proposition 1 summarizes the properties of the equilibrium labor share.

Proposition 1 *In the Atkeson-Burstein economy:*

(i) The firm labor share $\nu_{ij}(\theta_e)$ is independent of taxes and is decreasing in the markup $\mu_{ij}(\theta_e)$;

(ii) In the Laissez-faire economy,[22] the aggregate labor share ν decreases in market power, i.e., increase in I, when

$$\frac{1}{1+\varepsilon_e} + \frac{1}{\sigma - 1} > \xi. \tag{24}$$

Proof. See Online Appendix OA.3. ■

Part (i) of Proposition 1 already hints at the fact that taxes cannot "solve" the effect that market power has on both efficiency and inequality. To achieve the first-best, which we define below, the planner needs to tackle the problem at its root cause, either through antitrust enforcement or regulation of firms and industries. The objective of this paper is to show that optimal taxation can nonetheless restore efficiency and equality. Most importantly, we show that the optimal policy varies with market power and how.

This result also confirms a well-known theoretical property, namely that firms with higher individual markups have a lower labor share. This result is an immediate consequence of the firm's first-order condition. Higher markups mean that the firm sells and produces fewer units, even though sales are higher. Therefore, the firm needs fewer labor inputs, and the labor share falls. De Loecker et al. (2020) and Autor et al. (2020) show that the negative relation at the firm level between markups and the labor share is borne out in the data.

Part (ii) of Proposition 1 is strong in the sense that it is not dependent on the assumptions on $\eta(\theta_e)$. The parameter restriction (24) has intuitive economic interpretations. It ensures that with the increase of I the entrepreneurial effort of small firm increases as relative to the large firm, so that the market is not overly concentrated to cause an incerease of average markup. Besides, the condition guarantees that the entrepreneurial effort decreases in markup (increases in I).

The Laissez-faire Economy. We further analyze the properties of the model economy that we just laid out without government intervention: the government revenue R is zero and no taxes are levied. We ask what the effect is of market power on the equilibrium allocation. This serves as a benchmark to understand the workings of the model before we introduce the role of optimal taxation. In the Laissez-faire economy, we consider the comparative statics effect of a rise in the markup. We consider an increase in markups economy-wide by changing the number of competing firms I in all markets simultaneously. This comparative statics effect economy-wide affects individual firm outcomes, as well as aggregates. We summarize the results in the following proposition.

[22] See the following section 3.2 for details about the Laissez-faire economy.

Proposition 2 *Consider the Atkeson-Burstein economy with constant $\eta(\theta_e)$. Let conditions (24) and (25) hold:*

$$\frac{1+\varepsilon_w}{\varepsilon_w} - \xi\left(1+\varepsilon_e\right) > 0. \tag{25}$$

With the decrese of I, we have the following results:

(i) At the individual level, the labor share $\nu_{ij}(\theta_e)$, the quantity $Q_{ij}(\theta_e)$, sales $P_{ij}(\theta_e)Q_{ij}(\theta_e)$, entrepreneurial effort $l_{e,ij}(\theta_e)$, worker effort $l_w(\theta_w)$, income $y_w(\theta_w)$ and utility $V_w(\theta_w)$ decrease; The price $P_{ij}(\theta_e)$ remains unchanged; The effects on entrepreneur utility $V_{ij,e}(\theta_e)$ and entrepreneur profits $y_{e,ij}(\theta_e)$ are ambiguous;

(ii) At the aggregate level, the wage rate W, the aggregate labor share ν and output Q decrease. The effects on aggregate entrepreneur profits is ambiguous.

(iii) Individual and aggregate entrepreneur profits increase if and only if $\mu \leq \frac{\xi}{\frac{\varepsilon_e}{1+\varepsilon_e}+\frac{\varepsilon_w}{\varepsilon_w+1}\xi}$, and individual and aggregate entrepreneur utility increase if and only if $\mu \leq \frac{\xi+\frac{\varepsilon_e}{1+\varepsilon_e}}{\frac{\varepsilon_e}{1+\varepsilon_e}+\frac{\varepsilon_w}{1+\varepsilon_w}\xi}$.

Proof. See Online Appendix OA.4. ■

Condition (25) guarantees that the demand elasticity of $L_{w,ij}$ is smaller than the supply elasticity of l_w, so that the equilibrium wage increases in TFP (see e.g., (OA15)) and the labor demand decreases in markup. The two restrictions on the parameters are weak and are generally satisfied for the range of parameter values used in the quantitative literature.

Overall, the effect of the rise of market power is negative for workers and positive for entrepreneurs: market power lowers the income and the utility of workers and it increases the profits and the utility of entrepreneurs. In addition, the rise of market power has a negative impact on the aggregate economy: the wage rate declines, and aggregate output, sales, and labor share decline.

The conditions for increasing profits and increasing utility posited in the Proposition are satisfied for typical values used in quantitative studies. For example, with $\varepsilon_e = \varepsilon_w = 0.33$ and $\xi = 0.5$, the condition for increasing profits is satisfied for markup $\mu < 1.33$ and the second condition for increasing utility is satisfied for $\mu < 2$. When ξ is increased to 0.6, the condition is even looser, with the first and second conditions changed to $\mu < 1.5$ and $\mu < 2.125$, respectively.

4 The Planner's Problem

The planner's problem can be treated ins different ways. In the heuristic argument that follows, the planner adopts truth-telling mechanisms $\{c_w(\theta_w), y_w(\theta_w)\}_{\theta_w\in\Theta_w}$ and $\{c_e(\theta_e), y_e(\theta_e)\}_{\theta_e\in\Theta_e}$ to implement allocation rules that maximize social welfare subjecting to the information, resource, and policy constraints.

Specifically, the planner asks each of the entrepreneurs and workers to report their types and assigns a reward contingent on the announced type. A worker who reports θ'_w obtains $y_w(\theta'_w)$ in labor income, which results in $c_w(\theta'_w)$ of after-tax income. Similarly, an entrepreneur who reports θ'_e obtains $y_e(\theta'_e)$ in profit and $c_e(\theta'_e)$ in after-tax profit. The incentive-compatible conditions for workers are standard and can be found in any Mirrlessian tax literature. Additional clarification is required for the entrepreneur's incentive-compatible condition.

4.1 The First-order Approach

We follow the literature by applying the first-order approach (FOA) to simplify the incentive constraints and solve the planner's problem. In section OB.1, we discuss the validity of the FOA in our setting.

Worker. Workers report their types to maximize their gross utility $V_w(\theta_w)$. The worker's first-order incentive condition is similar to that in the literature:

$$V_w'(\theta_w) = l_w(\theta_w)\phi_w'\left(l_w(\theta_w)\right)\frac{x_w'\left(\theta_w\right)}{x_w\left(\theta_w\right)}, \forall\theta_w \in \Theta_w, \tag{26}$$

where $V_w\left(\theta_w\right) = c_w\left(\theta_w\right) - \phi_w\left(\frac{y_w(\theta_w)}{x_w(\theta_w)W}\right)$. When the Spence-Mirrlees condition is satisfied, the first-order incentive condition is not only necessary but also sufficient (see e.g., Salanié (2003) p.88-90).

Entrepreneur. In the subsequent analysis, we will focus on the entrepreneur's first-order incentive condition in the Nash equilibrium. That is, given that all other competitors are telling the truth, an entrepreneur also finds it optimal to report the truth. Formally, for any $j \in [0, J(\theta)]$ and $i = 1, ..., I$, truth-telling requires that entrepreneurs with type θ_e report a type θ_e' to maximize their gross utility $V_{e,ij}(\theta_e'|\theta_e;\theta_e)$.

As Cournot competitive equilibrium is a special case of Nash equilibrium, the above incentive condition is consistent with the SCCTE considered in this paper. Later, we will show that when the above incentive conditions together with the resource and policy constraints are satisfied, there exist a SCCTE, where the optimal allocation can be implemented by the policy system considered in our model. Under the above truth-telling condition and SCCTE, we can simplify $V_{e,ij}(\theta_e'|\theta_e;\theta_e)$ as $V_e(\theta_e'|\theta_e)$.

An entrepreneur's problem is to choose a reporting type θ_e' to maximize :

$$V_e\left(\theta_e\right) = \max_{\theta'\in\Theta_e} V_e(\theta_e'|\theta_e), \tag{27}$$

where $V_e(\theta_e'|\theta_e) = c_e\left(\theta_e'\right) - \phi_e\left(l_e\left(\theta_e'|\theta_e\right)\right)$ is the utility of the θ_e entrepreneur who reports θ_e' and $l_e\left(\theta_e'|\theta_e\right)$ is the entrepreneurial labor supply needed to finish the task:

$$\begin{aligned} l_e\left(\theta_e'|\theta_e\right) &= \min_{L_w, l_e}\ l_e \\ \text{s.t. } \left(1-t_s\right)P_{ij}\left(Q_{ij}, Q_{-ij}\left(\theta_e\right), \theta_e\right)Q_{ij} - WL_w &= y_e\left(\theta_e'\right). \end{aligned} \tag{28}$$

In what follows, we denote by $L_w\left(\theta_e'|\theta_e\right)$ the solution of the above problem, and let

$$Q_{ij}\left(\theta_e'|\theta_e\right) = Q_{ij}\left(x_e\left(\theta_e\right)l_e\left(\theta_e'|\theta_e\right), L_w\left(\theta_e'|\theta_e\right)\right);$$

$$P_{ij}\left(\theta_e'|\theta_e\right) = P_{ij}\left(Q_{ij}\left(\theta_e'|\theta_e\right), Q_{-ij}\left(\theta_e\right), \theta_e\right).$$

In Online Appendix OB.2, we show when the first-order conditions of the above problem is not only necessary but also sufficient. In the following analysis, we always assume that the first-order conditions of the above problem is both necessary and sufficient, where $l_e\left(\cdot|\cdot\right) : \Theta_e^2 \mapsto \mathbb{R}_+$ is differentiable. The first-order necessary incentive condition requires $\frac{\partial V_e(\theta_e'|\theta_e)}{\partial\theta_e'}|_{\theta_e'=\theta_e} = 0$. From this, Lemma 1 follows:

Lemma 1 *Under Assumption 1, the first-order necessary incentive condition is equivalent to:*

$$V_e'(\theta_e) = \phi_e'\left(l_e\left(\theta_e\right)\right) l_e\left(\theta_e\right) \left[\mu\left(\theta_e\right) \frac{d \ln P_{ij}\left(Q_{ij}, Q_{ij}\left(\theta_e\right), \theta_e\right)}{d\theta_e}|_{Q_{ij}=Q_{ij}(\theta_e)} + \frac{x_e'\left(\theta_e\right)}{x_e\left(\theta_e\right)}\right], \quad \forall \theta_e \in \Theta_e. \tag{29}$$

Proof. See Appendix B.1. ■

Lemma 1 is useful because it demonstrates that the incentive-compatible constraint of the entrepreneur boils down to condition (29). Two things worth noting here: First, the incentive condition depends on $\frac{d \ln P_{ij}\left(Q_{ij}, Q_{ij}(\theta_e), \theta_e\right)}{d\theta_e}|_{Q_{ij}=Q_{ij}(\theta_e)}$ instead of $\frac{d \ln P(\theta_e)}{d\theta_e}$. As the entrepreneurs can change the price of their own products by changing the firm-level output Q_{ij}, the tax has no first-order effect on the relative price through its effect on a firm's own output Q_{ij}. Therefore, the traditional indirect redistribution route is closed. Second, the tax exhibit an indirect redistribution effect through its influence on Q_{-ij}. The equation

$$\frac{d \ln P_{ij}\left(Q_{ij}, Q_{ij}\left(\theta_e\right), \theta_e\right)}{d\theta_e}|_{Q_{ij}=Q_{ij}(\theta_e)} = \frac{\partial \ln P_{ij}\left(Q_{ij}\left(\theta_e\right), Q_{ij}\left(\theta_e\right), \theta_e\right)}{\partial \theta_e} + \varepsilon^{P,cross}_{Q_{-ij}}\left(\theta_e\right) \frac{d \ln Q_{ij}\left(\theta_e\right)}{d\theta_e}$$

highlights the difference between Cournot competition and Monopoly. With strategic interaction between competitors in a submarket, the inverse demand of the firm depends on the competitors' outputs, which introduces a novel indirect redistribution route.

4.2 Implementability

In our benchmark, we consider the planner's third-best solution, where the sales tax is restricted to be uniform. Essentially, this constraint requires that the tax wedges between the marginal cost and benefit of labor inputs $L_w\left(\theta_e\right)$ must be uniform across the firms. In this Section, we show how the constrained optimal allocation can be implemented by the tax system $\mathcal{T}$. Besides, we demonstrate that t_s is redundant in implementing the constrained optimal allocation, so that we can focus on the tax system with $t_s = 0$.

Lemma 2 and 3 in Online Appendix OB.1 establishes a relaxed planner's problem, where the original incentive conditions are replaced by first-order incentive conditions. To this end, the planner now chooses the variables $\{c_w(\theta_w), l_w\left(\theta_w\right), c_e(\theta_e), L_w(\theta_e), l_e\left(\theta_e\right)\}_{\theta_e \in \Theta_e, \theta_w \in \Theta_w}$ to maximize the social welfare (5), subject to the incentive conditions (26) and (29), where the inverse demand function satisfies (15); the market clear conditions (12) to (14); and condition (16) with $t_s = 0$. Condition (16) can be treated as a policy constraint in the planner's problem. Essentially, it requires that the marginal revenue of labor inputs:

$$\varpi\left(\theta_e\right) \equiv \frac{\partial\left[P_{ij}\left(\theta_e\right) Q_{ij}\left(\theta_e\right)\right]}{\partial L_{w,ij}\left(\theta_e\right)} \tag{30}$$

to be equal for firms, $\theta_e \in \Theta_e$. The uniform-sales-tax policy constraint thus can be rewritten as $\frac{d \ln \varpi(\theta_e)}{d\theta_e} = 0$. Last, it turns out to be easier if we take $V_o\left(\theta_o\right)$ instead of $c_o(\theta_o)$ as the variable of the planner's problem. After solving for the planner's problem, we can construct the price system and tax system by the FOCs.

4.3 Useful Concepts

Tax Wedges Marginal distortions of taxes in agents' choices can be described with wedges. Entrepreneurs have three possible choices (consumption, effort, and hiring workers), while workers have two possible choices (consumption and working hours). In total, there are three tax wedges: (i) the tax wedge $\tau_s\left(\theta_e\right)$

between the marginal cost and marginal income of labor inputs $L_w(\theta_e)$, (ii) the tax wedge $\tau_w(\theta_w)$ between the marginal disutility and income of the labor supply l_w, and (iii) the tax wedge $\tau_e(\theta_e)$ between the marginal disutility and income of the entrepreneur's labor supply l_e. Specifically, we shall define the three types of tax wedges as:

$$\tau_s(\theta_e) = 1 - \frac{W}{\frac{\partial[P_{ij}(\theta_e)Q_{ij}(\theta_e)]}{\partial L_{w,ij}(\theta_e)}}, \tau_w(\theta_w) = 1 - \frac{\phi_w'(l_w(\theta_w))}{Wx_w(\theta_w)}, \tau_e(\theta_e) = 1 - \frac{\phi_e'(l_e(\theta_e))}{\frac{\partial[P_{ij}(\theta_e)Q_{ij}(\theta_e)]}{\partial l_{e,ij}(\theta_e)}}. \tag{31}$$

They are determined by the taxes. From the FOCs (15) - (18), we obtain $\tau_s(\cdot) = t_s$, $\tau_w(\theta_w) = T_w'(y_w(\theta_w))$ and $\tau_e(\theta_e) = 1 - (1 - t_s)[1 - T_e'(y_e(\theta_e))]$. Note that due to the policy constraint (the government cannot levy firm-specific or non-linear sales tax), $\tau_s(\theta_e)$ in our benchmark model is restricted to be uniform. In the model extension, we will loosen the policy constraint to provide a second-best optimum as a comparison.

As is known from the optimal tax literature, generally there are multiple tax systems that can implement the second-best allocation (see e.g., Chari and Kehoe (1999); Golosov et al. (2003)). In our model, as long as $\tau_s(\theta_e)$ is restricted to be uniform and income taxes are free, there is no need to enforce the sales taxes in addition to the direct taxes. Also note that, as long as $\tau_s(\cdot)$ is uniform, one can replace the sales tax by a consumptions tax, which suggests that the uniform sales tax should not be treated as a tax beared by the firm, but a uniform tax on the labor factors and entrepreneurial efforts. Without loss of generality, we set t_s at zero. Then $\tau_w(\cdot)$ and $\tau_e(\cdot)$ capture the effective tax rates on $l_w(\cdot)$ and $l_e(\cdot)$, respectively.

Income Elasticity. Set $\varepsilon_{1-\tau_w}^{y_w}(\theta_e) = \frac{1}{1+1/\varepsilon_w}\frac{d\ln y_w(\theta)}{d\ln x_w(\theta_w)}$ as the non-linear elasticity of labor income, which is familar from the optimal income tax literature (see e.g., Sachs et al. (2020)). As to the elasticity of profit, consider a marginal increase (i.e., $d\tau$) of the marginal tax rate faced by the θ_e-type entrepreneur. The tax reform has no first-order effects on the aggregate values. As in Scheuer and Werning (2017), the elasticity derived here is a micro elasticity. The entrepreneur treats the others' action as given, and react to the tax reform rationally. Define $\varepsilon_{1-\tau_e}^{y_e}(\theta_e) \equiv -\frac{dy_e(\theta_e)}{y_e(\theta_e)} / \frac{d\tau}{1-T_e'(y_e(\theta_e))}$ as the non-linear profit elasticity. We have the following result:

$$\varepsilon_{1-\tau_e}^{y_e}(\theta_e) = \frac{1}{\frac{1+\varepsilon_e}{\varepsilon_e}[\mu(\theta_e) - \xi] - \left[1 - \frac{y_e(\theta_e)T_e''(y_e(\theta_e))}{1-T_e'(y_e(\theta_e))}\right]}.$$

See Appendix A.2.2 for the details about the elasticity of profit.

Use the supscript o to denote variables in the Laissez-faire economy. Denote by $y_e^o : \Theta_e \mapsto \mathbb{R}_+$ the profit in the Laissez-faire economy. Denote by

$$\varepsilon_{1-\tau_e}^{y_e^o}(\theta_e) = \frac{1}{\frac{1+\varepsilon_e}{\varepsilon_e}[\mu(\theta_e) - \xi] - 1} \tag{32}$$

the elasticity of profit in the Laissez-faire economy. $\varepsilon_{1-\tau_e}^{y_e^o}(\theta_e)$ can be observed as long as $\varepsilon_{1-\tau_e}^{y_e}(\theta_e)$ and the progressivity of profit tax $\frac{y_e(\theta_e)T_e''(y_e(\theta_e))}{1-T_e'(y_e(\theta_e))}$ are observable.

Hazard Ratio. Denote by $F_{y_e}^o : \mathbb{R}_+ \mapsto \mathbb{R}_+$ the CDF of profit in the Laissez-faire economy. $f_{y_e}^o(y_e^o) = F_{y_e}^{o\prime}(y_e^o)$ is the PDF. Set

$$H(\theta_e) \equiv \frac{1 - F_{y_e}^o(y_e^o(\theta_e))}{f_{y_e}^o(y_e^o(\theta_e))} = \frac{d\ln y_e^o(\theta_e)}{d\theta_e}\frac{1 - F_e(\theta_e)}{f_e(\theta_e)}$$

as the hazard ratio of profit in the Laissez-faire economy. The real hazard ratio $\frac{1-F_{y_e}(y_e(\theta_e))}{f_{y_e}(y_e(\theta_e))}$ is determined by $H(\theta_e)$ and the progressivity of profit tax (see equations (OA23) to (OA25)). Therefore, $H(\theta_e)$, as well as $\frac{d\ln y_e^o(\theta_e)}{d\theta_e}$, can be observed by the real profit distribution. In particular, when the profit tax in the real economy is linear around θ_e, $H(\theta_e) = \frac{1-F_{y_e}(y_e(\theta_e))}{f_{y_e}(y_e(\theta_e))}$.

Skill Gap. At the heart of the canonical Mirleesian optimal taxation problem, is the term $\frac{x_w'(\theta_w)}{x_w(\theta_w)}$, which is a measure of the skill gap and captures how the worker's skill varies along the type distribution. The equivalent to $\frac{x_w'(\theta_w)}{x_w(\theta_w)}$ for the entrepreneur is the term $\frac{\gamma_e'(\theta_e)}{\gamma_e(\theta_e)}$, i.e. the skill gap that captures how the entrepreneur's skill varies along the type distribution. This term is derived in equation (OA24) and is a crucial element for the optimal tax formula:

$$\frac{\gamma_e'(\theta_e)}{\gamma_e(\theta_e)} \equiv \left[\kappa\frac{d\ln X_e(\theta_e)}{d\theta_e} - \kappa\frac{d\ln\mu(\theta_e)}{d\theta_e} + \frac{d\ln[\mu(\theta_e)-\xi]}{d\theta_e}\right]\left[\mu(\theta_e) - \xi - \frac{\varepsilon_e}{1+\varepsilon_e}\right], \tag{33}$$

where $\kappa = \frac{\frac{\sigma}{\sigma-1}\frac{1+\varepsilon_e}{\varepsilon_e}}{\frac{1+\varepsilon_e}{\varepsilon_e}\left(\frac{\sigma}{\sigma-1}-\xi\right)-1}$. Given preferences and technology, the skill gap for the entrepreneur is no only determined by the composite ability $X_e(\theta_e)$, but also by the markup $\mu(\theta_e)$. When markets are competitive and $\mu(\theta_e) = 1$, $\frac{\gamma_e'(\theta_e)}{\gamma_e(\theta_e)}$ equals $\frac{d\ln X_e(\theta_e)}{d\theta_e}$ multiplied by a constant. It is worth noting that the last term in the right hand side of (33), i.e., $\mu(\theta_e) - \xi - \frac{\varepsilon_e}{1+\varepsilon_e}$, is the inverse of $\varepsilon^{y_e^o}_{1-\tau_e}(\theta_e)$ multiplied by $\frac{\varepsilon_e}{1+\varepsilon_e}$. Rising market power affects the skill gap through multiple channels, one of the most important channels is via a reduction in the profit elasticity through an increase in $\mu(\theta_e)$.

Social Welfare Weights. We now introduce shorthand notation for the social welfare weights and useful elasticities. We denote $g_o(\theta_o)$ and $\bar{g}_o(\theta_o)$ as the marginal and weighted social welfare weights:

$$g_o(\theta_o) \equiv \frac{G'(V_o(\theta_o))\widetilde{f}_o(\theta_o)}{\lambda f_o(\theta_o)} \quad \text{and} \quad \bar{g}_o(\theta_o) \equiv \frac{\int_{\theta_o}^{\bar{\theta}_o} g(x)\widetilde{f}_o(x)\,dx}{1-F_o(\theta_o)}, \tag{34}$$

where $\lambda = \int_{\theta_o} G'(V_o(\theta_o))\widetilde{f}_o(\theta_o)d\theta_o$ is the shadow price of government revenue.

5 Main Results

We now analyze the properties of the economy that we have laid out under optimal taxation by the planner to solve for the benchmark model. We start by enunciating the most general result on the tax formula in section 5.1. Because of the complexity of the expression of the main result, we then show a series of results that pertain to special cases in section 5.2: (i) homogeneous agents, (ii) monopolistic competition ($I = 1$), (iii) oligopolistic competition with uniform markups ($\mu(\theta_e) = \mu$), and (iv) the general case of oligopolistic competition with heterogeneous markups. Each of these special cases gradually reveal the different components of the optimal tax wedges.

5.1 General Tax Formulas

The framework of this article is applicable to general, non-parametric technological specifications. In Online Appendix OC.3, we provide statistic-based optimal tax formulas which do not depend on specific technologies (see e.g., Theorem OC1). For analytical reasons, we now provide a parameter-based optimal taxation in the Atkeson-Burstein economy. It shows how the optimal profit income tax deviate from the optimal labor income tax due to market power.

Theorem 1 *For any $\theta_w \in \Theta_w$ and $\theta_e \in \Theta_e$, the optimal tax wedges satisfy the following equations in the Atkeson-Burstein economy:*

$$\frac{1}{1-\tau_w(\theta_w)} = \frac{1}{\mu}\left[1+[1-\bar{g}_w(\theta_w)]\frac{1-F_w(\theta_w)}{f_w(\theta_w)}\frac{x_w'(\theta_w)}{x_w(\theta_w)}\frac{1+\varepsilon_w}{\varepsilon_w}\right], \tag{35}$$

$$\frac{1}{1-\tau_e(\theta_e)} = \frac{\frac{1}{\mu(\theta_e)}\left[1+[1-\bar{g}_e(\theta_e)]\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{\gamma_e'(\theta_e)}{\gamma_e(\theta_e)}\frac{1+\varepsilon_e}{\varepsilon_e}\right]+\frac{\frac{\sigma}{\sigma-1}}{\frac{\sigma}{\sigma-1}-\xi}IRE(\theta_e)}{1-\frac{\xi}{\frac{\sigma}{\sigma-1}-\xi}RE(\theta_e)}. \tag{36}$$

where the Reallocation Effect $RE(\theta_e)$ and Indirect Redistribution Effect $IRE(\theta_e)$ are given by:

$$RE(\theta_e) \equiv \frac{\mu}{\mu(\theta_e)}-1, \tag{37}$$

$$IRE(\theta_e) \equiv \varepsilon_{Q-ij}^{P,cross}(\theta_e)\left\{[1-g_e(\theta_e)]-[1-\bar{g}_e(\theta_e)]H(\theta_e)\right\}. \tag{38}$$

The aggregate markup is $\mu = \int_{\theta_e}\mu(\theta_e)\,\omega(\theta_e)\,d\theta_e$, with $\omega(\theta_e) = \frac{L_w(\theta_e)f_e(\theta_e)}{\int_{\theta_e}L_w(\theta_e)f_e(\theta_e)d\theta_e}$:

$$\omega(\theta_e) = \frac{\left[[1-\tau_e(\theta_e)]\left(\frac{X_e(\theta_e)}{\mu(\theta_e)}\right)^{\frac{1+\varepsilon_e}{\varepsilon_e}\frac{\sigma}{\sigma-1}}\right]^{\frac{1}{\frac{1+\varepsilon_e}{\varepsilon_e}\left(\frac{\sigma}{\sigma-1}-\xi\right)-1}}f_e(\theta_e)}{\int_{\theta_e}\left[[1-\tau_e(\theta_e)]\left(\frac{X_e(\theta_e)}{\mu(\theta_e)}\right)^{\frac{1+\varepsilon_e}{\varepsilon_e}\frac{\sigma}{\sigma-1}}\right]^{\frac{1}{\frac{1+\varepsilon_e}{\varepsilon_e}\left(\frac{\sigma}{\sigma-1}-\xi\right)-1}}f_e(\theta_e)\,d\theta_e}. \tag{39}$$

Proof. See Appendix B.3. ■

Two things are worth noting here. First, equations (35), (36), together with the weights for firm-level markups, i.e., (equation 39), describe the optimal taxation as a solution to an integral equation. As an illustration, in Online Appendix OC.1, we solve the optimal tax rate explicitly for specific parameters (see Corollary 2). Second, the firm-level markup serves as a sufficient statistic capturing the influence of the market structure I and the elasticity of substitution $\eta(\theta_e)$. That is I and $\eta(\theta_e)$ only affect the optimal tax formulas and the equilibrium conditions via $\mu(\theta_e)$. The general tax formula can be decomposed into four elements, each of which is defined so that the tax rates increase in the element:

Mirrleesian Part. When the goods market is competitive, the tax rules reduce to the traditional Mirrleesian tax formulas. That is $\tau_o(\theta_o) = \tau_o^M(\theta_o)$, and

$$\begin{aligned}\frac{\tau_w^M(\theta_w)}{1-\tau_w^M(\theta_w)} &= [1-\bar{g}_w(\theta_w)]\frac{1-F_w(\theta_w)}{f_w(\theta_w)}\frac{x_w'(\theta_w)}{x_w(\theta_w)}\frac{1+\varepsilon_w}{\varepsilon_w}, \\ \frac{\tau_e^M(\theta_e)}{1-\tau_e^M(\theta_e)} &= [1-\bar{g}_e(\theta_e)]\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{\gamma_e'(\theta_e)}{\gamma_e(\theta_e)}\frac{1+\varepsilon_e}{\varepsilon_e}.\end{aligned} \tag{40}$$

We call $\frac{\tau_o^M(\theta_o)}{1-\tau_o^M(\theta_o)}$ the *Mirrleesian part* of optimal taxation, and it captures the trade-off between the direct redistribution and revenue effects of the profit tax.

One core component of the Mirrleesian part is the skill gap $\frac{\gamma_e'(\theta_e)}{\gamma_e(\theta_e)}$. Rising market power affects the optimal profit tax through multiple channels, the most important of which is increasing or decreasing the Mirrleesian part by raising or reducing $\frac{\gamma_e'(\theta_e)}{\gamma_e(\theta_e)}$. In the quantative analysis below, we show that the observed change in markups increases the top profit tax rate by increasing $\frac{\gamma_e'(\theta_e)}{\gamma_e(\theta_e)}$.

Pigouvian Part. The explicit impact of market power is illustrated with the appearance of the reciprocal of the markup, i.e., $\frac{1}{\mu}$ and $\frac{1}{\mu(\theta_e)}$. We call $\frac{1}{\mu}$ and $\frac{1}{\mu(\theta_e)}$ the *Pigouvian part* for labor income tax and profit tax, respectively. A noteworthy aspect is that the optimal tax rate depends on the employment-weighted average markup instead of other aggregate markups. The employment-weighted average markup is generally larger than the unweighted average markup because it assigns a higher weight to large firms that have higher markups. In addition, the change in the employment-weighted average markup is also larger than the change in unweighted average markup if the changes in markups mainly come from large firms. Moreover, the reallocation effect, $RE\left(\theta_e\right) = \frac{\mu}{\mu(\theta_e)} - 1$, of low skill segments is larger than the effect using the unweighted mean, while the reallocation effect of high skill segments tends to be be smaller.

Reallocation Effect and Indirect Redistribution Effect. In Theorem 1, we also have the *reallocation effect* and *indirect redistribution effect*, which together with the Pigouvian and Mirrleesian parts describe the optimal tax rates:

(i) The term $RE\left(\theta_e\right)$ captures the reallocation effect of taxes. Introducing $RE\left(\theta_e\right)$ increases the optimal tax rate if and only if $\mu\left(\theta_e\right) < \mu$. Otherwise, it decreases the optimal tax rate. This is because the labor demand of a high-markup firm is inefficiently lower than that of a low-markup firm. Thus, interventions in the product market should reallocate labor factors to the high-markup firms.

(ii) The term $IRE\left(\theta_e\right)$ captures the indirect redistribution effect of the profit tax. It contains two redistribution effects caused by a percentage change in $Q_{-ij}(\theta_e)$: a local redistribution effect captured by $\varepsilon_{Q_{-ij}}^{P,cross}\left(\theta_e\right)\left[1 - g_e(\theta_e)\right]$ and a cumulative redistribution effect $\varepsilon_{Q_{-ij}}^{P,cross}\left(\theta_e\right)\left[1 - \bar{g}_e(\theta_e)\right] H(\theta_e)$.

Depending on the distribution of skills and social welfare weights the indirect redistribution effect may either increase or decrease the optimal tax rate. As an illustration to the indirect redistribution effect, consider a decrease in $\tau_e\left(\theta_e\right)$. Such a tax reform increases the output of firms in submarket θ_e (i.e., $Q_{-ij}(\theta_e)$ increases), which in turn decreases the price of products in θ_e submarket. A decrease of $P\left(\theta_e\right)$ reduces the after-tax income of the θ_e entrepreneur, which promotes redistribution if $g_e(\theta_e) < 1$. However, to keep the entrepreneurial effort unchanged after the decrease of price,[23] $\tau_e\left(\theta_e\right)$ should decrease further. This decrease in $\tau_e\left(\theta_e\right)$ triggers a cumulative indirect redistribution effect because the tax liability of all entrepreneurs with ability higher than θ_e decreases. Such a a cumulative indirect redistribution effect hinders the enhancement of welfare because $\bar{g}_e(\theta_e) \leq 1$. In the end, suppose $g_e(\theta_e)$ is small enough such that $\left[1 - g_e(\theta_e)\right] - \left[1 - \bar{g}_e(\theta_e)\right] H(\theta_e) > 0$, the indirect redistribution effect $IRE\left(\theta_e\right)$ will reduce $\tau_e\left(\theta_e\right)$. Otherwise, it increases $\tau_e\left(\theta_e\right)$.

In summary, even without considering changes in social welfare weights, the impact of rising market power and uneven market power on the optimal tax rate is ambiguous. Next, we analyze the impact of changes in market power on the optimal tax rate.

[23] Note that the indirect redistribution is the influence of changes in price, and therefore should maintain the effort unchanged. The influence of tax on effort is captured by the traditional parts.

5.2 Market Structure and Optimal Taxation

We now study the impact of market structure on tax design. We gradually adjust the number of firms and evaluate the effect on the equilibrium outcome and the optimal tax.

(i) Homogeneous Agents To make the results more transparent, we first analyze the optimal taxation formulas when workers and entrepreneurs are homogeneous. As is in Akcigit et al. (2016), without asymmetric information, the government can achieve the first best and correct the externality by Pigovian taxes.

Proposition 3 *When worker and entrepreneur types are homogeneous, the optimal tax wedges satisfy the following:*

$$\tau_w = \tau_e = 1 - \mu. \tag{41}$$

Proof. In this case, $\bar{g}_o = g_o = 1$, and the optimal tax formulas (35) and (36) simplify to (41). ■

The tax wedge here plays the exclusive role of a Pigouvian tax. When agents are identical, the incentive constraints are muted because they are trivially satisfied. The only role the planner bestows on the tax wedges is to correct the externality or markup distortion. As a result, the optimal tax wedge, which is negative, exactly offsets the distortion due to the markup. This may seem surprising, but to counteract the externality from market power, the planner subsidizes entrepreneurs and workers in order to increase output. Note that Proposition 3 holds irrespective of the number of competitors I in each market and includes cases of both monopolistic and oligopolistic competition. Of course, the exact value of μ depends on the number of firms I.

(ii) Monopolistic Competition We now turn to an economy without strategic interaction within each market j, that is, with a monopolistic producer in each market where $I = 1$. In this case, it follows that markups are identical and equal to: $\mu = \frac{\sigma}{\sigma-1}$. Under the monopolistic competition with uniform markups, the solution to the planner's problem yields the following optimal tax rules:

Proposition 4 *When $I = 1$ the optimal profit tax can be simplified as:* [24]

$$\frac{1}{1-\tau_e(\theta_e)} = \frac{1+[1-\bar{g}_e(\theta_e)]\frac{1+\varepsilon_e}{\varepsilon_e}\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\left[\mu\frac{\chi'(\theta_e)}{\chi(\theta_e)}+\frac{x_e'(\theta_e)}{x_e(\theta_e)}\right]}{\mu}. \tag{42}$$

Proof. Ommited.[25] ■

Two things are worth noting. First, the Pigouvian part in the profit tax now corresponds to the Pigouvian part in the labor income tax, because the firm-level markups are uniform and equal to the average markup. Second, the after-tax retention rate of labor income may increase relative to the after-tax retention rate of profit. To see this, note that:

$$\frac{1-\tau_w(\theta_w)}{1-\tau_e(\theta_e)} = \frac{1+[1-\bar{g}_e(\theta_e)]\frac{1+\varepsilon_e}{\varepsilon_e}\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\left[\mu\frac{\chi'(\theta_e)}{\chi(\theta_e)}+\frac{x_e'(\theta_e)}{x_e(\theta_e)}\right]}{1+[1-\bar{g}_w(\theta_w)]\frac{1+\varepsilon_w}{\varepsilon_w}\frac{1-F_w(\theta_w)}{f_w(\theta_w)}\frac{x_w'(\theta_w)}{x_w(\theta_w)}}.$$

The term $\frac{1-\tau_w(\theta_w)}{1-\tau_e(\theta_e)}$ increases in μ if the distribution parameter increases in the skill, i.e., $\frac{\chi'(\theta_e)}{\chi(\theta_e)} > 0$. In this case the markup is determined by the elasticity of substitution between markets σ. So the increase of $\frac{1-\tau_w(\theta_w)}{1-\tau_e(\theta_e)}$ in markups means that $\frac{1-\tau_w(\theta_w)}{1-\tau_e(\theta_e)}$ increases as the elasticity of substitution elasticity between goods decreases.

[24] Note that when $I = 1$, one has $\mu\frac{\chi'(\theta_e)}{\chi(\theta_e)} + \frac{x_e'(\theta_e)}{x_e(\theta_e)} = \mu\frac{X'(\theta_e)}{X(\theta_e)}$, because $\mu = \frac{\sigma}{\sigma-1}$. Therefore, we have another form of equation (42): $\frac{1}{1-\tau_e(\theta_e)} = \frac{1+[1-\bar{g}_e(\theta_e)]\frac{1+\varepsilon_e}{\varepsilon_e}\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\mu\frac{X'(\theta_e)}{X(\theta_e)}}{\mu}$, $\forall\theta_e \in \Theta_e$. This finding again suggests that the optimal tax rate is irrelevant to the specific composition of $X(\theta_e)$.

[25] When $I = 1$, one has $\mu(\theta_e) = \mu$ and $\varepsilon^{P,cross}_{Q_{-ij}}(\theta_e) = 0$. In this case, equation (36)) reduces to (42).

Note also that $\mu\frac{\chi'(\theta_e)}{\chi(\theta_e)}+\frac{x_e'(\theta_e)}{x_e(\theta_e)}=\frac{X_e'(\theta_e)}{X_e(\theta_e)}$, which means the optimal profit tax is determined by the composite ability and not $x_e(\theta_e)$ and $\chi(\theta_e)$ individually.

(iii) Oligopolistic Competition with Uniform Markups In the previous two cases, the span of control ξ has no influence on $\tau_e(\theta_e)$. This is not so under oligopolistic competition. We now consider cases with $I>1$ but still restrict the markup to be uniform. That is, $\eta(\theta_e)$ is constant. This setting introduces inter-firm strategic action but still abstracts from the effect of markup inequality between firms. A planner who intends to take advantage of the general equilibrium price effect and ease the incentive constraint would like to decrease the relative price of goods produced by high-skilled entrepreneurs. However, whether the planner should encourage the factor inputs of high-skilled entrepreneurs remains ambiguous because there are two opposing forces. On the one hand, raising the labor inputs of competitors in the same submarket reduces the relative price of goods in the submarket; on the other hand, raising the labor inputs increases entrepreneurial effort's marginal productivity.

Proposition 5 *Let $\eta(\theta)=\eta$ be constant. Then result (i) below holds. In addition, let the social welfare weights be exogenous, then also results (ii) and (iii) hold:*

(i) For any $\theta_e\in\Theta_e$, the optimal profit tax wedge satisfies:

$$\frac{1}{1-\tau_e(\theta_e)}=\frac{1+[1-\bar{g}_e(\theta_e)]\,H(\theta_e)\frac{1}{\varepsilon^{y_e^o}_{1-\tau_e}}}{\mu}+\frac{\frac{\sigma}{\sigma-1}}{\frac{\sigma}{\sigma-1}-\xi}IRE(\theta_e),\tag{43}$$

where $\varepsilon^{y_e^o}_{1-\tau_e}=\frac{1}{\frac{1+\varepsilon_e}{\varepsilon_e}(\mu-\xi)-1}$.

(ii) For any $\theta_e\in\Theta_e$, $\tau_e(\theta_e)$ increases in μ iff

$$g_e(\theta_e)<\frac{\xi(\sigma-1)}{\sigma}\left[1+[1-\bar{g}_e(\theta_e)]\,H(\theta_e)\frac{1}{\varepsilon^{y_e^o}_{1-\tau_e}}\right].\tag{44}$$

$\frac{1-\tau_w(\theta_w)}{1-\tau_e(\theta_e)}$ increases in μ iff

$$g_e(\theta_e)<1+[1-\bar{g}_e(\theta_e)]\,H(\theta_e)\frac{1}{\varepsilon^{y_e^o}_{1-\tau_e}}.\tag{45}$$

(iii) In particular, $\frac{1-\tau_w(\theta_w)}{1-\tau_e(\theta_e)}$ increases in μ if $g_e(\theta_e)<1$ and $\tau_e(\theta_e)$ increases in μ if $\bar{g}_e(\theta_e)=g_e(\theta_e)\le\frac{\xi(\sigma-1)}{\sigma}$.

Proof. See Online Appendix OC.7. ■

Part (i) of Proposition 5 provides an optimal profit tax formula, which is explicit when the social welfare weights are exogenous. Compared to the tax formula under monopolistic competition (42), there is now an additional term, i.e., $\frac{\frac{\sigma}{\sigma-1}}{\frac{\sigma}{\sigma-1}-\xi}IRE(\theta_e)$, which captures the indirect redistribution effect of the profit tax.

As is illustrated before, $IRE(\theta_e)$ is the indirect redistribution effect of changing $Q_{ij}(\theta_e)$. We now illustrate what's $\frac{\frac{\sigma}{\sigma-1}}{\frac{\sigma}{\sigma-1}-\xi}$. It is the percentage change of $Q_{ij}(\theta_e)$ with one percentage increase of $l_e(\theta_e)$. To see this, note that one percentage increase of $l_e(\theta_e)$ induces one percentage increase of $Q_{ij}(\theta_e)$. Then, the labor demand $L_w(\theta_e)$ will increase proportionally by $-\xi\frac{\varepsilon^{\varpi}_{l_e}(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)}$ percent which renders the marginal productivity of L_w between firms uniform, where $\varepsilon^{\varpi}_{l_e}(\theta_e)$ and $\varepsilon^{\varpi}_{L_w}(\theta_e)$ are the elasticities of productivity $\varpi(\theta_e)$ with respect to $l_e(\theta_e)$ and $L_w(\theta_e)$. This crowding in effect of $l_e(\theta_e)$ on $L_w(\theta_e)$ induces the $-\xi \frac{\varepsilon^{\varpi}_{l_e}(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)}$ percentage change in $Q_{ij}(\theta_e)$. In sum, one percentage increase of $l_e(\theta_e)$ triggers $1 - \xi \frac{\varepsilon^{\varpi}_{l_e}(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)} = \frac{\frac{\sigma}{\sigma-1}}{\frac{\sigma}{\sigma-1}-\xi} > 1$ percentage increase of $Q_{ij}(\theta_e)$, and $\frac{\frac{\sigma}{\sigma-1}}{\frac{\sigma}{\sigma-1}-\xi} IRE(\theta_e)$ is the total indirect redistribution effect bringed by the increase of $l_e(\theta_e)$.

The literature has pointed out that the indirect redistribution effect generally makes the top tax rate lower (see e.g., Stiglitz (1982) and Rothschild and Scheuer (2013)).[26] To see this, suppose the social welfare weights are exogenous and the government has a preference for equality, where $g_e(\theta_e)$ and $\bar{g}_e(\theta_e)$ approach zero when $\theta_e \to \overline{\theta}_e$. Notice that in our oligopolistic competition with $\eta(\theta) = \eta$, $\varepsilon^{P,cross}_{Q_{-ij}}(\theta_e) \leq 0$ is constant. As $\theta_e \to \overline{\theta}_e$, $IRE(\theta_e)$ approaches $\varepsilon^{P,cross}_{Q_{-ij}}(\theta_e)[1 - H(\theta_e)]$, which is negative if $H(\theta_e) < 1$. The indirect redistribution effect generally requires a lower profit tax rate for high-skilled entrepreneurs in the United States, because for θ_e large enough it is empirically true that $H(\theta_e) < 1$.[27]

Part (ii) of Proposition 5 provides a condition under which $\frac{1-\tau_w(\theta_w)}{1-\tau_e(\theta_e)}$ increases in μ. We find that $\frac{1-\tau_w(\theta_w)}{1-\tau_e(\theta_e)}$ increases in μ if $g_e(\theta_e) < 1$, where the shadow price of government revenue is essentially 1. Part (iii) of Proposition 5 suggests that under reasonable $H(\theta_e)$, the top profit tax rate will increase with the rise of market power when $g_e(\theta_e)$ for the top is constant (so that for θ_e large enough, $\bar{g}_e(\theta_e)$ is also constant and $\bar{g}_e(\theta_e) = g_e(\theta_e)$) and low enough.

The above findings indicate that without considering market power inequality, the optimal profit tax rates for the top firms will increase with the rising markup.

(iv) Oligopolistic Competition with Heterogeneous Markups. Finally, we get to the full-blown tax formulas with both oligopolistic competition and heterogeneous markups from Theorem 1. For the workers, the tax formula (35) remains unchanged compared to the case with uniform markups. As for the entrepreneurs, now the planner can use the tax to implement an efficiency-enhancing reallocation of factors, which is captured by the denominator on the right side of (36), i.e., $1 - RE(\theta_e) \frac{\xi}{\frac{\sigma}{\sigma-1}-\xi} = 1 - \frac{\mu-\mu(\theta_e)}{\mu(\theta_e)} \frac{\xi}{\frac{\sigma}{\sigma-1}-\xi}$.

To understand the reallocation effect, we introduce two elasticities. Denote $\varepsilon^{\varpi}_{L_w}(\theta_e) \equiv \frac{\partial \ln \varpi(\theta_e)}{\partial \ln L_w(\theta_e)}$ and $\varepsilon^{\varpi}_{l_e}(\theta) \equiv \frac{\partial \ln \varpi(\theta_e)}{\partial \ln l_e(\theta_e)}$ as the own elasticities of wage with respect to labor inputs and effort. In the Atkeson-Burstein economy, $\varepsilon^{\varpi}_{L_w}(\theta_e) = \xi \frac{\sigma-1}{\sigma} - 1$ and $\varepsilon^{\varpi}_{l_e}(\theta_e) = \frac{\sigma-1}{\sigma}$ (see Appendix A.2.2).

Notice that $\frac{1}{\frac{\sigma}{\sigma-1}-\xi} = -\frac{\varepsilon^{\varpi}_{l_e}(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)}$ is the percentage increase in $L_w(\theta_e)$ needed to render the marginal productivities of labor inputs $(\varpi(\theta_e))$ between firms uniform when $l_e(\theta_e)$ is increased by one percent. $\frac{\xi}{\frac{\sigma}{\sigma-1}-\xi}$ is the percentage increase in Q_{ij} with one percentage increase in $L_{w,ij}$.

The increase of $L_w(\theta_e)$ comes from the reallocation of L_w at other firms. The influence of such a reallocation on the aggregate output is captured by $\frac{\mu(\theta_e)-\mu}{\mu(\theta_e)}$, where

$$[\mu(\theta_e) - \mu] W = \mu(\theta_e) W - \int_{\theta_e} P(\theta'_e) \frac{\partial Q_{ij}(\theta'_e)}{\partial L_w(\theta'_e)} \frac{L_w(\theta'_e) f_e(\theta'_e)}{\int L_w(\theta_e) f_e(\theta_e) d\theta_e} d\theta'_e$$

is the increase in total output by reallocating $\frac{L_w(\theta'_e)}{\int L_w(\theta_e) f_e(\theta_e) d\theta_e}$ units of labor factors from each of θ'_e-type firms to the θ_e-type firm. To see this, remember that $\mu = \int_{\theta_e} \mu(\theta'_e) \frac{L_w(\theta'_e) f_e(\theta'_e)}{\int L_w(\theta_e) f_e(\theta_e) d\theta_e} d\theta'_e$, where $\mu(\theta'_e) W = P(\theta'_e) \frac{\partial Q_{ij}(\theta'_e)}{\partial L_w(\theta'_e)}$ is the marginal output of $L_w(\theta'_e)$. Last, $\frac{\mu(\theta_e)-\mu}{\mu(\theta_e)} = \frac{[\mu(\theta_e)-\mu]W}{P(\theta_e)\frac{\partial Q_{ij}(\theta_e)}{\partial L_w(\theta_e)}}$ is the value of the increased output in terms

[26] Rothschild and Scheuer (2013) develop the notion of a so-called self-confirming policy equilibrium, used by Rothschild and Scheuer (2016), to illustrate the potential influence of neglecting the endogeneity of wages, therefore, the IRE.

[27] In 2007, the hazard ratio of top labor, capital, and total incomes in the United States are around 0.62, 0.76 and 0.71, respectively (see e.g., Saez and Stantcheva (2018)).

of $L_w(\theta_e)$. Notice that $1 - RE(\theta_e)\frac{\xi}{\frac{\sigma}{\sigma-1}-\xi} = 1 - \frac{\mu-\mu(\theta_e)}{\mu(\theta_e)}\frac{\xi}{\frac{\sigma}{\sigma-1}-\xi}$, where $1 = \frac{\partial \ln Q_{ij}(\theta_e)}{\partial \ln l_e(\theta_e)}$ and $\xi = \frac{\partial \ln Q_{ij}(\theta_e)}{\partial \ln L_w(\theta_e)}$. $1 - RE(\theta_e)\frac{\xi}{\frac{\sigma}{\sigma-1}-\xi}$ is exactually the output that increased with one percent increase of $l_e(\theta_e)$ and the resulting inter-firm reallocation of labor factors.

Our optimal tax formula suggests that, given the social welfare weights, with a marginal increase of the firm-level markup, the reallocation effect requires a lower tax rate on the firm with a markup higher than the average markup.[28] The above finding also provides a novel explanation (i.e., markup inequality) for why the profit tax in the real economy is less progressive (or not progressive at all) than the labor income tax (see e.g., Scheuer (2014)). It is worth noting here is that rising markup inequality not necessarily decreases the optimal profit tax rate. In fact, it depends on the rate at which the firms' markup increases compared to the average markup. When the growth rate of a firm's markup is faster than the average markup, changes in the reallocation effect often require a reduction in the firm's marginal profit tax rate.

5.3 Summary of Results

There are four elements determining the optimal profit tax rate, in addition to the social welfare weights. The Mirrleesian part reflects the trade-off between direct redistribution and the revenue effect of profit tax. The Pigouvian part restores the productive efficiency by offsetting the externality of labor supply. Meanwhile, the reallocation effect reduces the tax rate for entrepreneurs with relatively high markups in order to reduce the misallocation of labor inputs. Last, the indirect redistribution effect captures tax's redistribution effect through prices.

These four parts suggest that changes in market power has an amalgam of different, often opposing forces on the optimal profit tax rate. As an illustration, consider the increase of $\mu(\theta_e)$ in the Atkeson-Burstein economy:

1. Immediately, the Pigouvian part $\frac{1}{\mu(\theta_e)}$ decreases to reduce the tax rate.
2. The reallocation effect $RE(\theta_e) = \frac{\mu}{\mu(\theta_e)} - 1$ decreases if the firm-level markup $\mu(\theta_e)$ increases as relative to the average markup.
3. The indirect redistribution effect, i.e., $IRE(\theta_e) = \varepsilon^{P,cross}_{Q_{-ij}}(\theta_e)\{[1-g_e(\theta_e)] - [1-\bar{g}_e(\theta_e)]H(\theta_e)\}$, may either increase or decrease. The cross-inverse demand elasticity $\varepsilon^{P,cross}_{Q_{-ij}}(\theta_e) = -\frac{1}{\mu(\theta_e)} + \frac{\sigma-1}{\sigma}$ increases with the increase of $\mu(\theta_e)$. However, $[1-g_e(\theta_e)] - [1-\bar{g}_e(\theta_e)]H(\theta_e)$ may either be positive or negative (it is generally negative for the poor and positive for the rich).
4. Last, the core of Mirrleesian part – the skill gap $\frac{\gamma_e'(\theta_e)}{\gamma_e(\theta_e)}$ – generally increases because $\varepsilon^{y_e^o}_{1-\tau_e}(\theta_e) = \frac{1}{\mu(\theta_e)-\xi-\frac{\varepsilon_e}{1+\varepsilon_e}}$ decreases (see equation (33)).

In conclusion, our theoretical analysis identifies the different forces behind the effect of market power on the profit tax rate, especially for the top firms. To evaluate the net effect as well as the contribution of each individual force, below in Section 7 we perform a full quantitative exercise based on US micro-data in order to provide a policy prescription how the the tax authorities should react to the change in market power that we observe since the 1980s. We then establish the robustness of our main findings.

[28]Since there is no use to set a marginal tax rate larger than one, the right side of (36) is positive. Supposing $\tau_e(\theta_e) < 1$, the numerator of the right side of (36) is positive if the denominator $1 - \frac{\xi}{\frac{\sigma}{\sigma-1}-\xi}RE(\theta_e) > 0$, which is true because $\mu < \frac{\sigma}{\sigma-1}$.

6 Discussion, Policy-relevant Specifications and Robustness

In section 6.1, we discuss how our study relates to the literature. In section 6.2 we consider alternative specifications that are relevant for policy. In the Appendix C we discuss further Extensions and Robustness of our model setup and results.

6.1 Relation to the Literature

First, we review the literature on which our work builds. Then we enter into further detail of the analytic of the model to expand on the comparison with key results in the literature. Starting with Mirrlees (1971), an extensive and influential literature on optimal taxation has analyzed what determines the properties of income tax schedules. Within this literature, our paper speaks to three strands: (i) market power and optimal policies; (ii) endogenous price (wage) and optimal taxation; and (iii) entrepreneurship and optimal taxation. Our paper is further also related to the literature on externalities and optimal taxation, and to the literature on technology and optimal taxation.

In recent years, there is a growing policy literature on the relation between markups and inequality (see e.g., Stiglitz (2012); Atkinson (2015); Baker and Salop (2015); Khan and Vaheesan (2017)). This paper differs from existing research in several aspects. First, existing papers generally consider representative agents. These papers abstract from distributional concerns and focus on indirect taxes (see Stern (1987); Myles (1989); Cremer and Thisse (1994); Anderson et al. (2001); Colciago (2016); Atesagaoglu and Yazici (2021)). They assume that a lump-sum tax is not enforceable and study how can the government raise revenue efficiently. In a recent paper, Atesagaoglu and Yazici (2021) analyze the effect of optimal taxation on the labor share in a Ramsey problem with capital. They ask a different but related question, namely whether it is optimal to tax capital rather than labor when there is pure profit and the planner cannot distinguish capital income from profits. We study optimal taxation in the spirit of Mirrleesian taxations and consider indirect as well as direct taxes. Therefore, our approach highlights the trade-off between efficiency and equality in tax system design.

Second, for those papers considering redistribution in the presence of market power, the equilibrium often considered is monopolistic competition (see e.g., Gürer (2021) and Boar and Midrigan (2021)), or where markups are exogenous (see e.g., Kaplow (2019)). These studies miss out on the impact of market structure on tax systems. In our Atkeson and Burstein (2008) economy, there are a finite number of oligopolistic firms that have market power in their local market.[29] This setting allows us to model the influence of the market structure on the optimal design of the tax system, which is mostly absent in the literature. Our results lay bare the impact of market structure on the direct and indirect redistribution, as well as the capacity of taxation to promote production. The role of the market structure is not merely a theoretical frivolity. It is important for the overall conclusion and for the concrete policy prescription. While the reallocation effect (RE) is a force towards more concentration, the direct redistribution effect (Mirrleesian part) counteracts this force and calls for policies that lead to less concentration and lower market power. The indirect redistribution effect (IRE) calls for less progressive policies that lead to more concentration and lower market power. In models with monopolistic competition, the IRE and RE is absent (because of zero cross-wage elasticity and uniform markups), leading to the conclusion that policymakers should broker more concentrated mar-

[29] In our setup, we have a nested CES structure in inputs of production, instead of in preferences over consumption goods.

kets to benefit redistribution towards the poor. This is not so with oligopolistic competition, and in the quantitative exercise we find that policymakers want to implement higher, but less regressive profit taxes, in order to promote the production while enhancing redistribution.

When considering oligopolistic markets, current studies generally do not consider profit tax design and strategic pricing (see e.g., Kushnir and Zubrickas (2019) and Jaravel and Olivi (2019)). The profits received by the agents in these papers are taken as given. Thus, even if the profit tax is introduced in their models, it acts as lump-sum taxes. In contrast, we introduce entrepreneurs, and entrust the pricing behavior to these agents. The technology that an entrepreneur employs is as in Lucas (1978), where a skilled entrepreneur chooses the optimal amount of labor as an input to produce output. Unlike Lucas (1978), the entrepreneur has market power and chooses prices strategically when reporting their types. Therefore, the presence of market power in the principal-agent problem is notably distinct from the existing literature on optimal taxation and market power. In doing so, our paper also reveals the influence of market power on tax design through shaping the skill gap of entrepreneurs.

Most contributions to the secondary literature on endogenous price (wage) and optimal taxation considers competitive market (see e.g., Stiglitz (1982); Naito (1999) and Naito (2004); Saez (2004); Scheuer (2014); Sachs et al. (2020); Cui et al. (2021)). This literature emphasizes the general equilibrium effect of taxes on factor prices, which brings an indirect redistribution between agents providing different factors. We show that the indirect redistribution effect crucially depends on the market structure. Lowering the profit tax encourages entrepreneurial effort and output, thereby decreasing the price of the competitor's product which leads to the indirect redistribution. Interestingly, when there is no competitor in the submarket, i.e., under monopolistic competition, this effect disappears, because the entrepreneur's monopoly price-setting action eliminates the tax policy's first-order effect on the price. In the case of oligopoly, only part of the indirect redistributive effect is offset by the pricing behavior of entrepreneurs. Observe the difference between the intuitions for the indirect redistribution effects under monopolistic competition and the perfectly competitive case as in Rothschild and Scheuer (2013). This does get more subtle under oligopolistic competition.

A third stream of research works on entrepreneurship and optimal taxation (see e.g., Scheuer (2014); Ales and Sleet (2016), Ales et al. (2017); Scheuer and Werning (2017)). Within this setting, Boar and Midrigan (2021) is the only paper that also introduces market power. They consider an alternative incentive problem between the planner and the entrepreneur where a profit tax does not affect the entrepreneur's incentive constraint. As a result, their optimal policy prescription is quantity regulation instead of a profit tax. In addition, we consider different production technologies and market structures. The source of market power in our model is the number of firms that are in oligopolistic competition, instead of preferences, which in their setting works via the Kimball aggregator under monopolistic competition.

These different modeling choices have novel implications for policy. Market power affects the optimal policy not only through the Pigouvian channel, but also the Mirrleesian channel. More interestingly, the Mirrleesian channel leads to an increase of the marginal tax rate as markups increase. This is at the heart of the role that Mirrleesian taxes play as opposing forces of Pigouvian taxes.

In addition to the three streams of literature above, our paper also contributes to the literature on optimal taxation and technology (see e.g., Ales et al. (2015); Scheuer and Werning (2017); Costinot and Werning

(2023)). Scheuer and Werning (2017) find that the parametric optimal tax rate is independent of the span of control (the curvature of firm-level production with respect to labor inputs). Our results show that their findings extend to the monopolistically competitive economy. However, we find that in the oligopolistic economy, the span of control enters the parametric tax rule by enlarging the indirect redistribution effect.

Lastly, our paper is also related to the literature on optimal taxation in the presence of externalities (e.g., Sandmo (1975); Ng (1980); Bovenberg and van der Ploeg (1994); Kopczuk (2003); Farhi and Gabaix (2020)). As suggested by Kopczuk (2003), one striking finding of this literature is the "additivity property":[30] Optimal taxation in the presence externalities can be expressed additively by some Pigouvian taxes. However, we find that the additivity property generally does not hold in an economy where agents have heterogeneous market powers because of the reallocation effect.

6.2 Policy-relevant Specifications

We now consider alternative specifications that are relevant to put our results in perspective concerning concrete tax policy prescriptions.[31]

(i) Second-Best: Non-linear Sales Taxes. As we have emphasized before and considering the practicality of the tax system, our benchmark model is constrained to linear sales taxes and therefore corresponds to the planner's third-best solution. A comparison between the second- and third-best solutions is nonetheless useful for illustrating the influence of this policy constraint on the optimal profit tax. Set $\tau_w^E(\theta_w)$, $\tau_e^E(\theta_e)$ and $\tau_s^E(\theta_e)$ as the marginal labor income tax rate, profit tax rate, and *non-linear* sales income tax rate, respectively. By the definitions of tax wedges, $\tau_w(\theta_w) = \tau_w^E(\theta_w)$, $\tau_s(\theta_e) = \tau_s^E(\theta_e)$ and $1-\tau_e(\theta_e) = \left[1-\tau_e^E(\theta_e)\right]\left[1-\tau_s^E(\theta_e)\right]$. See Online Appendix OC.5.3 for additional explicit expressions for the tax wedges. Analogous to Theorem 1, Theorem 2 provides the most general result on the optimal tax formula in this extension with non-linear sales taxes.

Theorem 2 *The optimal tax rates in the second-best problem satisfy the following:*

$$\frac{1}{1-\tau_w^E(\theta_w)} = \frac{1+[1-\bar{g}_w(\theta_w)]\frac{1-F_w(\theta_w)}{f_w(\theta_w)}\frac{x_w'(\theta_w)}{x_w(\theta_w)}\frac{1+\varepsilon_w}{\varepsilon_w}}{\mu^*},$$

$$\frac{1}{1-\tau_e^E(\theta_e)} = \frac{1+[1-\bar{g}_e(\theta_e)]H(\theta_e)\left[[\mu(\theta_e)-\xi]\frac{1+\varepsilon_e}{\varepsilon_e}-1\right]}{\mu^*}+\frac{[1-\bar{g}_e(\theta_e)]\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\left[\frac{d\ln[\mu(\theta_e)-\xi]}{d\theta_e}+\frac{\mu(\theta_e)\varepsilon_{Q-ij}^{P,cross}(\theta_e)\left(1+\xi\frac{1+\varepsilon_e}{\varepsilon_e}\right)\left[\frac{d\ln\left[1-\tau_s^E(\theta_e)\right]}{d\theta_e}+\frac{d\ln\left[1-\tau_e^E(\theta_e)\right]}{d\theta_e}\right]}{1-\frac{\sigma-1}{\sigma}\left(\frac{\varepsilon_e}{1+\varepsilon_e}+\xi\right)}\right]}{\mu^*},$$

$$\frac{\tau_s^E(\theta_e)}{1-\tau_s^E(\theta_e)} = \left[\frac{\mu^*}{\mu(\theta_e)}-1\right]+\left[1-\tau_e^E(\theta_e)\right]\varepsilon_{Q-ij}^{P,cross}(\theta_e)\left\{[1-g_e(\theta_e)]-[1-\bar{g}_e(\theta_e)]H(\theta_e)\right\}$$
$$+\left[1-\tau_e^E(\theta_e)\right]\varepsilon_{Q-ij}^{P,cross}(\theta_e)\frac{[1-\bar{g}_e(\theta_e)][1-F_e(\theta_e)]}{f_e(\theta_e)}\left[\begin{array}{c}\frac{d\ln\left[\mu(\theta_e)\varepsilon_{Q-ij}^{P,cross}(\theta_e)\right]}{d\theta_e}-\frac{d\ln[\mu(\theta_e)-\xi]}{d\theta_e}\\+\frac{\frac{d\ln\left[1-\tau_s^E(\theta_e)\right]}{d\theta_e}+\left(1-\xi\frac{\sigma-1}{\sigma}\right)\frac{d\ln\left[1-\tau_e^E(\theta_e)\right]}{d\theta_e}}{1-\frac{\sigma-1}{\sigma}\left(\frac{\varepsilon_e}{1+\varepsilon_e}+\xi\right)}\end{array}\right],$$

[30] The additivity property can be treated as a special case of the "principle of targeting"proposed by Dixit (1985).

[31] To facilitate the analysis, in this section we assume that the relevant monotonicity hypothesis of the incentive problem are always tenable and we can rely on the first-order approach.

where $\mu^* \equiv \int_{\theta_e} \frac{\mu(\theta_e)}{1-\tau_s^E(\theta_e)} \omega(\theta_e)\, d\theta_e$ *is the optimal after-tax average markup.*

Proof. See Online Appendix OC.5. ■

μ^* is a generalized average markup. Intuitively, $\frac{\mu(\theta_e)}{1-\tau_s^E(\theta_e)}$ is the markup after tax and μ^* is the weighted average of the after-tax markup. In line with our benchmark model, μ^* is reduced to μ when the sales tax is zero. Given μ^* and social welfare weights, the expressions for $\tau_e^E(\theta_e)$ and $\tau_s^E(\theta_e)$ consists of a system of differential equations. These equations together with the boundary conditions $\frac{1}{1-\tau_e^E(\underline{\theta}_e)} = \frac{1}{\mu^*}$ and $\frac{\tau_s^E(\underline{\theta}_e)}{1-\tau_s^E(\underline{\theta}_e)} = \left[\frac{\mu^*}{\mu(\underline{\theta}_e)} - 1\right] + \mu^* \varepsilon_{Q_{-ij}}(\underline{\theta}_e)\left[1 - g_e(\underline{\theta}_e)\right]$ determine the optimal profit and sales taxes.

Due to the introduction of a non-linear sales tax (or factor tax), the optimal tax system is substantially more involved. The comparison between the second and third-best solutions become more transparent when we consider the following special case:

Corollary 1 *Suppose that at point* θ_e, $\tau_s^{E\prime}(\theta_e) = \tau_e^{E\prime}(\theta_e) = \mu'(\theta_e) = 0$, *then the optimal tax rates satisfy:*

$$\frac{1}{1-\tau_e^E(\theta_e)} = \frac{1 + \left[1 - \bar{g}_e(\theta_e)\right] H(\theta_e)\left[\frac{1+\varepsilon_e}{\varepsilon_e}\left(\mu(\theta_e) - \xi\right) - 1\right]}{\mu^*}, \tag{46}$$

and

$$\frac{\tau_s^E(\theta_e)}{1-\tau_s^E(\theta_e)} = \underbrace{\left[\frac{\mu^*}{\mu(\theta_e)} - 1\right]}_{RE^*(\theta_e)} + \left[1 - \tau_e^E(\theta_e)\right] \underbrace{\varepsilon_{Q-ij}^{P,cross}(\theta_e)\left\{\left[1 - g_e(\theta_e)\right] - \left[1 - \bar{g}_e(\theta_e)\right] H(\theta_e)\right\}}_{IRE^*(\theta_e)}. \tag{47}$$

Proof. Ommited. ■

Corollary 1 directly obtains from Theorem 2. $RE^*(\theta_e)$ and $IRE^*(\theta_e)$ are analogous in Theorem 1. In particular, $RE^*(\theta_e)$ ($IRE^*(\theta_e)$) equals $RE(\theta_e)$ ($IRE(\theta_e)$) under linear taxes. Comparing Theorem 1 to Corollary 1 clarifies the role of non-linear sales taxes. The design of sales taxes incorporates the reallocation effect and indirect redistribution effect. The profit tax now only covers the Mirrleesian and Pigouvian parts. Moreover, the Pigouvian part is based on the average markup instead of the firm-level markup. Note that the profit tax rate is not the tax wedge on entrepreneurial effort considered in our benchmark model (actually $1 - \tau_e(\theta_e) = \left[1 - \tau_e^E(\theta_e)\right]\left[1 - \tau_s^E(\theta_e)\right]$). The tax wedge on entrepreneurial effort still incorporates all of the four elements. Equation (46) suggests that rising markup inequality (increase in $\mu'(\theta)$) generally makes the profit tax more progressive. However, the profit tax rate may either increase or decrease depending on the relative change of firm-level markup to the average markup.

The above analysis invites the following three considerations. First, findings in our benchmark model regarding the optimal profit tax wedge are for the total tax rate on entrepreneurial effort enforced by the profit and sales taxes, instead of the nominal profit tax alone. Second, the main function of the non-linear sales tax is to shoulder the burden of reallocating factors between firms and sales-based indirect redistribution. Therefore the non-linear sales tax is generally positive for the small firms with low markups and negative for the large firms with high markups. Third, the optimal profit tax depends on the set of enforceable policies. That said, no matter whether the non-linear sales tax is enforced, tax design critically depends on the four elements highlighted in the benchmark model. Therefore, the total tax rate borne by factors, which we do in the benchmark analysis, deserves special attention. While there are multiple nominal tax systems that can achieve the same total tax rate on factors, the optimal total tax rate on factors is generally unique.

(ii) Conditioning Taxes on Markups. In our setup so far, the planner can not condition the tax on the firm's markup. We believe there are sound practical and empirical reasons for this assumption because markups are hard to measure. Markups are the ratio of price to marginal cost. Quality data on output prices are rare to come by. What is particularly challenging is obtaining measures of marginal costs. There are different ways to robustly calculate marginal costs – most notably through demand estimation (see for example Berry et al. (1995)) or through cost minimization (see for example De Loecker and Warzynski (2012) and De Loecker et al. (2020)) – but each method requires a theoretical and statistical model. It is plausible to assume that a taxation agency will not have the resources to do this estimation for all firms.

Nonetheless, we now derive the solution even if the planner has the ability to obtain these markup estimates. We show that with the Atkeson and Burstein (2008) technology, the optimal solution where taxes condition on markups as well as profits is equivalent to the solution in Section OC.5.1 with non-linear sales tax schedules. This equivalence leads us to conclude that the first-best cannot be achieved even with tax conditions on markups.

Formally, as in the non-linear sales tax case, we do not artificially impose policy constraints, so as to focus on the information problem itself. A planner who wants to regulate market power can enforce a markup-based punishment (a tax on markups for example). In particular, the planner can design the following mechanism: an entrepreneur who reports θ_e' should set the firm-level markup at $\mu\left(\theta_e'\right)$ and earn $y_e\left(\theta_e'\right)$ units of profit. Then the entrepreneur will receive $c_e\left(\theta_e'\right)$ units of consumption. The labor input $L_w\left(\theta_e'|\theta_e\right)$ and effort $l_e\left(\theta_e'|\theta_e\right)$ must satisfy:

$$\frac{P_{ij}}{W}\frac{\partial Q_{ij}}{\partial L_w} = \mu\left(\theta_e'\right) \quad \text{and} \quad P_{ij}Q_{ij} - WL_w = y_e\left(\theta_e'\right), \tag{48}$$

where $P_{ij} = P_{ij}\left(Q_{ij}, Q_{ij}\left(\theta_e\right), \theta_e\right)$. We suppose there is a unique solution to promise-keeping constraints.

The entrepreneur's problem can again be formulated as in equation (OC6). Thus, the incentive condition of the entrepreneur is the same as that in Section OC.5.1, if and only if (OC7) holds here too. A sufficient condition for (OC7) is that $L_w\left(\theta_e'|\theta_e\right)$ is independent of θ_e, which is true under our benchmark model. To see this, notice that $\frac{\partial Q_{ij}}{\partial L_w}\frac{1}{Q_{ij}} = \frac{\xi}{L_w}$ and combine the promise-keeping constraints (48) to derive $L_w\left(\theta_e'|\theta_e\right) = \frac{y_e(\theta_e')}{W[\mu(\theta_e')/\xi-1]}$, which is exactually independent of θ_e. Now that the incentive condition remains the same, we know that the first-best optimum is not achievable even if the markup is observable.[32]

(iii) Quantity Regulation. In our benchmark model, we consider profit tax as the policy instrument to incentivize entrepreneurs. In this subsection, we consider an alternative problem that instead uses quantity regulation as described in Boar and Midrigan (2021).[33] A natural question is whether there is any difference in considering these two different policy instruments. Interestingly, as long as the type is unobservable, the answer is no. Formally, the government designs the following mechanism: an entrepreneur who reports θ_e' should produce $Q_{ij}\left(\theta_e'\right)$ units of goods and pay $T_e\left(\theta_e'\right)$ units of tax (a subsidy, if negative). Thus, the entrepreneur's problem is formulated as below:

$$V_e\left(\theta_e\right) \equiv \max_{\theta_e'} V_e\left(\theta_e'|\theta_e\right) \tag{49}$$

[32] Under more general firm-level technology $L_w\left(\theta_e'|\theta_e\right)$ may depend on θ_e. Either way, the first-best optimum is not achievable.

[33] Boar and Midrigan (2021) cannot consider profit taxes because entrepreneurs provide no effort and profit taxes therefore have no effect on behavior.

$$\text{where} \quad V_e\left(\theta_e'|\theta_e\right) = \max_{L_w, l_e} P\left(Q_{ij}\left(\theta_e'\right), \theta_e\right) Q_{ij}\left(\theta_e'\right) - WL_w - T_e\left(\theta_e'\right) - \phi_e\left(l_e\right) \tag{50}$$

$$\text{s.t. } Q_{ij}\left(\theta_e'\right) = Q_{ij}\left(x_e\left(\theta_e\right) l_e, L_w\right). \tag{51}$$

To solve the above problem, one has the following incentive condition:

$$V_e'\left(\theta_e\right) = \frac{\partial V_e\left(\theta_e'|\theta_e\right)}{\partial \theta_e}|_{\theta_e'=\theta_e} = \frac{\partial P\left(Q_{ij}\left(\theta_e\right), \theta_e\right)}{\partial \theta_e} Q_{ij}\left(\theta_e\right) + \phi_e'\left(l_e\left(\theta_e\right)\right) l_e\left(\theta_e\right) \frac{x_e'\left(\theta_e\right)}{x_e\left(\theta_e\right)}.$$

The above incentive condition is equal to the original one if and only if $P\left(\theta_e\right) Q_{ij}\left(\theta_e\right) = \phi_e'\left(l_e\left(\theta_e\right)\right) l_e\left(\theta_e\right) \mu\left(\theta_e\right)$. Since $\phi_e'\left(l_e\left(\theta_e\right)\right) l_e\left(\theta_e\right) = \frac{WL_w(\theta_e)}{\xi}$,[34], the above condition naturally holds as the definition of a markup (see e.g., equation (20)) and implies $P\left(\theta_e\right) Q_{ij}\left(\theta_e\right) = \frac{WL_w(\theta_e)\mu(\theta_e)}{\xi}$ because $\phi_e'\left(l_e\left(\theta_e\right)\right) l_e\left(\theta_e\right) = \frac{WL_w(\theta_e)}{\xi}$. Also, note that in line with the first-order conditions of the above incentive problem, $\frac{W}{\frac{\partial Q_{ij}(\theta_e)}{\partial L_w(\theta_e)} \frac{P(\theta_e)}{\mu(\theta_e)}}$ must be a constant, which implies the policy constraint in our benchmark model. Therefore, the constraints faced by the government under these two different incentive problems are exactly the same. The above finding suggests that our main finding is independent of the policy instrument and that a quantity regulation generally can be replaced by a profit tax.

7 Quantitative Analysis

To provide direct guidance for policymakers and to approximate the theoretical results to reality, we perform a full quantitative exercise using microdata from the United States economy and estimating the model to match key moments regarding. In what follows, (i) we lay out the parameterize and underlying assumptions; (ii) describe the data; (iii) calibrate the distributions of $x_w\left(\theta_w\right)$, $X_e(\theta_e)$, and $\mu\left(\theta_e\right)$ using data on income and markups; (iv) solve for the optimal tax rates and tax revenue and perform counterfactual exercises. We execute the analysis in two different years, 1980 and 2019, to evaluate how optimal taxation changes during the the period in which market power has risen sharply.

7.1 Parameterization

For our benchmanrk economy, we consider a quasi-linear utility $c - \frac{l^{1+1/\varepsilon}}{1+1/\varepsilon}$ with $\varepsilon = 0.33$ (Chetty (2012)) and take the Atkeson-Burstein production technology specified in equations (1) to (4). To capture social preferences for redistribution, we consider a concave social welfare function $G\left(V\right) = \frac{V^{1-k}-1}{1-k}$, where the parameter k governs the preference for equality. We set the key parameters $k = 0.77$, following Heathcote et al. (2017) who find that for $k = 0.77$ marginal tax rates are in the range of those observed empirically. We take the benchmark value for $\sigma = 1.4$ from Katz and Murphy (1992). Their elasticity of substitution between different inputs of skilled labor is for a CES production function as in our model. To guarantee the concavity of the entrepreneur's incentive problem, we set $\xi = 0.5$. In section 7.3, we investigate the robustness of our main results to the the choice of $\{k, \sigma, \ \xi\}$.

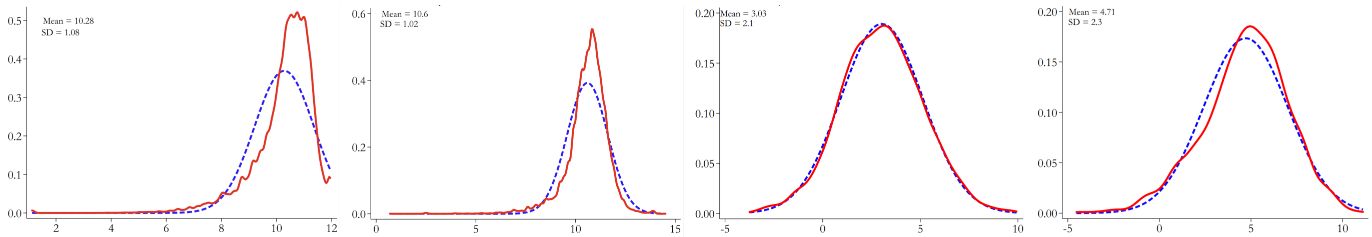
We assume θ_o is equal to the quantile of $y_o\left(\theta_o\right)$, which means $f_o\left(\theta_o\right) = 1$ and $\Theta_o = [0, 1]$. Since the functions $x_w\left(\theta_w\right)$ and $X_e(\theta_e)$ are used to govern the abilities there is no loss to assume that the distributions of skills are uniform. Following Saez and Stantcheva (2018), we set the average income tax rate to match the U.S. average tax rate on total income, which means the tax rate on labor and profit income is 25.6% in 1980

[34] This equation can be derived by the first-order conditions of the entrepreneur's problem.

and 25% in 2019. Tax revenue is returned to the agent through lump sum transfer payments.

7.2 Data and Calibration

We calibrate our economy to the distributions of labor income, profits, and markups for the US economy.



(a) Labor Income - 1980 (b) Labor Income – 2019 (c) Profits – 1980 (d) Profits – 2019

Figure 1: Labor Income and Profit Distributions

Labor Income. We obtain the income distribution of salaried workers in 2019 and 1980 from the Current Population Survey (CPS) and use the lognormal fit to estimate the model. Figures 1a and 1b plot the income distribution of salaried workers (red) and its lognormal fit (blue) for 1980 and 2019 on a log scale, and we report the mean and standard deviation of the log wages.[35]

Profits. We use profit data from De Loecker et al. (2020) based on the sample of publicly traded firms and use the lognormal fit to estimate the model. Figures 1c and 1d plot the distribution of log profit (red) and its lognormal fit (blue) for 1980 and 2019, and report the mean and standard deviation of log profits.[36]

Markups. We obtain the markup data using the method in De Loecker et al. (2020). The firm-level markup is defined as the ratio of the output price to the marginal cost and is estimated using the cost approach. From the cost-minimization decision of the firm, the firm-level markup can be expressed as $\mu_i = \alpha^V \frac{S_i}{E_i^V}$ where μ_i is firm i's markup, α^V is the output elasticity of the variable input, S_i is the sales of firm i, and E_i^V is its expenditure on the variable input. We use accounting data on sales and expenditures on different inputs (capital and variable inputs including labor) and estimate a firm-level production function to obtain the output elasticity α^V. We obtain the individual firm-level markup in each year, and hence the markup distribution for all firms.[37]

One insight from our theoretical analysis is that the firm-level markup is a sufficient statistic that captures the influence of market power on the equilibrium allocation and optimal taxation. There is no need to calibrate $\eta\left(\theta_e\right)$ and I separately, as long as we have $\mu(\theta_e)$.

The average markup μ enters the optimal tax formulas. Because heterogeneity in markups in our model is between markets and not within, we use the cost-weighted markup in the model and match it to the cost-weighted markup in the data.[38] Formally, we first calculate firm-level markups as described above, and rank the firms by their firm-level markups μ_{ij}. In our model, markups are increaseing in θ. Denote by

[35] We consider total pre-tax wage and salary income for the previous year. Armed forces and agriculture are excluded from the analysis. The sample includes those workers between 16 and 64 years old who were full-time employed during the full year and whose income was bigger than 0.

[36] Profits are in millions of dollars in 2019 prices, truncated at 0.

[37] We rank markups and truncate the sample below at 1 and winsorize the top at 0.8% to remove outliers.

[38] For a discussion on the distinction between alternative weighting of average markups, see De Loecker et al. (2020); Edmond et al. (2019).

$\vartheta_{ij} = \frac{\sum_{\mu_{i'j'} \leq \mu_{ij}} L_{w,i'j'}}{\sum_{i'j'} L_{w,i'j'}}$ the cost weight of the firms with markup not higher than μ_{ij}. We match the cost weighted markup schedule $\mu(\theta_e)$ from the model to the data, using a polynomial fit.

Figure 2 reports the data and the model fit of the cost-weighted markup distribution for 1980 and 2019. The cost-weighted average markup increases from 1.26 in 1980 to 1.38 in 2019. The plot shows the rise in the average markup is driven predominantly by the rise of the markup in the top percentiles of the distribution.

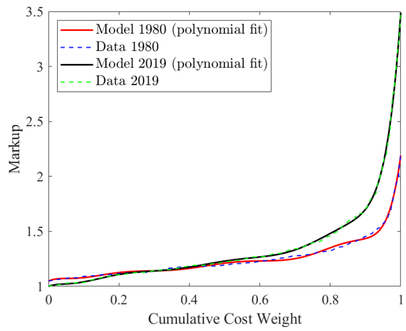
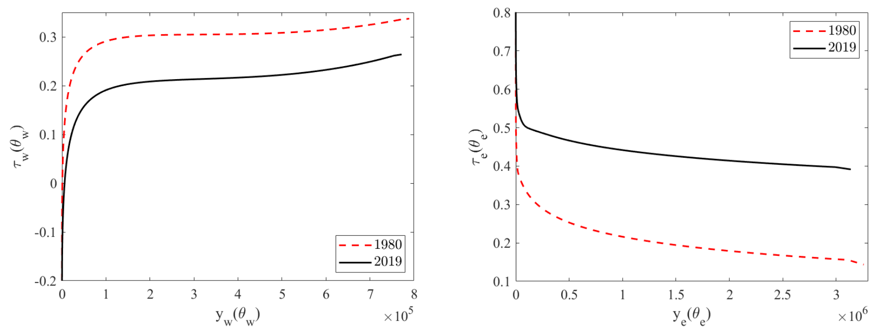


Figure 2: Markup Distribution: data and calibrated model



(a) Income-based Income Tax Wedges

(b) Profit-based Profit Tax Wedges

Figure 3: Optimal Tax Wedges in 1980 and 2019

Skill Gap. Finally, we calibrate $x_w(\cdot)$ and $X_e(\cdot)$ using the agents' first-order conditions. Our theoretical analysis suggests that the equilibrium allocation and optimal taxation do not depend on the decomposition of $X_e(\theta_e)$ into $x_e(\theta_e)$ and $\chi(\theta_e)$. Therefore, there is no need to calibrate $x_e(\theta_e)$ and $\chi(\theta_e)$ seperately, and instead we calibrate $X_e(\theta_e)$, or alternatively, $\gamma_e(\theta_e)$.[39] Equation (33) then gives us the expression for the skill gap $\frac{\gamma_e'(\theta_e)}{\gamma_e(\theta_e)}$, which we plot and discuss in Figure 5 below. Last, we set $N_e = 1$ and derive N_w by the labor market clearing condition, which is equivalent to (because $\frac{y_e(\theta_e)}{W}\frac{\xi}{\mu(\theta_e)-\xi} = L_w(\theta_e)$):

$$N_w \int_{\theta_w} y_w(\theta_w) f_{\theta_w}(\theta_w)\, d\theta_w = N_e \int_{\theta_e} \frac{y_e(\theta_e)}{W} \frac{\xi}{\mu(\theta_e) - \xi} f_{\theta_e}(\theta_e)\, d\theta_e. \tag{52}$$

7.3 Quantitative Results

With the calibrated parameters, we now report our main results.

Optimal Tax Rates. The existing tax regime may well be suboptimal. We therefore ask, within the context of the model and given the planner's objective, what the optimal tax rate is. Our main finding is that the optimal effective tax rate on labor income decreases between 1980 to 2019, while the optimal effective tax rate on profit increases between 1980 to 2019.

Figure 3 plots the tax rates on labor and profits against income y_o. Given our estimated economies, the optimal labor income tax rate in 2019 is lower than in 1980, while the profit tax rate is higher in 2019. In particular, the tax rate is higher for the top profits. The profit tax for large, high profit firms also becomes less regressive, while there is no significant change in the progressivity of labor income tax.

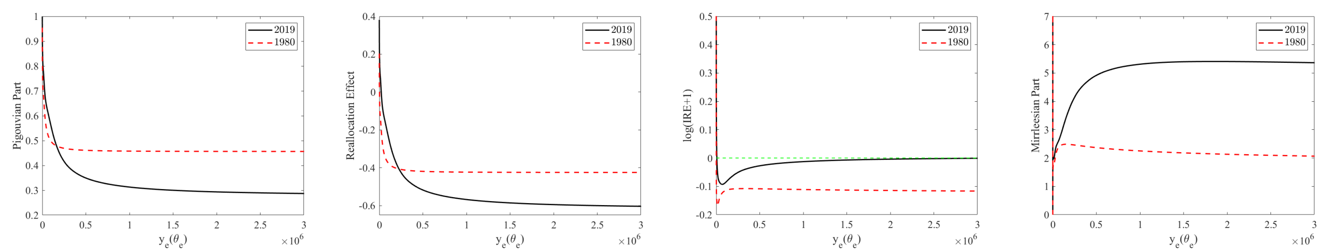
The optimal average labor income tax rate decreases from 21.4% in 1980 to 11.5% in 2019. Meanwhile,

[39] In Online Appendix OA.2, we derive the equilibrium solution as a function of parameters and tax rates. The equilibrium solution together with Theorem 1 implies the above findings.

the optimal average profit tax rate increases from 58.6% in 1980 to 61.3% in 2019.[40] When taxes transition from the initial level to the optimal level, the share of after-tax labor income in 1980 and 2019 increases by 16 and 20 percentage points, respectively.[41]

In addition to its role in redistribution, taxation also improves the efficiency of factor allocation among firms. The optimal taxation reallocates factors to high-markup firms and, as a result, improves allocative efficiency, but it also increases market concentration and the average markup.[42]

Decomposition of Optimal Taxation. The main reason for the decrease in the labor income tax rate is the rise in average markups, which reduces the Pigouvian part of the tax rule. The change in the profit tax rate is more intricate due to the different forces. In Figure 4, we decompose the profit tax rate into the Pigouvian part, the reallocation effect, the indirect redistribution effect, and the Mirrleesian part. The four elements are designed so that the optimal profit tax rate increases in all of the four components.



(a) Pigouvian Part (b) RE (c) IRE (d) Mirrleesian Part

Figure 4: Four Elements of the Profit Tax Wedge

Figure 3 establishes that the overall effect of the four components is an increase in the profit tax rate. In Figure 4 we observe that there is a *decline* in the top profit tax rate due to the Pigouvian part as well as the reallocation effect lower, whereas the indirect redistribution effect and the Mirrleesian part increase the top profit tax rate. Overall, the positive impact dominates. The Mirrleesian part contributes the majority to the increase of the top profit tax. It is worth noting that the Mirrleesian part is determined by the skill gap, which in turn depends on the ability distribution and the markup (see equation (33)). The markup enters the Mirrleesian part through the profit elasticity. Figure 5 demonstrates that the change in skill gap comes almost entirely from the changes in profit elasticity.

The solid lines in Figure 5 depict the logarithms of skill gap in 1980 and 2019. The dashed lines depict the logarithms of the inverse of profit elasticity in 1980 and 2019. Recall that the profit elasticity is a part of the skill gap (see equations (33) and (32)). The rise of skill gaps from 1980 to 2019 mainly originates from the decrease in profit elasticities because both patterns are remarkably similar. Since the change in profit elasticity is due to the change in markups (from equation (32)), this result suggests that rising markup generally increases the Mirrleesian part by decreasing the profit elasticity. That said, it is worth mentioning

[40] The after-tax labor income of θ_o refers to $y_o - T_o(y_o) - t_o$, where $t_o = -T_o(y_o(\underline{\theta}_o))$ is a lump-sum transfer (tax if negative). Then $T_o(y_o) = \int_{y_o(\underline{\theta}_o)}^{y_o} T_o'(y)\,dy - t_o$. We treat $\overline{y}_w = \int_{\underline{y}_w}^{\overline{y}_w} [y - T_w(y) - t_w] f_{y_w}(y)\,dy$ as the average after-tax labor income, and $\overline{T}_o = \frac{\int_{\underline{y}_o}^{\overline{y}_o} [T_o(y)+t_o] f_{y_o}(y_o)dy}{\int_{\underline{y}_o}^{\overline{y}_o} y_o f_{y_o}(y_o)dy}$ is the average tax rate, where $\underline{y}_o = y_o(\underline{\theta}_o)$ and $\overline{y}_o = y_o(\overline{\theta}_o)$.

[41] The original after-tax share of labor income to the total incomes in 1980 and 2019 are 39.8% and 36.4%. In the optimum, the after-tax share of labor income is 55.5% in 1980 and 56.5% in 2019. This is also the case considering transfer. The original after-tax-and-transfer share of labor income is 55.2% in 1980 and 52.3% in 2019, which are 75.0% and 75.3% in the optimum.

[42] The cost-weighted average markup in 2019 increases slightly from 1.375 in the initial economy to 1.382 in the optimum. In 1980, the increase is from 1.257 to 1.261.

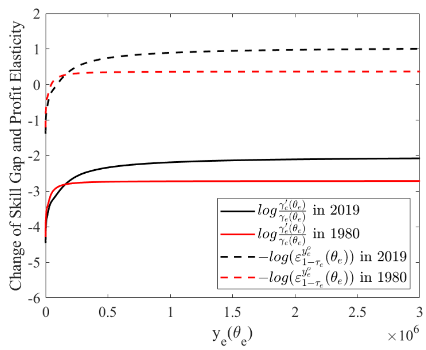


Figure 5: Skill Gap and Profit Elasticity

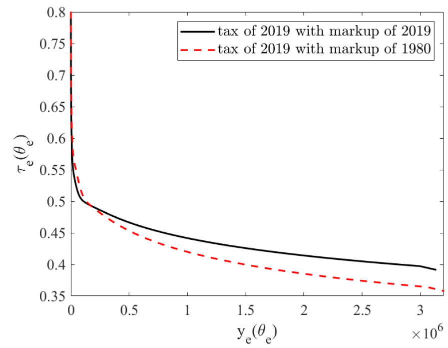


Figure 6: Counterfactual Profit Tax Wedges

that the markup enters all four factors.[43] Whether the rise of market power increases or decreases the optimal tax rate depends on all of these four components. It turns out that for the top profit tax rate, the impact of market forces on the Mirrleesian term is the most important. This means that as long as the government has a preference for equality, the recent changes in market powers require a higher top-profit tax rate.

Robustness. We consider $k \in \{0.7, 0.77, 1, 3\}$, $\xi \in \{0.4, 0.5, 0.6\}$, and $\sigma \in \{1.4, 1.3, 1.2\}$ to evaluate the robustness of our main results. Our choices of k covers common values used in the literature.[44] To guarantee the concavity of the entrepreneur's incentive problem, the value of the span of control in our model can not be too large, hence the range of values for ξ. Finally, the choice of σ can not be larger than 1.4, because the theoretical maximum value of firm-level markup decreases in σ. As the highest markup in our samples is 3.5, σ can not be higher than 1.4.

We find that our main conclusions hold within the broad parameter range mentioned above. That is, the labor income tax rates decrease and the profit tax rates increase. The optimal profit tax rate is insensitive to k. In all cases, the optimal average profit tax rate is around 58% in 1980 and 61% in 2019. The optimal average labor income tax rate increases in k. For $k = 0.7, 1$, and 3, it is 19.7%, 26.1%, and 44% in 1980; and 9.6%, 16.7%, and 37% in 2019. Figure OE2 in Online Appendix OE reports those results.

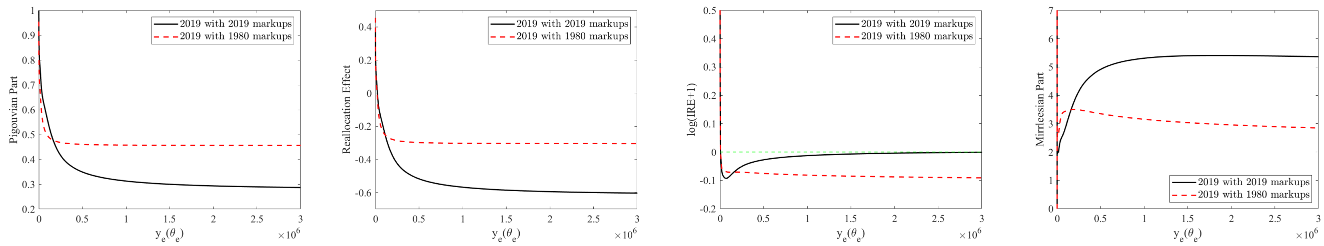
We also check the robustness of our result by changing the value of ξ and σ chosen for the calibration, while given markups. Note that changes in ξ and σ do not change $\tau_w(\cdot)$. However, it affects $T_w(\cdot)$ by changing $y_w(\cdot)$. In contrast, ξ and σ affect $T_e(\cdot)$ both directly and indirectly. Figure OE3 and OE4 in Online Appendix OE show that our conclusions hold, because the impact of ξ and σ on the optimal taxation of 1980 and 2019 works in the same direction.

Counterfactual Analysis and Policy Implications. To investigate how changes in the markups would affect the optimal profit tax wedges, we conduct a counterfactual analysis with respect to $\mu(\cdot)$. Specifically, we ask how the optimal tax rate in 2019 would change if the markups remains at the 1980 level? Figure 6 plots the profit tax wedges for 2019 with counterfactual 1980 level taxes, in addition to the 1980 and 2019 tax wedges. We see that taxes are lower for high profit incomes, and higher for the low profit incomes. This exercise isolates the effect of markups on the optimal tax rate from the effect changes in the productivity.

[43] The indirect redistribution effect also depends on the market structure because of the pricing power.

[44] For example, Saez (2001) considers $u = \log\left(c - \frac{l^{1-1/\varepsilon}}{1-1/\varepsilon}\right)$, which is equivalent to our case with $k = 1$. In their early version, Sachs et al. (2016) considered $k = 1, 3$.

As Figure 6 highlights, fixing markups to the lower 1980 level decreases the tax wedge for the high profit entrepreneurs. Next, we again decompose the elements of optimal tax rates to show how the top profit tax rate increases in this counterfactual exercise where only the markups vary and the productivity distribution is kept at the 2019 level.



(a) Pigouvian Part (b) RE (c) IRE (d) Mirrleesian Part

Figure 7: Counterfactual: Four Elements of Profit Tax Wedges in 2019 with 1980 Markups

Figure 7 decomposes the tax change for the counterfactual. With markups held at the 1980 level, the Pigouvian part and reallocation effect are larger for high incomes and lower for low incomes. These two elements require the optimal profit to be more regressive. However, the other two elements have opposite effects. The change in the Indirect Redistribution Effect increases the tax rates of low and middle types and decreases the tax rates of high types. While it significantly decreases the top profit tax with markups of 1980, it has minor effects on the the top tax rate in 2019 because the cross-inverse demand elasticity becomes smaller for the top. Finally, we see that the increase in the Mirrleesian part due to the increase of the markup is primarily responsible for the increase in the top profit tax.

Policy Implications. Our quantitative results provide three key *policy implications*. First, with the changing market structure that we have observed in recent years, the labor the income tax rate should be appropriately reduced, while the profit tax rate should be appropriately increased, especially for firms at the top of the distribution. Second, for large firms, the profit tax rate should be appropriately regressive to improve production efficiency through the reallocation effect. Third, while the optimal profit tax rate should be regressive for large firms, it becomes less regressive during the years that market power increases.

8 Conclusion

The most effective way to address market power is to eliminate the root cause with competition policy. In its absence, we ask what the role is for income taxation in addressing the inefficiency and inequality due to market power. In a standard partial equilibrium setting, taxing profits redistributes resources but does not affect optimal production. In a Mirrleesian setting, however, income and profit taxes do affect optimal production via the incentive constraint, endogenous labor supply, and the general equilibrium wage effect.

How should a policymaker design optimal taxation rules to balance the distributional and efficiency considerations, in an economy where incentives for production and market power interact? Optimal taxation cannot achieve the first-best, but it can improve welfare by enhancing the allocation efficiency while simultaneously redistributing income to the poor.

Our theory seamlessly merges the Mirrleesian approach to optimal income taxation with an Atkeson-Burstein inspired model of oligopoly pricing, which gives rise to a tractable framework. We derive tax wedges for labor and entrepreneurial income taxes that can be decomposed into four channels: (i) a Mirrleesian channel; (ii) a Pigouvian correction of the externality from market power; (iii) an indirect redistribution effect; and (iv) a reallocation effect towards more productive firms.

We conduct a detailed quantitative analysis estimating our model economy to match the key moments of the US economy in 1980 and 2019, a period of rising market power. Our estimates allow us to decompose the optimal tax rules into the four channels that we identify in the theory. Our main insights for policymakers are that optimal labor income taxes in 2019 are lower than in 1980 due to the rise of market power, and that optimal profit taxes are higher on average. We also find that optimal profit tax is regressive, in order to raise allocative efficiency, but less so in 2019 than in 1980, due to the trade-off between efficiency and equality.

Optimal income taxation in the presence of market power is far from first-best, yet optimal income taxation reduces inequality and incentivizes production by reducing taxes on labor, which increases the after-tax labor share. Meanwhile, profit tax rates are positive and increased to raise taxes for transfer. Last, policymakers should also use taxation to reduce misallocation between firms with low and high markups, which results in a relatively regressive profit tax to the labor income tax.

APPENDIX

A Environment

A.1 The Cournot Competitive Tax Equilibrium

When first-order conditions are both necessary and sufficient to both the individuals' and final good producer's problems, the equilibrium allocations are determined by (12) to (18) and the individuals' budget constraints. Under the technology considered in this paper and $\phi_o(l_o) = \frac{l_o^{1+\frac{1}{\varepsilon_o}}}{1+\frac{1}{\varepsilon_o}}$, we have the following conditions in the symmetric equilibrium:

1. First-order conditions

$$P(\theta_e) = [N_e f_e(\theta_e)]^{-\frac{1}{\sigma}} \zeta(\theta_e) A^{\frac{\sigma-1}{\sigma}} Q_{ij}(\theta_e)^{-\frac{1}{\sigma}} Q^{\frac{1}{\sigma}}, \tag{A1}$$

and

$$\left[1 + \frac{\partial \ln P_{ij}(Q_{ij}(\theta_e), Q_{-ij}(\theta_e), \theta_e)}{\partial \ln Q_{ij}(\theta_e)}\right] \frac{\xi P(\theta_e) Q_{ij}(\theta_e)}{L_w(\theta_e)} - \frac{W}{1-t_s} = 0 \tag{A2}$$

and

$$W x_w(\theta_w) \left[1 - T_w'(W x_w(\theta_w) l_w(\theta_w))\right] = l_w(\theta_w)^{\frac{1}{\varepsilon_w}}, \tag{A3}$$

and

$$\left[1 + \frac{\partial \ln P_{ij}(Q_{ij}(\theta_e), Q_{-ij}(\theta_e), \theta_e)}{\partial \ln Q_{ij}(\theta_e)}\right] P(\theta_e) Q_{ij}(\theta_e)(1-t_s)\left[1 - T_e'(y_e(\theta_e))\right] = l_e(\theta_e)^{1+\frac{1}{\varepsilon_e}}, \theta_o \in \Theta_o.$$

Combination of (A2) and (19) (i.e., $\mu(\theta_e) = \frac{P(\theta_e)}{W/\left[\frac{\partial Q_{ij}(\theta_e)}{\partial L_w(\theta_e)}(1-t_s)\right]} = \frac{\xi P(\theta_e) Q_{ij}(\theta_e)(1-t_s)}{W L_w(\theta_e)}$) delivers (20). Substituting $1 + \frac{\partial \ln P\left(Q_{ij}(\theta_e), \theta_e\right)}{\partial \ln Q_{ij}(\theta_e)}$ by (20), we have:

$$W L_w(\theta_e) = \frac{\xi(1-t_s)}{\mu(\theta_e)} P(\theta_e) Q_{ij}(\theta_e), \tag{A4}$$

and

$$\frac{P(\theta_e) Q_{ij}(\theta_e)(1-t_s)}{\mu(\theta_e)}\left[1 - T_e'(y_e(\theta_e))\right] = l_e(\theta_e)^{1+\frac{1}{\varepsilon_e}}, \theta_e \in \Theta_e. \tag{A5}$$

2. Inverse demand function

$$P_{ij}(Q_{ij}, Q_{-ij}(\theta_e), \theta_e) = \chi(\theta_e) A^{\frac{\sigma-1}{\sigma}} Q_{ij}^{-\frac{1}{\eta(\theta_e)}} I^{-\left[\frac{1}{\eta(\theta_e)} - \frac{1}{\sigma}\right]\frac{\eta(\theta_e)}{\eta(\theta_e)-1}} \begin{bmatrix} (I-1)\, Q_{-ij}(\theta_e)^{\frac{\eta(\theta_e)-1}{\eta(\theta_e)}} \\ +Q_{ij}^{\frac{\eta(\theta_e)-1}{\eta(\theta_e)}} \end{bmatrix}^{\left[\frac{1}{\eta(\theta_e)} - \frac{1}{\sigma}\right]\frac{\eta(\theta_e)}{\eta(\theta_e)-1}} \left[\frac{Q}{N}\right]^{\frac{1}{\sigma}}, \tag{A6}$$

where $Q_{ij}(\theta_e)$ is treated as given by the entrepreneurs.

3. Labor market clear condition

$$\int_{\theta_w} x_w(\theta_w)\, l_w(\theta_w)\, f_w(\theta_w) d\theta_w = W^{\varepsilon_w} \int_{\theta_w} \zeta(\theta_w)^{\varepsilon_w+1} \left[1-\tau_w(\theta_w)\right]^{\varepsilon_w} f_w(\theta_w) d\theta_w \tag{A7}$$

4. Meanwhile, in the equilibrium, we have:

$$Q = \int_{\theta_e} N_e f_e(\theta_e)\, P(\theta_e) Q_{ij}(\theta_e) d\theta_e. \tag{A8}$$

The above parts 1 to 4 solve the symmetric equilibrium allocation $\{L_w(\theta_e), l_e(\theta_e), l_w(\theta_w)\}$, price system $\{P(\theta_e), W\}$, and total output Q. Then one can derive other allocations with individuals' budget constraints.

A.2 Elasticities in the Equilibrium

A.2.1 Definitions

Price Elasticities. We define the elasticity of firm-level outputs with respect to the entrepreneurial effort $l_e(\theta_e)$ and labor inputs $L_w(\theta_e)$ respectively as:

$$\varepsilon_{l_e}^{Q_{ij}}(\theta_e) \equiv \frac{\partial \ln Q_{ij}(\theta_e)}{\partial \ln l_e(\theta_e)} \quad \text{and} \quad \varepsilon_{L_w}^{Q_{ij}}(\theta_e) \equiv \frac{\partial \ln Q_{ij}(\theta_e)}{\partial \ln L_w(\theta_e)}.$$

In the Atkeson-Burstein economy, $\varepsilon_{l_e}^{Q_{ij}}(\theta_e) = 1$ and $\varepsilon_{L_w}^{Q_{ij}}(\theta_e) = \xi$ are constants.

We define the sales elasticity and price elasticity respectively as

$$\varepsilon_{Q_{ij}}^{S}(\theta_e) \equiv \frac{\partial \ln \left[P_{ij}\left(Q_{ij}(\theta_e), Q_{ij}(\theta_e), \theta_e\right) Q_{ij}(\theta_e)\right]}{\partial \ln Q_{ij}(\theta_e)} \quad \text{and} \quad \varepsilon_{Q_{ij}}^{P}(\theta_e) \equiv \frac{\partial \ln P(\theta_e)}{\partial \ln Q_{ij}(\theta_e)}.$$

Define the own-inverse demand elasticity and cross-inverse demand elasticity as:

$$\varepsilon_{Q_{ij}}^{P,own}(\theta_e) \equiv \frac{\partial \ln P_{ij}\left(Q_{ij}, Q_{ij}(\theta_e), \theta_e\right)}{\partial \ln Q_{ij}}\Big|_{Q_{ij}=Q_{ij}(\theta_e)} \quad \text{and} \quad \varepsilon_{Q_{-ij}}^{P,cross}(\theta_e) \equiv \frac{\partial \ln P_{ij}\left(Q_{ij}, Q_{ij}(\theta_e), \theta_e\right)}{\partial \ln Q_{ij}(\theta_e)}\Big|_{Q_{ij}=Q_{ij}(\theta_e)}.$$

By definitions, $\varepsilon_{Q_{ij}}^{P}(\theta_e) = \varepsilon_{Q_{ij}}^{P,own}(\theta_e) + \varepsilon_{Q_{-ij}}^{P,cross}(\theta_e)$ and $\varepsilon_{Q_{ij}}^{S}(\theta_e) = 1 + \varepsilon_{Q_{ij}}^{P}(\theta_e)$. Moreover, notice that $\mu(\theta_e) = \frac{1}{1+\varepsilon_{Q_{ij}}^{P,own}(\theta_e)}$. We have

$$\varepsilon_{Q_{ij}}^{S}(\theta_e) = \frac{1}{\mu(\theta_e)} + \varepsilon_{Q_{-ij}}^{P,cross}(\theta_e), \quad \forall \theta_e \in \Theta_e. \tag{A9}$$

Following Sachs et al. (2020), we denote by $\varepsilon_{L_w}^{\varpi}(\theta_e', \theta_e)$ and $\varepsilon_{l_e}^{\varpi}(\theta_e', \theta_e)$ the cross elasticities of wage with respect to $L_w(\theta_e)$ and $l_e(\theta_e)$ for any $(\theta_e, \theta_e') \in \Theta_e^2$:

$$\varepsilon_{L_w}^{\varpi}(\theta_e', \theta_e) = \begin{cases} \frac{\partial \ln \varpi(\theta_e')}{\partial \ln L_w(\theta_e)}, \; \theta_e' \neq \theta_e, \\ \lim_{\theta_e' \to \theta_e} \frac{\partial \ln \varpi(\theta_e')}{\partial \ln L_w(\theta_e)}, \; \theta_e' = \theta_e \end{cases}$$

and

$$\varepsilon_{l_e}^{\varpi}(\theta_e',\theta_e) = \begin{cases} \frac{\partial \ln \varpi(\theta_e')}{\partial \ln l_e(\theta_e)}, \; \theta_e' \neq \theta_e, \\ \lim_{\theta_e' \to \theta_e} \frac{\partial \ln \varpi(\theta_e')}{\partial \ln l_e(\theta_e)}, \; \theta_e' = \theta_e. \end{cases}$$

We denote $\varepsilon_{L_w}^{\varpi}(\theta_e)$ and $\varepsilon_{l_e}^{\varpi}(\theta_e)$ as the own elasticities of wages with respect to $L_w(\theta_e)$ and $l_e(\theta_e)$. These own elasticities of wages are defined by the following relationships:

$$\frac{\partial \ln \varpi(\theta_e)}{\partial \ln L_w(\theta_e)} = \varepsilon_{L_w}^{\varpi}(\theta_e,\theta_e) + \varepsilon_{L_w}^{\varpi}(\theta_e)\delta\left(\theta_e' - \theta_e\right) \quad \text{and} \quad \frac{\partial \ln \varpi(\theta_e)}{\partial \ln l_e(\theta_e)} = \varepsilon_{l_e}^{\varpi}(\theta_e,\theta_e) + \varepsilon_{l_e}^{\varpi}(\theta_e)\delta\left(\theta_e' - \theta_e\right),$$

where δ denotes the Dirac delta function, $(\theta_e,\theta_e') \in \Theta_e^2$. See Appendix A.2.2 for details about the above elasticities.

A.2.2 Elasticities in the Atkeson-Burstein Economy

Wage Elasticity. Remember that in the Atkeson-Burstein economy, we have:

$$\begin{aligned} P_{ij}(\theta_e) &= N^{-\frac{1}{\sigma}} A^{\frac{\sigma-1}{\sigma}} \chi(\theta_e) Q_{ij}(\theta_e)^{-\frac{1}{\sigma}} Q^{\frac{1}{\sigma}}, \\ Q_{ij}(\theta_e) &= x_e(\theta_e)\, l_e(\theta_e)\, L_w(\theta_e)^{\xi}, \\ \varpi(\theta_e) &= \frac{\chi(\theta_e) N^{-\frac{1}{\sigma}} A^{\frac{\sigma-1}{\sigma}} Q_{ij}(\theta_e)^{-\frac{1}{\sigma}} Q^{\frac{1}{\sigma}}}{\mu(\theta_e)} \xi x_e(\theta_e)\, l_e(\theta_e)\, L_w(\theta)^{\xi-1}, \quad \forall \theta_e \in \Theta_e. \end{aligned}$$

It's easy to see that $\varepsilon_{L_w}^{\varpi}(\theta_e',\theta_e)$ and $\varepsilon_{l_e}^{\varpi}(\theta_e',\theta_e)$ are independent of θ_e'. By definition,

$$\varepsilon_{L_w}^{\varpi}(\theta_e) = \xi\left(1 - \frac{1}{\sigma}\right) - 1 < 0, \quad \text{and} \quad \varepsilon_{l_e}^{\varpi}(\theta) = 1 - \frac{1}{\sigma} > 0. \tag{A10}$$

Note that both $\varepsilon_{L_w}^{\varpi}(\theta_e)$ and $\varepsilon_{l_e}^{\varpi}(\theta)$ are constants.

Price Elasticity. Solving the final good producer's problem, we immediately have the price equation (A1) and the inverse demand function (A6). By the definitions of the price elasticities, we have the following results in this economy:

$$\begin{aligned} \varepsilon_{Q_{ij}}^{P}(\theta_e) &= -\frac{1}{\sigma}, \quad \varepsilon_{Q_{ij}}^{S}(\theta_e) = \frac{\sigma-1}{\sigma}, \\ \varepsilon_{Q_{ij}}^{P,own}(\theta_e) &= -\left[\frac{1}{\eta(\theta_e)}\frac{I-1}{I} + \frac{1}{\sigma}\frac{1}{I}\right], \\ \varepsilon_{Q_{-ij}}^{P,cross}(\theta_e) &= \left[\frac{1}{\eta(\theta_e)} - \frac{1}{\sigma}\right]\frac{I-1}{I}, \quad \forall \theta_e \in \Theta_e. \end{aligned}$$

Notice that $\mu(\theta_e) = \frac{1}{1+\varepsilon_{Q_{ij}}^{P,own}(\theta_e)}$, we have:

$$\varepsilon_{Q_{-ij}}^{P,cross}(\theta_e) = -\frac{1}{\mu(\theta_e)} + \frac{\sigma-1}{\sigma}, \quad \forall \theta_e \in \Theta_e. \tag{A11}$$

Under our production technology, we have

$$\begin{aligned}
&\frac{d\ln P_{ij}\left(Q_{ij},Q_{-ij}\left(\theta_e\right),\theta_e\right)}{d\theta_e}|_{Q_{ij}=Q_{ij}(\theta_e)} \qquad \text{(A12)}\\
=\;& \frac{d\ln P(\theta_e)}{d\theta_e}-\varepsilon^{P,own}_{Q_{ij}}\left(\theta_e\right)\frac{d\ln Q_{ij}\left(\theta_e\right)}{d\theta_e}\\
=\;& \frac{\chi'(\theta_e)}{\chi(\theta_e)}-\frac{1}{\sigma}\frac{Q'_{ij}\left(\theta_e\right)}{Q_{ij}\left(\theta_e\right)}+\left[\frac{I-1}{I}\frac{1}{\eta\left(\theta_e\right)}+\frac{1}{I}\frac{1}{\sigma}\right]\frac{Q'_{ij}\left(\theta_e\right)}{Q_{ij}\left(\theta_e\right)}\\
=\;& \frac{\chi'(\theta_e)}{\chi(\theta_e)}+\varepsilon^{P,cross}_{Q_{-ij}}\left(\theta_e\right)\frac{d\ln Q_{ij}\left(\theta_e\right)}{d\theta_e}, \quad \forall\theta_e\in\Theta_e.
\end{aligned}$$

Specially, when $I=1$, we have $\frac{d\ln P_{ij}\left(Q_{ij},Q_{-ij}(\theta_e),\theta_e\right)}{d\theta_e}|_{Q_{ij}=Q_{ij}(\theta_e)}=\frac{\chi'(\theta_e)}{\chi(\theta_e)}$.

Elasticity of Profit. Consider a small increase (i.e., $d\tau$) in the marginal tax rate faced by the θ_e-type entrepreneur. In this case, the tax reform has no first-order effects on the aggregate values and the actions of other types. Therefore, aggregate variables including outputs of final goods and price of labor factors are unchanged. As in Scheuer and Werning (2017), the elasticity derived here is a micro elasticity.

The optimal choice of the θ_e-type entrepreneur (i.e., l_e and L_w) satisfy the following first-order conditions in the equilibrium:

$$WL_w=\left(1-t_s\right)\frac{P_{ij}Q_{ij}}{\mu\left(\theta_e\right)}\frac{\partial\ln Q_{ij}}{\partial\ln L_w}, \tag{A13}$$

and

$$\phi'_e\left(l_e\right)=\left[1-T'_e(P_{ij}Q_{ij}\left(1-t_s\right)-WL_w)-d\tau\right]\frac{P_{ij}Q_{ij}\left(1-t_s\right)}{l_e\mu\left(\theta_e\right)}, \tag{A14}$$

where P_{ij} and Q_{ij} refere to $P_{ij}\left(Q_{ij},Q_{-ij}\left(\theta_e\right),\theta_e\right)$ and $Q_{ij}\left(x_e(\theta_e)l_e,L_w\right)$, respectively. Note that $Q_{ij}\left(\theta_e\right)$ in $P_{ij}\left(Q_{ij},Q_{-ij}\left(\theta_e\right),\theta_e\right)$ is treated as given by the agent. Equation (A13) is derived by (16) and (20), and equation (A14) is derived by (18) and (20).

The cases we considered has constant $\frac{\partial\ln Q_{ij}}{\partial\ln L_w}$ and exogenous $\mu\left(\theta_e\right)$. Set $\xi=\frac{\partial\ln Q_{ij}}{\partial\ln L_w}$. Combination of (A13) and (A14) gives:

$$WL_w=\left(1-t_s\right)\frac{P_{ij}Q_{ij}}{\mu\left(\theta_e\right)}\xi, \tag{A15}$$

and

$$\phi'_e\left(l_e\right)=\left[1-T'_e\left(\left(\frac{\mu\left(\theta_e\right)}{\xi}-1\right)WL_w\right)-d\tau\right]\frac{WL_w}{\xi}\frac{1}{l_e}. \tag{A16}$$

The θ_e-type entrepreneur's reaction to the tax reform can be described by differential equations of the first-order conditions. Total differential of (A16) gives:

$$\frac{1+\varepsilon_e}{\varepsilon_e}\frac{dl_e}{l_e}=\frac{dL_w}{L_w}\left[1-\frac{y_eT''_e(y_e)}{1-T'_e(y_e)}\right]-\frac{d\tau}{1-T'_e(y_e)}. \tag{A17}$$

Total differential of (A13) gives:

$$\frac{dL_w}{L_w}=\frac{1}{\mu\left(\theta_e\right)}\left[\frac{dl_e}{l_e}+\xi\frac{dL_w}{L_w}\right].$$

Note that $Q_{-ij}(\theta_e)$ also changes with the tax reform, which is captured by $\varepsilon_{Q_{-ij}}^{P,cross}(\theta_e)$.

A combination of the above two equations gives

$$-\frac{\frac{dL_w}{L_w}}{\frac{d\tau}{1-T_e'(y_e)}} = \frac{1}{\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu(\theta_e)-\xi\right]-\left[1-\frac{y_e T_e''(y_e)}{1-T_e'(y_e)}\right]}. \tag{A18}$$

Notice that $y_e = PQ_{ij}(1-t_s) - WL_w$ and $WL_w = (1-t_s)PQ_{ij}\frac{\xi}{\mu(\theta_e)}$. We have $y_e = WL_w\left(\frac{\mu(\theta_e)}{\xi}-1\right)$ and

$$\frac{dy_e}{y_e} = \frac{dL_w}{L_w}. \tag{A19}$$

Define $\varepsilon_{1-\tau_e}^{y_e}(\theta_e) \equiv -\frac{dy_e(\theta_e)}{y_e(\theta_e)} / \frac{d\tau}{1-T_e'(y_e(\theta_e))}$ as the non-linear profit elasticity. We have:

$$\varepsilon_{1-\tau_e}^{y_e}(\theta_e) = \frac{1}{\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu(\theta_e)-\xi\right]-\left[1-\frac{y_e(\theta_e)T_e''(y_e(\theta_e))}{1-T_e'(y_e(\theta_e))}\right]}. \tag{A20}$$

General Elasticity of Profit. Again, consider a small increase (i.e., $d\tau$) in the marginal tax rate faced by the θ_e-type entrepreneur. The tax reform has no first-order effects on the aggregate values and the actions of other types. Aggregate variables including outputs of final goods and price of labor factors are unchanged. However, unlike the profit elasticity considered in the previous, now we consider a profit elasticity in the general equilibrium, where the change of $Q_{-ij}(\theta_e)$ should be taken into consideration. Set $\tilde{\varepsilon}_{1-\tau_e}^{y_e}(\theta_e) \equiv -\frac{dy_e(\theta_e)}{y_e(\theta_e)} / \frac{d\tau}{1-T_e'(y_e(\theta_e))}$ as the elasticity of profit in this case. We call $\tilde{\varepsilon}_{1-\tau_e}^{y_e}(\theta_e)$ the *general elasticity of profit* and $\varepsilon_{1-\tau_e}^{y_e}(\theta_e)$ the *partial elasticity of profit* to distinguish them from each other.

Still, we have (A15) and (A16). However, when taking total differential, one should include the change of $Q_{-ij}(\theta_e)$. Notice that $Q_{-ij}(\theta_e) = Q_{ij}(\theta_e)$. Total differential of (A15) delivers:

$$\begin{aligned}
d\ln L_w &= \left[\frac{d\ln P_{ij}\left(Q_{ij}, Q_{ij}(\theta_e), \theta_e\right)}{d\ln Q_{ij}}+1\right] d\ln Q_{ij} \\
&= \left[1+\varepsilon_{Q_{ij}}^{P,own}(\theta_e)+\varepsilon_{Q_{-ij}}^{P,cross}(\theta_e)\right]\left[d\ln l_e + \xi d\ln L_w\right] \\
&= \left[\frac{1}{\mu(\theta_e)}+\varepsilon_{Q_{-ij}}^{P,cross}(\theta_e)\right]\left[d\ln l_e + \xi d\ln L_w\right] \\
&= \varepsilon_{Q_{ij}}^{S}(\theta_e)\left[d\ln l_e + \xi d\ln L_w\right]
\end{aligned}$$

i.e.,

$$\frac{dl_e}{l_e} = \frac{dL_w}{L_w}\frac{1-\varepsilon_{Q_{ij}}^{S}(\theta_e)\xi}{\varepsilon_{Q_{ij}}^{S}(\theta_e)},$$

where $\varepsilon_{Q_{ij}}^{S}(\theta_e) = \frac{1}{\mu(\theta_e)} + \varepsilon_{Q_{-ij}}^{P,cross}(\theta_e) = \frac{\sigma-1}{\sigma}$.

$$\frac{dl_e}{l_e} = \frac{dL_w}{L_w}\left[\frac{1}{\varepsilon_{Q_{ij}}^{S}(\theta_e)}-\xi\right],$$

On the other hand, we still have (A17). Substitute $\frac{dl_e}{l_e}$ in (A17) by the above equation:

$$\tilde{\varepsilon}^{y_e}_{1-\tau_e}(\theta_e) = \frac{1}{\frac{1+\varepsilon_e}{\varepsilon_e}\frac{\partial \ln Q_{ij}(\theta_e)}{\partial \ln l_e(\theta_e)}\left[\frac{1}{\varepsilon^S_{Q_{ij}}(\theta_e)\frac{\partial \ln Q_{ij}(\theta_e)}{\partial \ln l_e(\theta_e)}} - \xi\right] - \left[1 - \frac{y_e(\theta_e)T''_e(y_e(\theta_e))}{1-T'_e(y_e(\theta_e))}\right]}. \tag{A21}$$

The general profit elasticity shows to what extent the marginal profit tax reform changes the firm-level profits by changing the firm's own decisions and its competitors' outputs, while the partial elasticity reflects the reaction of the firm to the tax reform.

B Solution

B.1 Proof of Lemma 1

To simplify notation, in the following analysis, we set $P\left(Q_{ij}, \theta_e\right) = P_{ij}\left(Q_{ij}, Q_{-ij}\left(\theta_e\right), \theta_e\right)$ for any $\theta_e \in \Theta_e$ and $Q_{ij} \in \mathbb{R}_+$.

We first prove part (i) of Lemma 1 (i.e., given (A22), (A26) is satisfied if and only if (29) is satisfied). According to the definition of $V_e(\theta)$, we have

$$V_e(\theta_e) = c_e\left(\theta_e\right) - \phi_e\left(l_e(\theta_e)\right), \quad \forall \theta_e \in \Theta_e, \tag{A22}$$

Notice that

$$V_e(\theta'_e|\theta_e) = c_e\left(\theta'_e\right) - \phi_e\left(l_e\left(\theta'_e|\theta_e\right)\right),$$

where $l_e\left(\theta'_e|\theta_e\right)$ is the effort needed to finish the θ'_e task:

$$l_e\left(\theta'_e|\theta_e\right) = \arg\min_{l_e, L_w}\left\{l_e | P\left(Q_{ij}\left(x_e\left(\theta_e\right)l_e, L_w\right), Q_{-ij}\left(\theta_e\right), \theta_e\right) \cdot Q_{ij}\left(x_e\left(\theta_e\right)l_e, L_w\right)\left(1-t_s\right) - WL_w = y_e\left(\theta'_e\right)\right\}. \tag{A23}$$

Obviously, $P\left(Q_{ij}\left(x_e\left(\theta_e\right)l_e, L_w\right), Q_{-ij}\left(\theta_e\right), \theta_e\right) \cdot Q_{ij}\left(x_e\left(\theta_e\right)l_e, L_w\right)\left(1-t_s\right) - WL_w$ increases in l_e. Denote by $L_w\left(\theta_e|l_e\right)$ the solution to

$$L_w\left(\theta_e|l_e\right) = \arg\max_{L_w}\left\{P\left(Q_{ij}\left(x_e\left(\theta_e\right)l_e, L_w\right), Q_{-ij}\left(\theta_e\right), \theta_e\right) \cdot Q_{ij}\left(x_e\left(\theta_e\right)l_e, L_w\right)\left(1-t_s\right) - WL_w\right\}$$

for $l_e > 0$. In Online Appendix OB.2, we show that for any $l_e > 0$, the first-order condition for solving $L_w\left(\theta_e|l_e\right)$ is not only necessary but also sufficient and there is a unique solution. Meanwhile, for $L_w > 0$,

$$P\left(Q_{ij}\left(x_e\left(\theta_e\right)l_e, L_w\right), Q_{-ij}\left(\theta_e\right), \theta_e\right) Q_{ij}\left(x_e\left(\theta_e\right)l_e, L_w\right)\left(1-t_s\right) - WL_w$$

strictly increases in l_e. Therefore, there must exist a unqiue solution to problem (A23). Denote by $L_w\left(\theta'_e|\theta_e\right)$ the optimal labor input given that θ_e entrepreneur reports θ'_e. When $y_e\left(\theta'_e\right) = 0$, $L_w\left(\theta'_e|\theta_e\right) = l_e\left(\theta'_e|\theta_e\right) = 0$. Otherwise, $L_w\left(\theta'_e|\theta_e\right) > 0$ and $l_e\left(\theta'_e|\theta_e\right) > 0$ are determined by the first-order conditions. In particular,

$L_w\left(\theta_e'|\theta_e\right)$ and $l_e\left(\theta_e'|\theta_e\right)$ satisfy:

$$\begin{aligned}
& P\left(Q_{ij}\left(x_e\left(\theta_e\right)l_e\left(\theta_e'|\theta_e\right), Q_{-ij}\left(\theta_e\right), L_w\left(\theta_e'|\theta_e\right)\right), \theta_e\right)\cdot Q_{ij}\left(x_e\left(\theta_e\right)l_e\left(\theta_e'|\theta_e\right), L_w\left(\theta_e'|\theta_e\right)\right)\left(1-t_s\right) \\
= \; & WL_w\left(\theta_e'|\theta_e\right) + y_e\left(\theta_e'\right),
\end{aligned} \tag{A24}$$

Equation (A24) and problem (A23) implies:

$$\begin{aligned}
\frac{\partial l_e\left(\theta_e'|\theta_e\right)}{\partial\theta_e} &= -\frac{\frac{\partial\left[P_{ij}\left(Q_{ij},Q_{-ij}(\theta_e),\theta_e\right)Q_{ij}\right]}{\partial Q_{ij}}\frac{\partial Q_{ij}}{\partial\theta_e} + \frac{dP_{ij}\left(Q_{ij},Q_{-ij}(\theta_e),\theta_e\right)}{d\theta_e}Q_{ij}}{\frac{\partial\left[P_{ij}\left(Q_{ij},Q_{-ij}(\theta_e),\theta_e\right)Q_{ij}\right]}{\partial Q_{ij}}\frac{\partial Q_{ij}}{\partial l_e}} \\
&= -\frac{\frac{\partial Q_{ij}}{\partial\theta_e} + \frac{\frac{dP_{ij}\left(Q_{ij},Q_{-ij}(\theta_e),\theta_e\right)}{d\theta_e}Q_{ij}}{P\left(Q_{ij},\theta_e\right)\left[1+\frac{\partial\ln P_{ij}\left(Q_{ij},Q_{-ij}(\theta_e),\theta_e\right)}{\partial\ln Q_{ij}}\right]}}{\frac{\partial Q_{ij}}{\partial l_e}} \\
&= -\left[\frac{x_e'\left(\theta_e\right)}{x_e\left(\theta_e\right)} + \frac{\frac{d\ln P_{ij}\left(Q_{ij},Q_{-ij}(\theta_e),\theta_e\right)}{d\theta_e}}{1+\frac{\partial\ln P_{ij}\left(Q_{ij},Q_{-ij}(\theta_e),\theta_e\right)}{\partial\ln Q_{ij}}}\right] l_e\left(\theta_e'|\theta_e\right) < 0.
\end{aligned} \tag{A25}$$

The first-order incentive condition ($\frac{\partial V_e(\theta_e'|\theta_e)}{\partial\theta_e'}|_{\theta_e'=\theta_e} = 0$) can be expressed as

$$0 = \left[c_e'\left(\theta_e'\right) - \phi_e'\left(l_e\left(\theta_e'|\theta_e\right)\right)\frac{\partial l_e\left(\theta'|\theta\right)}{\partial\theta'}\right]|_{\theta'=\theta}, \quad \forall\theta_e\in\Theta_e. \tag{A26}$$

First, note that by

$$V_e(\theta_e) = \max_{\theta_e'} V_e(\theta_e'|\theta_e),$$

we have

$$V_e'(\theta_e) = \frac{\partial V_e(\theta_e^*\left(\theta_e\right)|\theta_e)}{\partial\theta_e^*\left(\theta_e\right)}\frac{d\theta_e^*\left(\theta_e\right)}{d\theta_e} + \frac{\partial V_e(\theta_e^*\left(\theta_e\right)|\theta_e)}{\partial\theta_e} \tag{A27}$$

where we use $\theta_e^*\left(\theta_e\right)$ to denote the optimal choice of θ_e entrepreneur.

Second, by the definition of $V_e(\theta_e'|\theta_e)$, we have

$$\frac{\partial V_e(\theta_e^*\left(\theta_e\right)|\theta_e)}{\partial\theta_e} = -\phi_e'\left(l_e\left(\theta_e^*\left(\theta_e\right)|\theta\right)\right)\frac{\partial l_e\left(\theta_e^*\left(\theta_e\right)|\theta_e\right)}{\partial\theta_e}, \tag{A28}$$

where by (A25), we have

$$\begin{aligned}
\frac{\partial l_e\left(\theta_e^*\left(\theta_e\right)|\theta_e\right)}{\partial\theta_e} &= -\frac{x_e'\left(\theta_e\right)}{x_e\left(\theta_e\right)}l_e\left(\theta_e^*\left(\theta_e\right)|\theta_e\right) \\
&\quad -\frac{\partial\ln P\left(Q_{ij}\left(x_e\left(\theta_e\right)l_e\left(\theta_e^*\left(\theta_e\right)|\theta_e\right), L_w\left(\theta_e^*\left(\theta_e\right)|\theta_e\right)\right), \theta_e\right)}{\partial\theta_e}\frac{l_e\left(\theta_e^*\left(\theta_e\right)|\theta_e\right)}{1+\varepsilon_{Q_{ij}}^{P,own}\left(\theta_e\right)}.
\end{aligned} \tag{A29}$$

Combining (A27), (A28), and (A29) gives

$$V_e'(\theta_e) = \frac{\partial V_e(\theta_e^*(\theta_e)|\theta_e)}{\partial \theta_e^*(\theta_e)} \frac{d\theta_e^*(\theta_e)}{d\theta_e} - \phi_e'(l_e(\theta_e^*(\theta_e)|\theta)) \left[\begin{array}{c} -\frac{x_e'(\theta_e)}{x_e(\theta_e)} l_e(\theta_e^*(\theta_e)|\theta_e) - \\ \frac{\partial \ln P\left(Q_{ij}(x_e(\theta_e)l_e(\theta_e^*(\theta_e)|\theta_e), L_w(\theta_e^*(\theta_e)|\theta_e)), \theta_e\right)}{\partial \theta_e} \frac{l_e(\theta_e^*(\theta_e)|\theta_e)}{1+\varepsilon_{Q_{ij}}^{P,own}(\theta_e)} \end{array} \right]$$

which implies that for any $\theta_e \in \Theta$,

$$V_e'(\theta_e) = \phi_e'(l_e(\theta_e^*(\theta_e)|\theta_e)) l_e(\theta_e^*(\theta_e)|\theta_e) \left[\begin{array}{c} \frac{x_e'(\theta_e)}{x_e(\theta_e)} + \mu(\theta_e) \\ \times \frac{d \ln P_{ij}\left(Q_{ij}, Q_{-ij}(\theta_e), \theta_e\right)}{d\theta_e} |_{Q_{ij}=Q_{ij}(x_e(\theta_e)l_e(\theta_e^*(\theta_e)|\theta_e), L_w(\theta_e^*(\theta_e)|\theta_e))} \end{array} \right] \tag{A30}$$

if and only if $\frac{\partial V_e(\theta_e^*(\theta_e)|\theta_e)}{\partial \theta_e^*(\theta_e)} \frac{d\theta_e^*(\theta_e)}{d\theta_e} = 0$.

Notice that mass points are ruled out by Assumption 1: $\frac{\partial V_e(\theta_e^*(\theta_e)|\theta_e)}{\partial \theta_e^*(\theta_e)} \frac{d\theta_e^*(\theta_e)}{d\theta_e} = 0$ if and only if $\frac{\partial V_e(\theta_e^*(\theta_e)|\theta_e)}{\partial \theta_e^*(\theta_e)} = 0$. Therefore, (29), i.e., (A30) when $\theta_e^*(\theta_e) = \theta_e$, implies the first-order necessary condition $\frac{\partial V_e(\theta_e'|\theta_e)}{\partial \theta_e'}|_{\theta_e'=\theta_e} = 0$; and the first-order necessary condition $\frac{\partial V_e(\theta_e'|\theta_e)}{\partial \theta_e'}|_{\theta_e'=\theta_e} = 0$, i.e., $\frac{\partial V_e(\theta_e^*(\theta_e)|\theta_e)}{\partial \theta_e^*(\theta_e)}|_{\theta_e^*(\theta_e)=\theta_e} = 0$, implies (29) when the agent reports the true types. In conclusion, (29) is a necessary condition for the truth-telling strategy to be the optimum choice of agents and it implies $\frac{\partial V_e(\theta_e'|\theta_e)}{\partial \theta_e'}|_{\theta_e'=\theta_e} = 0$.■

B.2 Optimal Taxation

B.2.1 Lagrangian and First-order Conditions

We now take Lagrange multipliers to solve the planner's optimization problem.[45] The Lagrangian function for the planner's problem is:

$$\begin{aligned}
&£\left(L_w, l_w, l_e, V_w, V_e, \delta; \lambda, \lambda', \psi_w, \psi_e, \kappa, \varphi\right) \\
=\ & \sum_{o\in\{w,e\}} N_o \int_{\theta_o} G(V_o(\theta_o)) \widetilde{f}_o(\theta_o) d\theta_o + \lambda \left[Q - \sum_{o\in\{w,e\}} N_o \int_{\theta_o} [V_o(\theta_o) + \phi_o(l_o(\theta_o))] f_o(\theta_o) d\theta_o - R \right] \\
& + \lambda' \left[N_w \int_{\theta_w} x_w(\theta_w) l_w(\theta_w) f_w(\theta_w) d\theta_w - N_e \int_{\theta_e} L_w(\theta_e) f_e(\theta_e) d\theta_e \right] + \int_{\theta_e} \varphi(\theta_e) \frac{d \ln \varpi(\theta_e)}{d\theta_e} d\theta_e \\
& + \int_{\theta_e} \kappa(\theta_e) \left[\delta(\theta_e) - \frac{d \ln Q_{ij}(\theta_e)}{d\theta_e} \right] d\theta_e + \int_{\theta_w} \psi_w(\theta_w) \left[l_w(\theta_w) \phi_w'(l_w(\theta_w)) \frac{x_w'(\theta_e)}{x_w(\theta_e)} - V_w'(\theta_w) \right] d\theta_w \\
& + \int_{\theta_e} \psi_e(\theta_e) \left[\phi_e'(l_e(\theta_e)) l_e(\theta_e) \left[\frac{x_e'(\theta_e)}{x_e(\theta_e)} + \mu(\theta_e) \left(\frac{\chi'(\theta_e)}{\chi(\theta_e)} + \varepsilon_{Q_{-ij}}^{P,cross}(\theta) \delta(\theta_e) \right) \right] - V_e'(\theta_e) \right] d\theta_e,
\end{aligned}$$

where $\frac{\chi'(\theta_e)}{\chi(\theta_e)} + \varepsilon_{Q_{-ij}}^{P,cross}(\theta_e) \delta(\theta_e) = \frac{d \ln P_{ij}\left(Q_{ij}, Q_{-ij}(\theta_e), \theta_e\right)}{d\theta_e}|_{Q_{ij}=Q_{ij}(\theta_e)}$, and $\varpi(\theta_e) = \frac{\partial\left[P_{ij}\left(Q_{ij}(\theta_e), Q_{-ij}(\theta_e), \theta_e\right) Q_{ij}(\theta_e)\right]}{\partial L_w(\theta_e)}$. Note that we have introduced $\delta(\theta_e) = \frac{d \ln Q_{ij}(\theta_e)}{d\theta_e}$ as a control value and that $\ln Q_{ij}(\theta_e)$ can be treated as a state variable. Constraint $\frac{d \ln \varpi(\theta_e)}{d\theta_e} = 0$ is used to guarantee that $\varpi(\theta_e) = \frac{\partial\left[P_{ij}\left(Q_{ij}(\theta_e), Q_{-ij}(\theta_e), \theta_e\right) Q_{ij}(\theta_e)\right]}{\partial L_w(\theta_e)}$ (equivalently $\varpi(\theta_e) = \frac{P(\theta_e)}{\mu(\theta_e)} \frac{\partial Q_{ij}(\theta_e)}{\partial L_w(\theta_e)}$) is constant, which is a result of uniform sales taxes on the goods produced by firms.

[45]See Luenberger (1997) for details about the Lagrangian techniques, and Mirrlees (1976), Golosov et al. (2016), Findeisen and Sachs (2017) for its application in the field of public economics.

Taking partial integrals yields the following

$$-\int_{\theta_e} \kappa(\theta_e) \frac{d \ln Q_{ij}(\theta_e)}{d\theta_e} d\theta_e = \ln Q_{ij}(\underline{\theta}_e) \kappa(\underline{\theta}_e) - \ln Q_{ij}(\overline{\theta}_e) \kappa(\overline{\theta}_e) + \int_{\theta_e} \kappa'(\theta_e) \ln Q_{ij}(\theta_e) d\theta_e,$$

and

$$\int_{\theta_e} \varphi(\theta_e) \frac{d \ln \varpi(\theta_e)}{d\theta_e} d\theta_e = \varphi(\overline{\theta}_e) \ln \varpi(\overline{\theta}_e) - \varphi(\underline{\theta}_e) \ln \varpi(\underline{\theta}_e) - \int_{\theta_e} \varphi'(\theta_e) \ln \varpi(\theta_e) d\theta_e,$$

and

$$-\int_{\theta_o} \psi_o(\theta_e) V_o'(\theta_o) d\theta_o = V_o(\underline{\theta}_o)\psi_o(\underline{\theta}_o) - V_o(\overline{\theta}_o)\psi_o(\overline{\theta}_o) + \int_{\theta_o} \psi_o'(\theta_o) V_o(\theta_o) d\theta_o.$$

The derivatives with respect to the endpoint conditions yield boundary conditions:

$$\kappa(\underline{\theta}_e) = \kappa(\overline{\theta}_e) = \varphi(\overline{\theta}_e) = \varphi(\underline{\theta}_e) = \psi_o(\underline{\theta}_o) = \psi_o(\overline{\theta}_o) = 0, \quad o \in \{w, e\}. \tag{A31}$$

Thus,

$$\int_{\theta_e} \varphi'(\theta_e) d\theta_e = 0, \tag{A32}$$

Substituting the above conditions into the Lagrangian function yields the following first-order conditions:

$$\frac{\partial \pounds}{\partial V_o(\theta_o)} = G'(V_o(\theta_o)) N_o \widetilde{f}_o(\theta_o) + \psi_o'(\theta_o) - \lambda N_o f_o(\theta_o) = 0, \tag{A33}$$

$$\frac{\partial \pounds}{\partial \delta(\theta_e)} = \kappa(\theta_e) + \psi_e(\theta_e) \phi_e'(l_e(\theta_e)) l_e(\theta_e) \mu(\theta_e) \varepsilon_{Q-ij}^{P,cross}(\theta_e) = 0, \tag{A34}$$

$$\frac{\partial \pounds}{\partial l_w(\theta_w)} = \left[-\lambda \phi_w'(l_w(\theta_w)) + \lambda' x_w(\theta_w)\right] N_w f_w(\theta_w) + \psi_w(\theta_w) \frac{\phi_w'(l_w(\theta_w))}{x_w(\theta_w)} \frac{1+\varepsilon_w}{\varepsilon_w} = 0, \tag{A35}$$

$$\frac{\partial \pounds}{\partial L_w(\theta_e)} = \left[\lambda P(\theta_e) \frac{\partial Q_{ij}(\theta_e)}{\partial L_w(\theta_e)} - \lambda'\right] N_e f_e(\theta_e) + \left[\begin{array}{c} \frac{\kappa'(\theta_e)}{L_w(\theta_e)} \frac{\partial \ln Q_{ij}(\theta_e)}{\partial \ln L_w(\theta_e)} \\ -\frac{\int_{\Theta_e} \varphi'(\theta_e') \frac{\partial \ln \varpi(\theta_e')}{\partial \ln L_w(\theta_e)} d\theta_e'}{L_w(\theta_e)} \end{array} \right] = 0, \tag{A36}$$

and

$$\begin{aligned} \frac{\partial \pounds}{\partial l_e(\theta_e)} &= \psi_e(\theta_e) \phi_e'(l_e(\theta_e)) \frac{1+\varepsilon_e}{\varepsilon_e} \left[\mu(\theta_e) \frac{d \ln P_{ij}(Q_{ij}, Q_{-ij}(\theta_e), \theta_e)}{d\theta_e}|_{Q_{ij}=Q_{ij}(\theta_e)} + \frac{x_e'(\theta_e)}{x_e(\theta_e)}\right] \\ &+ \lambda \left[P(\theta_e) \frac{\partial Q_{ij}(\theta_e)}{\partial l_e(\theta_e)} - \phi_e'(l_e(\theta_e))\right] N_e f_e(\theta_e) \\ &+ \frac{\kappa'(\theta_e)}{l_e(\theta_e)} \frac{\partial \ln Q_{ij}(\theta_e)}{\partial \ln l_e(\theta_e)} - \frac{\int_{\Theta} \varphi'(\theta_e') \frac{\partial \ln \varpi(\theta_e')}{\partial \ln l_e(\theta_e)} d\theta_e'}{l_e(\theta_e)} = 0, \quad \forall \theta_o \in \Theta_o. \end{aligned} \tag{A37}$$

B.2.2 Social Welfare Weight

Unless otherwise specified, the following equations in this subsection are derived for any $\theta_o \in \Theta_o$. According to $\frac{\partial \pounds}{\partial V_o(x)}$ and $\psi_o(\underline{\theta}_o) = \psi_o(\overline{\theta}_o) = 0$, we have:

$$\lambda = \int_{\theta_o} G'(V_o(\theta_o)) \widetilde{f}_o(\theta_o) d\theta_o. \tag{A38}$$

Set

$$g_o(\theta_o) = \frac{G'(V_o(\theta_o))\widetilde{f}_o(\theta_o)}{\lambda f_o(\theta_o)} \tag{A39}$$

as the monetary marginal social welfare weight for θ_o agent of o occupation. Set:

$$\bar{g}_o(\theta_o) = \frac{\int_{\theta_o}^{\overline{\theta}_o} g(x) f_o(x)\,dx}{1 - F_o(\theta_o)} \tag{A40}$$

as the weighted monetary social welfare weight for agents whose abilities are higher than θ_e.

Substituting $g_o(\theta_o)$ into $\frac{\partial \pounds}{\partial V_o(\theta_o)}$ gives

$$\frac{\psi_o'(\theta_o)}{\lambda N_o f_o(\theta_o)} = 1 - g_o(\theta_o) \tag{A41}$$

Taking integration and using the boundary conditions gives

$$\begin{aligned} -\frac{\psi_o(\theta_o)}{\lambda N_o} &= \int_{\theta_o}^{\overline{\theta}_o} [1 - g_o(x)]\, f_o(x) dx \\ &= [1 - \bar{g}_o(\theta_o)]\,[1 - F_o(\theta_o)]\,. \end{aligned} \tag{A42}$$

In addition, based on $\frac{\partial \pounds}{\partial \delta(\theta_e)}$, we have:

$$\begin{aligned} \kappa(\theta_e) &= -\psi_e(\theta_e)\,\phi_e'(l_e(\theta_e))\,l_e(\theta_e)\,\mu(\theta_e)\varepsilon_{Q-ij}^{P,cross}(\theta_e) \\ &= -\psi_e(\theta_e)\,P(\theta_e)\,Q_{ij}(\theta_e)\,[1 - \tau_e(\theta_e)]\,(1 - \tau_s)\,\varepsilon_{Q-ij}^{P,cross}(\theta_e)\,, \end{aligned} \tag{A43}$$

where the second equation is derived by

$$\phi_e'(l_e(\theta_e))\,l_e(\theta_e) = \frac{P(\theta_e)\,Q_{ij}(\theta_e)}{\mu(\theta_e)}\,[1 - \tau_e(\theta_e)]\,(1 - \tau_s)\,. \tag{A44}$$

In addition, we have:

$$\begin{aligned} \kappa'(\theta_e) &= -\frac{d\left[\psi_e(\theta_e)\,\phi_e'(l_e(\theta_e))\,l_e(\theta_e)\,\mu(\theta_e)\varepsilon_{Q-ij}^{P,cross}(\theta_e)\right]}{d\theta_e} \\ &= -\begin{bmatrix} \psi_e'(\theta_e)\,\phi_e'(l_e(\theta_e))\,l_e(\theta_e)\,\mu(\theta_e)\varepsilon_{Q-ij}^{P,cross}(\theta_e) + \\ \psi_e(\theta_e)\,\phi_e'(l_e(\theta_e))\,\frac{1+\varepsilon_e}{\varepsilon_e} l_e'(\theta_e)\,\mu(\theta_e)\varepsilon_{Q-ij}^{P,cross}(\theta_e) + \\ \psi_e(\theta_e)\,\phi_e'(l_e(\theta_e))\,l_e(\theta_e)\,\frac{d\ln\left[\mu(\theta_e)\varepsilon_{Q-ij}^{P,cross}(\theta_e)\right]}{d\theta_e} \end{bmatrix} \\ &= -\phi_e'(l_e(\theta_e))\,l_e(\theta_e)\,\mu(\theta_e)\varepsilon_{Q-ij}^{P,cross}(\theta_e)\begin{bmatrix} \psi_e(\theta_e)\,\frac{1+\varepsilon_e}{\varepsilon_e}\frac{l_e'(\theta_e)}{l_e(\theta_e)} + \\ \psi_e'(\theta_e) + \psi_e(\theta_e)\,\frac{d\ln\left[\mu(\theta_e)\varepsilon_{Q-ij}^{P,cross}(\theta_e)\right]}{d\theta_e} \end{bmatrix} \\ &= -P(\theta_e)\,Q_{ij}(\theta_e)\,[1 - \tau_e(\theta_e)]\,(1 - \tau_s)\,\varepsilon_{Q-ij}^{P,cross}(\theta_e)\begin{bmatrix} \psi_e(\theta_e)\,\frac{1+\varepsilon_e}{\varepsilon_e}\frac{l_e'(\theta_e)}{l_e(\theta_e)} + \\ \psi_e'(\theta_e) + \psi_e(\theta_e)\,\frac{d\ln\left[\mu(\theta_e)\varepsilon_{Q-ij}^{P,cross}(\theta_e)\right]}{d\theta_e} \end{bmatrix}. \end{aligned} \tag{A45}$$

Substituting $\psi_e(\theta_e)$ and $\psi_e'(\theta_e)$ in (A45) and (A43) with (A41) and (A42), we have:

$$\begin{aligned}\kappa(\theta_e) &= \lambda N_e\left[1-\bar{g}_e(\theta_e)\right]\left[1-F_e(\theta_e)\right]\phi_e'\left(l_e(\theta_e)\right)l_e(\theta_e)\,\mu(\theta_e)\varepsilon_{Q-ij}^{P,cross}(\theta_e)\\ &= \lambda N_e\left[1-\bar{g}_e(\theta_e)\right]\left[1-F_e(\theta_e)\right]P(\theta_e)\,Q_{ij}(\theta_e)\left[1-\tau_e(\theta_e)\right](1-\tau_s)\,\varepsilon_{Q-ij}^{P,cross}(\theta_e),\end{aligned}$$

and

$$\frac{\kappa'(\theta_e)}{\lambda N_e f_e(\theta_e)} = -P(\theta_e)\,Q_{ij}(\theta_e)\left[1-\tau_e(\theta_e)\right](1-\tau_s)\,\varepsilon_{Q-ij}^{P,cross}(\theta_e)\left[\begin{array}{c}\left[1-g_e(\theta_e)\right]-\frac{[1-\bar{g}_e(\theta_e)][1-F_e(\theta_e)]}{f_e(\theta_e)}\\ \times\left[\frac{1+\varepsilon_e}{\varepsilon_e}\frac{l_e'(\theta_e)}{l_e(\theta_e)}+\frac{d\ln\left[\mu(\theta_e)\varepsilon_{Q-ij}^{P,cross}(\theta_e)\right]}{d\theta_e}\right]\end{array}\right]. \tag{A46}$$

B.3 Proof of Theorem 1

Unless otherwise specified, the following equations in this subsection are derived for any $\theta_o \in \Theta_o$ and $\tau_s = 0$.

(i) $\frac{\partial \pounds}{\partial L_w(\theta_e)} = 0$ implies:

$$\begin{aligned}P(\theta_e)\frac{\partial Q_{ij}(\theta_e)}{\partial L_w(\theta_e)} &= \frac{\lambda'}{\lambda}-\frac{\kappa'(\theta_e)}{\lambda L_w(\theta_e)N_e f_e(\theta_e)}\frac{\partial\ln Q_{ij}(\theta_e)}{\partial\ln L_w(\theta_e)}+\frac{\int_{\Theta_e}\varphi'(\theta_e')\frac{\partial\ln\varpi(\theta_e')}{\partial\ln L_w(\theta_e)}d\theta_e'}{\lambda L_w(\theta_e)N_e f_e(\theta_e)}\\ &= \frac{\lambda'}{\lambda}-\frac{\kappa'(\theta_e)\,\xi}{\lambda L_w(\theta_e)N_e f_e(\theta_e)}+\frac{\varphi'(\theta_e)\,\varepsilon_{L_w}^{\varpi}(\theta_e)}{\lambda L_w(\theta_e)N_e f_e(\theta_e)},\end{aligned}$$

where $\int_{\Theta_e}\varphi'(\theta_e')\frac{\partial\ln\varpi(\theta_e')}{\partial\ln L_w(\theta_e)}d\theta_e' = \varphi'(\theta_e)\,\varepsilon_{L_w}^{\varpi}(\theta_e)$ since $\varepsilon_{L_w}^{\varpi}(\theta_e',\theta_e)$ is independent of θ_e' and $\int_{\Theta_e}\varphi'(\theta_e')\,d\theta' = 0$. Substituting $P(\theta_e)\frac{\partial Q_{ij}(\theta_e)}{\partial L_w(\theta_e)}$ by $W\mu(\theta_e)$ gives:

$$W\mu(\theta_e) = \frac{\lambda'}{\lambda}-\frac{\kappa'(\theta_e)\,\xi}{\lambda L_w(\theta_e)N_e f_e(\theta_e)}+\frac{\varphi'(\theta_e)\,\varepsilon_{L_w}^{\varpi}(\theta_e)}{\lambda L_w(\theta_e)N_e f_e(\theta_e)}. \tag{A47}$$

Dividing both sides of the above equation by $\frac{\varepsilon_{L_w}^{\varpi}(\theta_e)}{L_w(\theta_e)N_e f_e(\theta_e)}$ and integrating across θ_e gives:

$$W\int_{\Theta_e}\mu(\theta_e)\frac{L_w(\theta_e)N_e f_e(\theta_e)}{\varepsilon_{L_w}^{\varpi}(\theta_e)}d\theta_e = \frac{\lambda'}{\lambda}\int_{\Theta_e}\frac{L_w(\theta_e)N_e f_e(\theta_e)}{\varepsilon_{L_w}^{\varpi}(\theta_e)}d\theta_e-\int_{\Theta_e}\frac{\kappa'(\theta_e)}{\lambda\varepsilon_{L_w}^{\varpi}(\theta_e)}\xi d\theta_e,$$

where we use $\int_{\Theta_e}\varphi'(\theta_e')\,d\theta' = 0$ again. Reformation of the above equation gives:

$$\begin{aligned}1 &= \frac{\frac{\lambda'}{\lambda}\int_{\Theta_e}\frac{L_w(\theta_e)N_e f_e(\theta_e)}{\varepsilon_{L_w}^{\varpi}(\theta_e)}d\theta_e}{W\int_{\Theta_e}\mu(\theta_e)\frac{L_w(\theta_e)N_e f_e(\theta_e)}{\varepsilon_{L_w}^{\varpi}(\theta_e)}d\theta_e}-\frac{\int_{\Theta_e}\frac{\kappa'(\theta_e)}{\lambda\varepsilon_{L_w}^{\varpi}(\theta_e)}\xi d\theta_e}{W\int_{\Theta_e}\mu(\theta_e)\frac{L(\theta_e)N_e f_e(\theta_e)}{\varepsilon_{L_w}^{\varpi}(\theta_e)}d\theta_e}\\ &= \frac{\frac{\lambda'}{\lambda}\int_{\Theta_e}\frac{L_w(\theta_e)N_e f_e(\theta_e)}{\varepsilon_{L_w}^{\varpi}(\theta_e)}d\theta_e}{W\int_{\Theta_e}\mu(\theta_e)\frac{L_w(\theta_e)N_e f_e(\theta_e)}{\varepsilon_{L_w}^{\varpi}(\theta_e)}d\theta_e}+\int_{\Theta_e}\frac{\kappa(\theta_e)}{\lambda}\frac{d\frac{\xi}{\varepsilon_{L_w}^{\varpi}(\theta_e)}/d\theta_e}{W\int_{\Theta_e}\mu(\theta_e)\frac{L_w(\theta_e)N_e f_e(\theta_e)}{\varepsilon_{L_w}^{\varpi}(\theta_e)}d\theta_e}d\theta_e,\end{aligned} \tag{A48}$$

where the second equation is derived by $\kappa(\underline{\theta}_e) = \kappa(\overline{\theta}_e) = 0$ and integration by parts. Note that under our production function $\varepsilon^{\varpi}_{L_w}(\theta_e)$ is independent of θ_e (see e.g., (A10)). Thus, $d\frac{\xi}{\varepsilon^{\varpi}_{L_w}(\theta_e)}/d\theta_e = 0$ and (A48) implies:

$$1 = \frac{\lambda'}{\lambda W \mu}. \tag{A49}$$

According to (A35), we have:

$$\frac{1}{\frac{\phi'_w(l_w(\theta_w))}{x_w(\theta_w)}} = \frac{\lambda}{\lambda'}\left[1 - \frac{x'_w(\theta_w)}{x_w(\theta_w)}\frac{\psi_w(\theta_w)}{\lambda N_w f_w(\theta_w)}\frac{1+\varepsilon_w}{\varepsilon_w}\right].$$

Substitute $\frac{\phi'_w(l_w(\theta_w))}{x_w(\theta_w)}$ by $[1-\tau_w(\theta_w)]\,W$:

$$\frac{1}{1-\tau_w(\theta_w)} = \frac{W\lambda}{\lambda'}\left[1 - \frac{x'_w(\theta_w)}{x_w(\theta_w)}\frac{\psi_w(\theta_w)}{\lambda N_w f_w(\theta_w)}\frac{1+\varepsilon_w}{\varepsilon_w}\right]. \tag{A50}$$

Use (A42), and (A49) to substitute $\frac{\psi_w(\theta_w)}{\lambda N_w f_w(\theta_w)}$ and $\frac{\lambda}{\lambda'}$ in (A50):

$$\frac{1}{1-\tau_w(\theta_w)} = \frac{1}{\mu}\left[1 + [1-\bar{g}_w(\theta_w)]\frac{1-F_w(\theta_w)}{f_w(\theta_w)}\frac{x'_w(\theta_w)}{x_w(\theta_w)}\frac{1+\varepsilon_w}{\varepsilon_w}\right]. \tag{A51}$$

(ii) In the following analysis, we first derive an optimal profit tax formula in part (a). Then we simplify the expression in parts (b) and (c).

(a) Divide both sides of (A37) by $\lambda N_e f_e(\theta_e)\,P(\theta_e)\,\frac{\partial Q_{ij}(\theta_e)}{\partial L_e(\theta_e)}$:

$$\begin{aligned}
& 1 - \frac{\phi'_e(l_e(\theta_e))}{P(\theta_e)\frac{\partial Q_{ij}(\theta_e)}{\partial l_e(\theta_e)}} \\
= \quad & -\frac{\psi_e(\theta_e)}{\lambda N_e f_e(\theta_e)}\frac{\phi'_e(l_e(\theta_e))}{P(\theta_e)\frac{\partial Q_{ij}(\theta_e)}{\partial l_e(\theta_e)}}\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu(\theta_e)\frac{d\ln P_{ij}\left(Q_{ij}, Q_{-ij}(\theta_e), \theta_e\right)}{d\theta_e}\Big|_{Q_{ij}=Q_{ij}(\theta_e)} + \frac{x'_e(\theta_e)}{x_e(\theta_e)}\right] \\
& -\frac{\kappa'(\theta_e)}{\lambda l_e(\theta_e)\,P(\theta_e)\frac{\partial Q_{ij}(\theta_e)}{\partial l_e(\theta_e)}N_e f_e(\theta_e)} + \frac{\varphi'(\theta_e)\,\varepsilon^{\varpi}_{l_e}(\theta_e)}{\lambda l_e(\theta_e)\,P(\theta_e)\frac{\partial Q_{ij}(\theta_e)}{\partial l_e(\theta_e)}N_e f_e(\theta_e)},
\end{aligned}$$

where we use $\frac{\partial \ln Q_{ij}(\theta_e)}{\partial \ln l_e(\theta_e)} = 1$ and $\int_\Theta \varphi'(\theta'_e)\frac{\partial \ln \varpi(\theta'_e)}{\partial \ln l_e(\theta_e)}d\theta_e = \varphi'(\theta_e)\,\varepsilon^{\varpi}_{l_e}(\theta_e)$ to simplify the expression. Moreover, from the definitions of the elasticities,

$$\int_\Theta \varphi'(\theta'_e)\frac{\partial \ln \varpi(\theta'_e)}{\partial \ln l_e(\theta_e)}d\theta_e = \varphi'(\theta_e)\,\varepsilon^{\varpi}_{l_e}(\theta_e)$$

since $\varepsilon^{\varpi}_{l_e}(\theta'_e, \theta_e)$ is independent of θ'_e and $\int_{\theta_e}\varphi'(\theta'_e)\,d\theta' = 0$.

For the convenience of derivation, we define:

$$1 - \widetilde{\tau}_e(\theta_e) \equiv \frac{[1-\tau_e(\theta_e)]\,(1-\tau_s)}{\mu(\theta_e)} = \frac{\phi'_e(l_e(\theta_e))}{P(\theta_e)\frac{\partial Q_{ij}(\theta_e)}{\partial l_e(\theta_e)}}.$$

Then one has

$$\widetilde{\tau}_e(\theta_e) = -\frac{\psi_e(\theta_e)}{\lambda N_e f_e(\theta_e)} \frac{1+\varepsilon_e}{\varepsilon_e} \left[\mu(\theta_e) \frac{d \ln P_{ij}\left(Q_{ij}, Q_{-ij}(\theta_e), \theta_e\right)}{d\theta_e} |_{Q_{ij}=Q_{ij}(\theta_e)} + \frac{x_e'(\theta_e)}{x_e(\theta_e)} \right] [1 - \widetilde{\tau}_e(\theta_e)]$$
$$- \frac{\kappa'(\theta_e)}{\lambda P(\theta_e) Q_{ij}(\theta_e) N_e f_e(\theta_e)} + \frac{\varphi'(\theta_e)\, \varepsilon^{\varpi}_{l_e}(\theta_e)}{\lambda P(\theta_e) Q_{ij}(\theta_e) N_e f_e(\theta_e)},$$

where we use $\frac{\partial \ln Q_{ij}(\theta_e)}{\partial \ln l_e(\theta_e)} = 1$ to simplify the expression. In the same vein, we have

$$\frac{\widetilde{\tau}_e(\theta_e)}{1-\widetilde{\tau}_e(\theta_e)} = -\frac{\psi_e(\theta_e)}{\lambda N_e f_e(\theta_e)} \frac{1+\varepsilon_e}{\varepsilon_e} \left[\mu(\theta_e) \frac{d \ln P_{ij}\left(Q_{ij}, Q_{-ij}(\theta_e), \theta_e\right)}{d\theta_e} |_{Q_{ij}=Q_{ij}(\theta_e)} + \frac{x_e'(\theta_e)}{x_e(\theta_e)} \right]$$
$$- \frac{1}{1-\widetilde{\tau}_e(\theta_e)} \frac{1}{P(\theta_e) Q_{ij}(\theta_e)} \left[\frac{\kappa'(\theta_e)}{\lambda N_e f_e(\theta_e)} - \frac{\varphi'(\theta_e)}{\lambda N_e f_e(\theta_e)} \varepsilon^{\varpi}_{l_e}(\theta_e) \right] \tag{A52}$$

or

$$\frac{1}{1-\widetilde{\tau}_e(\theta_e)} = 1 - \frac{\psi_e(\theta_e)}{\lambda N_e f_e(\theta_e)} \frac{1+\varepsilon_e}{\varepsilon_e} \left[\mu(\theta_e) \frac{d \ln P_{ij}\left(Q_{ij}, Q_{-ij}(\theta_e), \theta_e\right)}{d\theta_e} |_{Q_{ij}=Q_{ij}(\theta_e)} + \frac{x_e'(\theta_e)}{x_e(\theta_e)} \right]$$
$$- \frac{1}{1-\widetilde{\tau}_e(\theta_e)} \frac{1}{P(\theta_e) Q_{ij}(\theta_e)} \left[\frac{\kappa'(\theta_e)}{\lambda N_e f_e(\theta_e)} - \frac{\varphi'(\theta_e)}{\lambda N_e f_e(\theta_e)} \varepsilon^{\varpi}_{l_e}(\theta_e) \right]. \tag{A53}$$

Combining (A52) and (A42) gives:

$$\frac{\widetilde{\tau}_e(\theta_e)}{1-\widetilde{\tau}_e(\theta_e)} = [1-\bar{g}_e(\theta_e)] \frac{1-F_e(\theta_e)}{f_e(\theta_e)} \frac{1+\varepsilon_e}{\varepsilon_e} \left[\mu(\theta_e) \frac{d \ln P_{ij}\left(Q_{ij}, Q_{-ij}(\theta_e), \theta_e\right)}{d\theta_e} |_{Q_{ij}=Q_{ij}(\theta_e)} + \frac{x_e'(\theta_e)}{x_e(\theta_e)} \right]$$
$$- \frac{1}{1-\widetilde{\tau}_e(\theta_e)} \frac{1}{P(\theta_e) Q_{ij}(\theta_e)} \frac{\kappa'(\theta_e)}{\lambda N_e f_e(\theta_e)}$$
$$+ \frac{1}{1-\widetilde{\tau}_e(\theta_e)} \frac{1}{P(\theta_e) Q_{ij}(\theta_e)} \frac{\varphi'(\theta_e)\, \varepsilon^{\varpi}_{l_e}(\theta_e)}{\lambda N_e f_e(\theta_e)}.$$

Using (A47) to substitute $\frac{\varphi'(\theta_e)\varepsilon^{\varpi}_{l_e}(\theta_e)}{\lambda N_e f_e(\theta_e)}$ in the above equation,[46] we have:

$$\frac{\widetilde{\tau}_e(\theta_e)}{1-\widetilde{\tau}_e(\theta_e)} = [1-\bar{g}_e(\theta_e)] \frac{1-F_e(\theta_e)}{f_e(\theta_e)} \frac{1+\varepsilon_e}{\varepsilon_e} \left[\mu(\theta_e) \frac{d \ln P_{ij}\left(Q_{ij}, Q_{-ij}(\theta_e), \theta_e\right)}{d\theta_e} |_{Q_{ij}=Q_{ij}(\theta_e)} + \frac{x_e'(\theta_e)}{x_e(\theta_e)} \right]$$
$$- \frac{1}{1-\widetilde{\tau}_e(\theta_e)} \frac{1}{P(\theta_e) Q_{ij}(\theta_e)} \frac{\kappa'(\theta_e)}{\lambda N_e f_e(\theta_e)} \left[1 - \xi \frac{\varepsilon^{\varpi}_{l_e}(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)} \right]$$
$$- \frac{1}{1-\widetilde{\tau}_e(\theta_e)} \frac{L_w(\theta_e)}{P(\theta_e) Q_{ij}(\theta_e)} \frac{\lambda'}{\lambda} \left[1 - \frac{\lambda}{\lambda'} \frac{W\mu(\theta_e)}{1-\tau_s} \right] \frac{\varepsilon^{\varpi}_{l_e}(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)}. \tag{A54}$$

We now transform the three terms on the right side of the above equations one by one. First, substituting

[46] Equation (A47) suggests that $\frac{\varphi'(\theta_e)\varepsilon^{\varpi}_{l_e}(\theta_e)}{\lambda N_e f_e(\theta_e)} = \left[\left[\frac{W\mu(\theta)}{1-\tau_s} - \frac{\lambda'}{\lambda} \right] L_w(\theta_e) + \frac{\kappa'(\theta_e)\xi}{\lambda N_e f_e(\theta_e)} \right] \frac{\varepsilon^{\varpi}_{l_e}(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)}$.

$\kappa'(\theta_e)$ with (A46), we have the following equation:[47]

$$-\frac{1}{1-\widetilde{\tau}_e(\theta_e)}\frac{1}{P(\theta_e)Q_{ij}(\theta_e)}\frac{\kappa'(\theta_e)}{\lambda N_e f_e(\theta_e)}=\frac{1-\tau_e(\theta_e)}{1-\widetilde{\tau}_e(\theta_e)}\varepsilon^{P,cross}_{Q_{-ij}}(\theta_e)\left[\begin{array}{c}[1-g_e(\theta_e)]-\frac{[1-\bar{g}_e(\theta_e)][1-F_e(\theta_e)]}{f_e(\theta_e)}\\ \times\left[\begin{array}{c}\frac{1+\varepsilon_e}{\varepsilon_e}\frac{l'_e(\theta_e)}{l_e(\theta_e)}+\\ \frac{d\ln\left[\mu(\theta_e)\varepsilon^{P,cross}_{Q_{-ij}}(\theta_e)\right]}{d\theta_e}\end{array}\right]\end{array}\right]. \quad \text{(A55)}$$

Second, notice that $\frac{L_w(\theta_e)W}{P(\theta_e)Q_{ij}(\theta_e)}=\frac{\xi}{\mu(\theta_e)}$ and $\frac{\lambda'}{\lambda W}=\mu$ (see e.g., (A49)). The last term of (A54) equals:

$$\begin{aligned}&-\frac{1}{1-\widetilde{\tau}_e(\theta_e)}\frac{L_w(\theta_e)}{P(\theta_e)Q_{ij}(\theta_e)}\frac{\lambda'}{\lambda}\left[1-\frac{\lambda}{\lambda'}\frac{W\mu(\theta_e)}{1-\tau_s}\right]\frac{\varepsilon^{\varpi}_{l_e}(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)} \qquad \text{(A56)}\\ =\;&-\frac{1-\tau_s}{1-\widetilde{\tau}_e(\theta_e)}\frac{\xi}{\mu(\theta_e)}\mu\left[1-\frac{\mu(\theta_e)}{\mu}\right]\frac{\varepsilon^{\varpi}_{l_e}(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)}\\ =\;&-\frac{\xi}{1-\widetilde{\tau}_e(\theta_e)}\left[\frac{\mu}{\mu(\theta_e)}-1\right]\frac{\varepsilon^{\varpi}_{l_e}(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)}.\end{aligned}$$

Substituting the second and third terms of the right side of (A54) by (A55) and (A56) gives:

$$\begin{aligned}&\frac{1}{1-\tau_e(\theta_e)} \qquad \text{(A57)}\\ =\;&\frac{1+[1-\bar{g}_e(\theta_e)]\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu(\theta_e)\frac{d\ln P_{ij}\left(Q_{ij},Q_{-ij}(\theta_e),\theta_e\right)}{d\theta_e}|_{Q_{ij}=Q_{ij}(\theta_e)}+\frac{x'_e(\theta_e)}{x_e(\theta_e)}\right]}{\mu(\theta_e)}\\ &+\varepsilon^{P,cross}_{Q_{-ij}}(\theta_e)\left[\begin{array}{c}[1-g_e(\theta_e)]-\\ \frac{[1-\bar{g}_e(\theta_e)][1-F_e(\theta_e)]}{f_e(\theta_e)}\left[\begin{array}{c}\frac{1+\varepsilon_e}{\varepsilon_e}\frac{l'_e(\theta_e)}{l_e(\theta_e)}\\ +\frac{d\ln\left[\mu(\theta_e)\varepsilon^{P,cross}_{Q_{-ij}}(\theta_e)\right]}{d\theta_e}\end{array}\right]\end{array}\right]\left[1-\xi\frac{\varepsilon^{\varpi}_{l_e}(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)}\right]\\ &+\frac{1}{1-\tau_e(\theta_e)}\left[1-\frac{\mu}{\mu(\theta_e)}\right]\xi\frac{\varepsilon^{\varpi}_{l_e}(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)},\end{aligned}$$

where we have substituted $1-\widetilde{\tau}_e(\theta_e)$ by $\frac{1-\tau_e(\theta_e)}{\mu(\theta_e)}$.

Using $RE(\theta_e)\equiv\frac{\mu}{\mu(\theta_e)}-1$ and

$$\widetilde{IRE}(\theta_e)\equiv\varepsilon^{P,cross}_{Q_{-ij}}(\theta_e)\left[[1-g_e(\theta_e)]-\frac{[1-\bar{g}_e(\theta_e)][1-F_e(\theta_e)]}{f_e(\theta_e)}\left[\begin{array}{c}\frac{1+\varepsilon_e}{\varepsilon_e}\frac{l'_e(\theta_e)}{l_e(\theta_e)}+\\ \frac{d\ln\left[\mu(\theta_e)\varepsilon^{P,cross}_{Q_{-ij}}(\theta_e)\right]}{d\theta_e}\end{array}\right]\right], \qquad \text{(A58)}$$

[47] Note that we consider the case with $\tau_s=0$.

we have

$$
\begin{aligned}
\frac{1}{1-\tau_e\left(\theta_e\right)} \quad = \quad & \frac{1+\left[1-\bar{g}_e(\theta_e)\right]\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu\left(\theta_e\right)\frac{d\ln P_{ij}\left(Q_{ij},Q_{-ij}(\theta_e),\theta_e\right)}{d\theta_e}|_{Q_{ij}=Q_{ij}(\theta_e)}+\frac{x_e'(\theta_e)}{x_e(\theta_e)}\right]}{\mu\left(\theta_e\right)} \\
& +\widetilde{IRE}\left(\theta_e\right)\left[1-\xi\frac{\varepsilon_{l_e}^{\varpi}(\theta_e)}{\varepsilon_{L_w}^{\varpi}(\theta_e)}\right] \\
& -\frac{1}{1-\tau_e\left(\theta_e\right)}\xi RE\left(\theta_e\right)\frac{\varepsilon_{l_e}^{\varpi}(\theta_e)}{\varepsilon_{L_w}^{\varpi}(\theta_e)},
\end{aligned}
$$

which is equivalent to

$$
\frac{1}{1-\tau_e\left(\theta_e\right)}=\frac{\frac{1+\left[1-\bar{g}_e(\theta_e)\right]\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu(\theta_e)\frac{d\ln P_{ij}\left(Q_{ij},Q_{-ij}(\theta_e),\theta_e\right)}{d\theta_e}|_{Q_{ij}=Q_{ij}(\theta_e)}+\frac{x_e'(\theta_e)}{x_e(\theta_e)}\right]}{\mu(\theta_e)}+\widetilde{IRE}\left(\theta_e\right)\left[1-\xi\frac{\varepsilon_{l_e}^{\varpi}(\theta_e)}{\varepsilon_{L_w}^{\varpi}(\theta_e)}\right]}{1+RE\left(\theta_e\right)\xi\frac{\varepsilon_{l_e}^{\varpi}(\theta_e)}{\varepsilon_{L_w}^{\varpi}(\theta_e)}}. \tag{A59}
$$

(b) We now try to express the right side of the above equation in terms of parameters. Using (OA29), we have:

$$
\begin{aligned}
& \frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu\left(\theta_e\right)\frac{d\ln P_{ij}\left(Q_{ij},Q_{ij}\left(\theta_e\right),\theta_e\right)}{d\theta_e}|_{Q_{ij}=Q_{ij}(\theta_e)}+\frac{x_e'\left(\theta_e\right)}{x_e\left(\theta_e\right)}\right] \\
= \quad & \frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{1+\varepsilon_e}{\varepsilon_e}\left\{\begin{array}{c}\frac{\mu(\theta_e)\varepsilon_{Q_{-ij}}^{P,cross}(\theta_e)+1}{\varepsilon_{Q_{ij}}^{S}(\theta_e)}\frac{d}{d\theta_e}\left[\ln\frac{X_e(\theta_e)}{\mu(\theta_e)}\right]+\mu\left(\theta_e\right)\frac{d\ln\mu(\theta_e)}{d\theta_e} \\ +\mu\left(\theta_e\right)\varepsilon_{Q_{-ij}}^{P,cross}\left(\theta_e\right)\frac{\left(1+\xi\frac{1+\varepsilon_e}{\varepsilon_e}\right)\frac{d}{d\theta_e}\left[\ln\frac{X_e(\theta_e)}{\mu(\theta_e)}\right]+\frac{d\ln[1-\tau_e(\theta_e)]}{d\theta_e}}{\frac{1+\varepsilon_e}{\varepsilon_e}-\varepsilon_{Q_{ij}}^{S}(\theta_e)\left(1+\frac{1+\varepsilon_e}{\varepsilon_e}\xi\right)}\end{array}\right\} \\
= \quad & \frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{d\ln\left[\mu\left(\theta_e\right)-\xi\right]}{d\theta_e}\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu\left(\theta_e\right)-\xi\right] \\
& +\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{\frac{1+\varepsilon_e}{\varepsilon_e}\frac{d}{d\theta_e}\ln\frac{X_e(\theta_e)}{\mu(\theta_e)}}{\frac{1+\varepsilon_e}{\varepsilon_e}-\varepsilon_{Q_{ij}}^{S}\left(\theta_e\right)\left(1+\frac{1+\varepsilon_e}{\varepsilon_e}\xi\right)}\left[\begin{array}{c}\frac{\mu(\theta_e)\varepsilon_{Q_{-ij}}^{P,cross}(\theta_e)+1}{\varepsilon_{Q_{ij}}^{S}(\theta_e)}\frac{1+\varepsilon_e}{\varepsilon_e} \\ -\left(1+\xi\frac{1+\varepsilon_e}{\varepsilon_e}\right)\end{array}\right] \\
& +\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{1+\varepsilon_e}{\varepsilon_e}\frac{\mu\left(\theta_e\right)\varepsilon_{Q_{-ij}}^{P,cross}\left(\theta_e\right)\frac{d\ln[1-\tau_e(\theta_e)]}{d\theta_e}}{\frac{1+\varepsilon_e}{\varepsilon_e}-\varepsilon_{Q_{ij}}^{S}\left(\theta_e\right)\left(1+\frac{1+\varepsilon_e}{\varepsilon_e}\xi\right)},
\end{aligned}
$$

where the third equation is derived by $\mu\left(\theta_e\right)\frac{d\ln\mu(\theta_e)}{d\theta_e}=\left[\mu\left(\theta_e\right)-\xi\right]\frac{d\ln[\mu(\theta_e)-\xi]}{d\theta_e}$ and combine terms multiplied by $\frac{d}{d\theta_e}\ln\frac{X_e(\theta_e)}{\mu(\theta_e)}$.

Notice that:

$$
\frac{1-F_e(\theta_e)}{f_e\left(\theta_e\right)}\frac{\frac{1+\varepsilon_e}{\varepsilon_e}\frac{d}{d\theta_e}\ln\frac{X_e(\theta_e)}{\mu(\theta_e)}}{\frac{1+\varepsilon_e}{\varepsilon_e}-\frac{\sigma-1}{\sigma}\left(1+\frac{1+\varepsilon_e}{\varepsilon_e}\xi\right)}=H(\theta_e)-\frac{1-F_e(\theta_e)}{f_e\left(\theta_e\right)}\frac{d\ln\left[\mu\left(\theta_e\right)-\xi\right]}{d\theta_e},
$$

where $H(\theta_e)$ is given by (OA25). We have

$$
\begin{aligned}
&\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu\left(\theta_e\right)\frac{d\ln P_{ij}\left(Q_{ij},Q_{ij}\left(\theta_e\right),\theta_e\right)}{d\theta_e}|_{Q_{ij}=Q_{ij}(\theta_e)}+\frac{x_e'\left(\theta_e\right)}{x_e\left(\theta_e\right)}\right]\\
=\;&\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{d\ln\left[\mu\left(\theta_e\right)-\xi\right]}{d\theta_e}\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu\left(\theta_e\right)+\frac{\varepsilon_e}{1+\varepsilon_e}-\frac{\mu\left(\theta_e\right)\varepsilon_{Q_{-ij}}^{P,cross}+1}{\varepsilon_{Q_{ij}}^{S}\left(\theta_e\right)}\right]\\
&+H(\theta_e)\frac{1+\varepsilon_e}{\varepsilon_e}\left[\frac{\mu\left(\theta_e\right)\varepsilon_{Q_{-ij}}^{P,cross}+1}{\varepsilon_{Q_{ij}}^{S}\left(\theta_e\right)}-\frac{\varepsilon_e}{1+\varepsilon_e}-\xi\right]\\
&+\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{1+\varepsilon_e}{\varepsilon_e}\frac{\mu\left(\theta_e\right)\varepsilon_{Q_{-ij}}^{P,cross}\frac{d\ln[1-\tau_e(\theta_e)]}{d\theta_e}}{\frac{1+\varepsilon_e}{\varepsilon_e}-\varepsilon_{Q_{ij}}^{S}\left(\theta_e\right)\left(1+\frac{1+\varepsilon_e}{\varepsilon_e}\xi\right)},
\end{aligned}
$$

where, according to $\varepsilon_{Q_{-ij}}^{P,cross}=-\frac{1}{\mu(\theta_e)}+\varepsilon_{Q_{ij}}^{S}\left(\theta_e\right)$, one has $\frac{\mu(\theta_e)\varepsilon_{Q_{-ij}}^{P,cross}+1}{\varepsilon_{Q_{ij}}^{S}(\theta_e)}=\mu\left(\theta_e\right)$. Therefore,

$$
\begin{aligned}
&\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu\left(\theta_e\right)\frac{d\ln P_{ij}\left(Q_{ij},Q_{ij}\left(\theta_e\right),\theta_e\right)}{d\theta_e}|_{Q_{ij}=Q_{ij}(\theta_e)}+\frac{x_e'\left(\theta_e\right)}{x_e\left(\theta_e\right)}\right]\\
=\;&\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{d\ln\left[\mu\left(\theta_e\right)-\xi\right]}{d\theta_e}+H(\theta_e)\left[\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu\left(\theta_e\right)-\xi\right]-1\right]\\
&+\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{1+\varepsilon_e}{\varepsilon_e}\frac{\mu\left(\theta_e\right)\varepsilon_{Q_{-ij}}^{P,cross}\frac{d\ln[1-\tau_e(\theta_e)]}{d\theta_e}}{\frac{1+\varepsilon_e}{\varepsilon_e}-\varepsilon_{Q_{ij}}^{S}\left(\theta_e\right)\left(1+\frac{1+\varepsilon_e}{\varepsilon_e}\xi\right)}.
\end{aligned}
\tag{A60}
$$

Substituting $\frac{l_e'(\theta_e)}{l_e(\theta_e)}$ in $\widetilde{IRE}\left(\theta_e\right)$ by (OA22) and utilizing (OA25), i.e.,

$$
H(\theta_e)=\frac{1-F_e(\theta_e)}{f_e\left(\theta_e\right)}\left[\frac{\frac{1+\varepsilon_e}{\varepsilon_e}\frac{d\ln[X_e(\theta_e)/\mu(\theta_e)]}{d\theta_e}}{\frac{1+\varepsilon_e}{\varepsilon_e}\left[1-\xi\varepsilon_{Q_{ij}}^{S}\left(\theta_e\right)\right]-\varepsilon_{Q_{ij}}^{S}\left(\theta_e\right)}+\frac{d\ln\left[\mu\left(\theta_e\right)-\xi\right]}{d\theta_e}\right],
$$

we have:

$$
\widetilde{IRE}\left(\theta_e\right)=\varepsilon_{Q_{-ij}}^{P,cross}\left(\theta_e\right)\left[\left[1-g_e(\theta_e)\right]-\left[1-\bar{g}_e(\theta_e)\right]\left[\begin{array}{c}\frac{\frac{1+\varepsilon_e}{\varepsilon_e}\left[1-\xi\varepsilon_{Q_{ij}}^{S}(\theta_e)\right]\frac{d\ln[1-\tau_e(\theta_e)]}{d\theta_e}}{\frac{1+\varepsilon_e}{\varepsilon_e}-\varepsilon_{Q_{ij}}^{S}(\theta_e)\left(1+\frac{1+\varepsilon_e}{\varepsilon_e}\xi\right)}\frac{1-F_e(\theta_e)}{f_e(\theta_e)}+H(\theta_e)\\-\left[1-\frac{\frac{d\ln\left[\mu(\theta_e)\varepsilon_{Q_{-ij}}^{P,cross}(\theta_e)\right]}{d\theta_e}}{\frac{d\ln[\mu(\theta_e)-\xi]}{d\theta_e}}\right]\frac{[1-F_e(\theta_e)]}{f_e(\theta_e)}\frac{d\ln[\mu(\theta_e)-\xi]}{d\theta_e}\end{array}\right]\right].
$$

Notice that $\mu(\theta_e)\varepsilon^{P,cross}_{Q_{-ij}}(\theta_e)+1=\mu(\theta_e)\varepsilon^{S}_{Q_{ij}}(\theta_e)$. We have $\frac{\frac{d\ln\left[\mu(\theta_e)\varepsilon^{P,cross}_{Q_{-ij}}(\theta_e)\right]}{d\theta_e}}{\frac{d\ln[\mu(\theta_e)-\xi]}{d\theta_e}}=\frac{\varepsilon^{S}_{Q_{ij}}(\theta_e)[\mu(\theta_e)-\xi]}{\mu(\theta_e)\varepsilon^{P,cross}_{Q_{-ij}}(\theta_e)}$ and:

$$\widetilde{IRE}(\theta_e)=\varepsilon^{P,cross}_{Q_{-ij}}(\theta_e)\left\{[1-g_e(\theta_e)]-[1-\bar{g}_e(\theta_e)]\left[\begin{array}{c} H(\theta_e)+\frac{\left[1-\xi\varepsilon^{S}_{Q_{ij}}(\theta_e)\right]\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{d\ln[\mu(\theta_e)-\xi]}{d\theta_e}}{\mu(\theta_e)\varepsilon_{Q_{-ij}}(\theta_e)} \\ +\frac{\left[1-\xi\varepsilon^{S}_{Q_{ij}}(\theta_e)\right]\frac{1+\varepsilon_e}{\varepsilon_e}\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{d\ln[1-\tau_e(\theta_e)]}{d\theta_e}}{\frac{1+\varepsilon_e}{\varepsilon_e}-\varepsilon^{S}_{Q_{ij}}(\theta_e)\left(1+\frac{1+\varepsilon_e}{\varepsilon_e}\xi\right)}\end{array}\right]\right\}. \tag{A61}$$

Last, substituting $\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu(\theta_e)\frac{\partial\ln P\left(Q_{ij}(\theta_e),\theta_e\right)}{\partial\theta_e}+\frac{x_e'(\theta_e)}{x_e(\theta_e)}\right]$ and $\widetilde{IRE}(\theta_e)$ in (A59) by (A60) and (A61), respectively, we have, for any $\theta_e\in\Theta_e$:

$$\begin{aligned}&\frac{1+RE(\theta_e)\,\xi\frac{\varepsilon^{\varpi}_{l_e}(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)}}{1-\tau_e(\theta_e)} \\ =\;&\frac{1+[1-\bar{g}_e(\theta_e)]\left[\begin{array}{c}H(\theta_e)\left[\frac{1+\varepsilon_e}{\varepsilon_e}[\mu(\theta_e)-\xi]-1\right]+\\ \frac{1-F_e(\theta_e)}{f_e(\theta_e)}\left[\frac{d\ln[\mu(\theta_e)-\xi]}{d\theta_e}+\frac{\frac{1+\varepsilon_e}{\varepsilon_e}\mu(\theta_e)\varepsilon^{P,cross}_{Q_{-ij}}(\theta_e)\frac{d\ln[1-\tau_e(\theta_e)]}{d\theta_e}}{\frac{1+\varepsilon_e}{\varepsilon_e}-\varepsilon^{S}_{Q_{ij}}(\theta_e)\left(1+\frac{1+\varepsilon_e}{\varepsilon_e}\xi\right)}\right]\end{array}\right]}{\mu(\theta_e)} \\ &+\left[1-\xi\frac{\varepsilon^{\varpi}_{l_e}(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)}\right][1-g_e(\theta_e)]\,\varepsilon^{P,cross}_{Q_{-ij}}(\theta_e) \\ &-\left[1-\xi\frac{\varepsilon^{\varpi}_{l_e}(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)}\right]\varepsilon^{P,cross}_{Q_{-ij}}(\theta_e)\,[1-\bar{g}_e(\theta_e)]\left\{\begin{array}{c}H(\theta_e)+\frac{\left[1-\xi\varepsilon^{S}_{Q_{ij}}(\theta_e)\right]\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{d\ln[\mu(\theta_e)-\xi]}{d\theta_e}}{\mu(\theta_e)\varepsilon^{P,cross}_{Q_{-ij}}(\theta_e)} \\ +\frac{\left[1-\xi\varepsilon^{S}_{Q_{ij}}(\theta_e)\right]\frac{1+\varepsilon_e}{\varepsilon_e}\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{d\ln[1-\tau_e(\theta_e)]}{d\theta_e}}{\left[\frac{1+\varepsilon_e}{\varepsilon_e}-\varepsilon^{S}_{Q_{ij}}(\theta_e)\left(1+\frac{1+\varepsilon_e}{\varepsilon_e}\xi\right)\right]}\end{array}\right\}.\end{aligned} \tag{A62}$$

Notice that $1-\xi\frac{\varepsilon^{\varpi}_{l_e}(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)}=\frac{1}{1-\xi\varepsilon^{S}_{Q_{ij}}(\theta_e)}$. One can see the sum of terms multiplied by $\frac{d\ln[1-\tau_e(\theta_e)]}{d\theta_e}$ of the above equation equals zero. Moreover, the sum of terms multiplied by $\frac{d\ln[\mu(\theta_e)-\xi]}{d\theta_e}$ also equals zero. Last, using the definition of $IRE(\theta_e)$ (see e.g., (38)), we have (A63):

$$\begin{aligned}\frac{1+RE(\theta_e)\,\xi\frac{\varepsilon^{\varpi}_{l_e}(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)}}{1-\tau_e(\theta_e)}\;=\;&\frac{1+[1-\bar{g}_e(\theta_e)]\left\{H(\theta_e)\left[\frac{1+\varepsilon_e}{\varepsilon_e}[\mu(\theta_e)-\xi]-1\right]\right\}}{\mu(\theta_e)} \\ &+\left[1-\xi\frac{\varepsilon^{\varpi}_{l_e}(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)}\right][1-g_e(\theta_e)]\,\varepsilon^{P,cross}_{Q_{-ij}}(\theta_e) \\ &-\left[1-\xi\frac{\varepsilon^{\varpi}_{l_e}(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)}\right]\varepsilon^{P,cross}_{Q_{-ij}}(\theta_e)\,[1-\bar{g}_e(\theta_e)]\,H(\theta_e),\end{aligned} \tag{A63}$$

Notice that $1-\xi\frac{\varepsilon^{\varpi}_{l_e}(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)}=\frac{1}{1-\xi\frac{\sigma-1}{\sigma}}$. Equation (A63) is equivalent to (36) in the Atkeson-Burstein economy.■

C Extensions and Robustness

C.1 Alternative Market and Taxation Specifications

In what follows, we address the relation of our results to 4 distinct issue that have been studied in the literature. We do this through the lens of our model, and where needed, using the notation and features of our setup to derive new results.

(i) Top Incomes: Market Power and Optimal Profit Tax. Taxes payed by top earners account for the vast majority of income tax. Proposition 6 provides an analytic optimal top profit tax formula. We call entrepreneurs with $\theta_e \geq \widehat{\theta}_e$ the top entrepreneurs. We assume that the profit and sales tax rates in the real economy for the top entrepreneurs are linear and lower than one. Then, we have the following result:

Proposition 6 *Suppose that there exist $\widehat{\theta}_e \in \Theta_e$ such that for $\theta_e \geq \widehat{\theta}_e$, $H(\theta_e) = \widehat{H}$, $\mu(\theta_e) = \widehat{\mu}$ and $g_e(\theta_e) = \widehat{g}_e$ are constants. Then we have the following results in the Atkeson-Burstein economy:*

(i) The optimal profit tax rate for $\theta_e \geq \widehat{\theta}_e$ is constant and satisfy:

$$\frac{1}{1-\widehat{\tau}_e} = \frac{\frac{1+(1-\widehat{g}_e)\widehat{H}_e\left[\frac{1+\varepsilon_e}{\varepsilon_e}(\widehat{\mu}-\xi)-1\right]}{\widehat{\mu}} + \frac{1-\frac{\sigma}{\sigma-1}\frac{1}{\widehat{\mu}}}{\frac{\sigma}{\sigma-1}-\xi}(1-\widehat{g}_e)\left(1-\widehat{H}\right)}{1-\frac{\xi}{\frac{\sigma}{\sigma-1}-\xi}\frac{\mu-\widehat{\mu}}{\widehat{\mu}}}; \tag{A64}$$

(ii) Given σ, $\widehat{\tau}_e$ decreases in I (increases in $\widehat{\mu}$) if:

$$\widehat{g}_e < 1 - \frac{1}{2 \cdot \widehat{H}}. \tag{A65}$$

Proof. See Online Appendix OC.6. ■

Formula (A64) generalizes the traditional top income tax formula. It generalizes the familiar top tax rate result of Saez (2001) (where $\xi = 0$, $\mu = 1$, $I = 1$, $\sigma \to \infty$) by extending the technology and market structure. Compared to Corollary 5 of Sachs et al. (2020) (where $\xi = 0$, $\mu = 1$ and $I = 1$), this result highlights the effect of market structure and superstar effects on the optimal top tax rate. Compared to Scheuer and Werning (2017) (where $\mu = 1$, $I = 1$ and $\sigma \to \infty$), it demonstrates the interaction between superstar effect and market structure on the optimal top tax rate.

Proposition 6 is powerful in the sense that it suggests that under some reasonable assumptions, one can use original observable statistics like $\widehat{H}_e$ to derive the optimal top profit tax rate and judge whether the top profit tax should be increased with the changes of technology and market structure. A key question here, considering the reallocation and indirect redistribution effects, is: should the government increase the top profit tax rate with the rise of market power? Condition (A64) shows that whether the optimal top profit tax rate increases crucially depends on the value of $\widehat{H}_e$. Specifically, if $\widehat{g}_e \to 0$, condition (A65) is equivalent to $\widehat{H}_e > \frac{1}{2}$. Condition (A65) is a sufficient but not necessary condition for $\widehat{\tau}_e$ to be increasing in $\widehat{\mu}$.[48] At this

[48] In Online Appendix OC.2, we also provide a looser sufficient condition: $\widehat{g}_e < 1 - \frac{1}{\left[\xi\left(1-\mu\frac{\sigma-1}{\sigma}\right)\frac{1+\varepsilon_e}{\varepsilon_e}+1+\frac{\frac{\sigma}{\sigma-1}-\xi\mu\frac{\sigma-1}{\sigma}}{\frac{\sigma}{\sigma-1}-\xi}\right]\widehat{H}}$, where the term in the bracket of the right side of the above inequality is larger than or equal to 2, because $\mu \leq \frac{\sigma}{\sigma-1}$.

stage, it is important to recall that the hazard ratios of top income in the United States is around 0.5 in 1992 and 1993 (Saez (2001)) and around $\frac{2}{3}$ in 2005 (Diamond and Saez (2011)). In 2007, the hazard ratio of top labor, capital and total incomes in the United States are around 0.62, 0.76 and 0.71, respectively (see e.g., Saez and Stantcheva (2018)). In conclusion, empirical results suggests that in the U.S. condition (A65) is satisfied and the top profit tax rate should increase when market power increases.

(ii) Span of Control and Optimal Top Profit Tax. Like in our model, Scheuer and Werning (2017) also considers optimal income taxation in a Lucas (1978) span of control setting. They study the superstar effect induced by the positive assortative matching between the ability of entrepreneurs and the scale of firms. In their model, the span of control captures the magnitude of the superstar effect as it represents the extent to which can entrepreneurs leverage the productivity of workers.

In our setting, Proposition 6 delivers interesting insights about how the span of control affects optimal taxation. In the light of Scheuer and Werning (2017), an increase in ξ changes the elasticity of profits with respect to the tax rate as well as the hazard ratio of profit. When $I = 1$ these two effects cancel each other out such that the top profit tax rate is independent of ξ. To see this, notice that when $I = 1$, the markup is uniform ($\mu = \widehat{\mu} = \frac{\sigma}{\sigma-1}$) and formula (A64) is reduced to:

$$\frac{1}{1-\widehat{\tau}_e} = \frac{1 + (1-\widehat{g}_e)\,\widehat{H}\left[\frac{1+\varepsilon_e}{\varepsilon_e}\left(\widehat{\mu}-\xi\right)-1\right]}{\widehat{\mu}} = \frac{1}{\widehat{\mu}}\left[1 + (1-\widehat{g}_e)\,\frac{1-F_e(\widehat{\theta}_e)}{f_e\left(\widehat{\theta}_e\right)}\frac{d\ln X_e(\widehat{\theta}_e)}{d\widehat{\theta}_e}\frac{\sigma-1}{\sigma}\frac{1+\varepsilon_e}{\varepsilon_e}\right], \quad \text{(A66)}$$

where the second equality follows from the fact that $\widehat{H} = \frac{1-F_e(\widehat{\theta}_e)}{f_e\left(\widehat{\theta}_e\right)}\frac{\frac{\sigma-1}{\sigma}\frac{1+\varepsilon_e}{\varepsilon_e}\frac{d\ln X_e(\widehat{\theta}_e)}{d\widehat{\theta}_e}}{\frac{1+\varepsilon_e}{\varepsilon_e}\left(\frac{\sigma}{\sigma-1}-\xi\right)-1}$. The equation above establishes that $\widehat{\tau}_e$ is independent of ξ in a monopoly competitive economy. In other words, the "neutrality" of the span of control found by Scheuer and Werning (2017) still holds.

The "neutrality" of the span of control does not necessarily hold in a more general setting. As an illustration, we consider a special case where the top markup is $\widehat{\mu} = \frac{\sigma}{\sigma-1}$. We consider this case because $\frac{\sigma}{\sigma-1}$ is the theoretical maximum value of markup in our model. Substitute $\widehat{H}_e$ in formula (A64) by $\widehat{H} = \frac{1}{\widehat{\mu}-\xi-\frac{\varepsilon_e}{1+\varepsilon_e}}\frac{\gamma_e'\left(\widehat{\theta}_e\right)}{\gamma_e\left(\widehat{\theta}_e\right)}\frac{1-F_e(\widehat{\theta}_e)}{f_e\left(\widehat{\theta}_e\right)}$. As the markup for $\theta_e > \widehat{\theta}_e$ is assumed to be constant and $\widehat{\mu} = \frac{\sigma}{\sigma-1}$, one has:

$$\frac{1}{1-\widehat{\tau}_e} = \frac{\frac{1}{\widehat{\mu}}\left[1 + (1-\widehat{g}_e)\,\frac{1-F_e(\widehat{\theta}_e)}{f_e\left(\widehat{\theta}_e\right)}\frac{d\ln X_e(\widehat{\theta}_e)}{d\widehat{\theta}_e}\frac{\sigma-1}{\sigma}\frac{1+\varepsilon_e}{\varepsilon_e}\right]}{1+\left[\frac{\frac{\sigma}{\sigma-1}}{\frac{\sigma}{\sigma-1}-\xi}-1\right]\frac{\widehat{\mu}-\mu}{\widehat{\mu}}}.$$

It can be seen that in this case $\widehat{\tau}_e$ decreases in ξ, because the rising span of control enlarges the influence of reallocation effect, which can be seen from the multiplier on $RE\left(\theta_e\right)$ in formula (35), i.e., $\frac{\xi}{\frac{\sigma}{\sigma-1}-\xi}$.

(iii) Market Structure, Indirect Redistribution, and Optimal Taxation. One interesting finding of this paper is that the market structure is crucial for the optimal tax, and in particular for the indirect redistribution effect of taxation. In many previous studies on endogenous prices and optimal taxation, taxes have a first-order effect on relative prices and can thus be used to ease the incentive constraints and improve income distribution (see e.g., Naito (1999); Stiglitz (2018); Sachs et al. (2020); Cui et al. (2021)). Specifically, when the marginal productivity of the labor factor (wage) decreases with labor inputs, the planner can compress the wage distribution by reducing the marginal tax rate of high-skilled agents and enhancing the high-skilled agents' labor supply. Saez (2004) argues that tax's indirect redistribution effect collapses when agents make endogenous human capital investments. In that case, agents determine their wages, and the tax's effect on prices becomes second order. In response to Saez (2004), Naito (2004) shows that when human capital is imperfectly substitutable, the indirect redistribution effect is still in place.

Our findings contribute to this debate by demonstrating that the indirect redistribution of tax depends on the market structure. Even when there is market power, there is indirect redistribution, but the amount decreases as market power increases. In fact, the IRE completely disappears under monopoly, i.e., when the firms set their prices alone. As long as prices are determined completely (under competition) or partially (under oligopoly) outside the firm, there is a role for IRE. This discovery explains the seemingly contradictory conclusions obtained by Saez (2004) and Naito (2004). Saez (2004) considers the case where the agent is the monopolistic supplier of its own factors, Naito (2004) considers a competitive labor market.

(iv) Endogenous Social Welfare Weights. In most of the analysis above, we assume that the social welfare weights are exogenous. We thus abstract from the influence of market power on the optimal taxation through the social welfare weights. This allows us to highlight the other four elements.

Our analysis of the Laissez-faire economy suggests that the gross utility of entrepreneurs generally increases with market power. As we see from Proposition 2, a rise in market power accompanied by an increase in μ leads to a redistribution of income from workers to firms (mainly through a lower wage rate W) as well as a decrease in welfare. If this is also the case under optimal taxation, then the marginal social welfare weights for entrepreneurs will decrease which in turn increases the optimal profit tax rates. This finding has been emphasized by previous studies (see e.g., Kushnir and Zubrickas (2019)). Moreover, the non-linear tax system facilitates a transfer between entrepreneurs and workers, which also depends on the social welfare function. Under a utilitarian social welfare function, the burden is indeterminate, whereas it is determinate under a concave social welfare function, in which case market power plays a key role.

A reason for segregating the effect of endogenous social welfare weights is that a generalized social welfare weight may depend on factors other than the gross utility, and those factors may also change with the market structure (see Saez and Stantcheva (2016)). For example, a generalized social welfare weight may depend on the revenue contribution of an entrepreneur relative to that of a worker (see Scheuer (2014)). More generally, it depends on the gap between the social and private values of being an entrepreneur. By segregating the endogeneity of the social welfare weights, the criterion in Proposition 5 becomes applicable. Besides, we demonstrate that empirically, the optimal top profit tax rate increases with the markup. This finding is in line with previous studies.

C.2 Alternative Technology Specifications

(i) Capital Investment. We do not explicitly model capital in our benchmark model. However, the the problem can be modeled equivalently with capital in place of entrepreneurial effort. The most relevant assumption is that part of the cost (or benefit) from factors cannot be deducted before the profit tax (either because the cost is unobservable or legally excluded from the deductible costs). Formally, consider an economy where the entrepreneur chooses labor inputs L_w and capital investment K, instead of effort:

$$\max_{K,L_w} \; P_{ij}\left(Q_{ij}\left(K,L_w\right),Q_{-ij}\left(\theta_e\right),\theta_e\right)Q_{ij}\left(K,L_w\right)-WL_w-rK-\phi_K\left(K,\theta_e\right)-T_e\left(y_e\right)$$

$Q_{ij}\left(K,L_w\right)$ is the firm-level production function of capital and labor inputs, r is the market price of capital,[49] and $\phi_K\left(K,\theta_e\right)$ is the unobservable cost of investment, which may depend on the entrepreneur's type.

In the real economy, although the market price of capital (i.e., r) can be observed, the total opportunity costs of investments are typically hard to measure. The unobservable part of cost is captured by $\phi_K\left(K,\theta_e\right)$, which may include the cost of raising and managing funds.[50] An alternative explanation for $\phi_K\left(K,\theta_e\right)$ is the preference for asset (wealth). In that case, $\phi_K\left(K,\theta_e\right)$ can be negative, which means investment directly generates positive utility. The common ground in these situations is that the elasticity of investment may be finite, which is the key point of Saez and Stantcheva (2018), in which case, $y_e=P_{ij}\left(Q_{ij}\left(K,L_w\right),Q_{-ij}\left(\theta_e\right),\theta_e\right)Q_{ij}\left(K,L_w\right)-WL_w-rK$.

The incomplete deductibility of investment is relevant to the real economy. For example, the interest of debt is deductible before tax, but the equity investment is not. Equity investments affect the cash flow of shareholders and generate costs, in which case, $y_e=P_{ij}\left(Q_{ij}\left(K,L_w\right),Q_{-ij}\left(\theta_e\right),\theta_e\right)Q_{ij}\left(K,L_w\right)-WL_w$. Then, even if $\phi_K=0$ there are non-deductible capital costs before tax. That's why the profits tax has often been interpreted as a tax on capital on the production side (see e.g., chapter 8 in Myles (2008)).

In all the cases above, our main results continue to hold. Essentially, the key to the incentive problem is the unobservability of inputs. Moreover, we can model both deductible and non-deductible inputs. The model therefore captures the main elements behind the profit tax. It is worth noting that the optimal profit tax formula provided in this paper is independent of factor inputs and therefore has a wider application.

(ii) Performance Pay and Optimal Profit Tax. In the real economy, the entrepreneur may only obtain a part of the profit. We now show how profit sharing affects optimal taxation. Assuming that a portion (s) of the company's profits are paid to entrepreneurs through performance pay. The entrepreneur's problem is:

$$\begin{aligned} V_{e,ij}\left(\theta_e\right) &\equiv \max_{l_{e,ij},L_{w,ij}} \; c_e-\phi_e\left(l_e\right) \\ \text{s.t. } c_{e,ij} &= \left[y_{e,ij}-T_e\left(y_{e,ij}\right)\right]\cdot sy_{e,ij} \\ &= \left(1-t_s\right)P_{ij}\left(Q_{ij},\left\{Q_{-ij}\left(\theta_e\right)\right\}_{-i\neq i},\theta_e\right)Q_{ij}-WL_{w,ij}, \end{aligned}$$

where s is the share of profit to the entrepreneur. We assume that the remaining profits are evenly distributed among taxpayers (or households).

In this case, the planner's problem remains the same and so is the constrained optimal allocation. The

[49]The model can easily be extended to be dynamic, where the introduction of K and r will be more intuitive (e.g., see Cui et al. (2021)). Alternatively, one can consider a small open economy, where r is exogenous, or one can introduce a technology for the production of capital, which will also fix r. In the latter case, we can assume that the final goods can be used as either consumption goods or investments and the conversion rate between consumption and investment is one. Then $r=1$, and the social resource constraint is transformed to be

$$Q-N_e\int_{\theta_e}K(\theta_e)f_e(\theta_e)d\theta_e-\sum_{o\in\{e,w\}}N_o\int_{\theta^o}c_o(\theta_e)f_o(\theta_e)d\theta_e-R\geq 0, \tag{A67}$$

where $K\left(\theta_e\right)$ is the investment of θ_e firms.

[50]Under this illustration, $\phi_K\left(K,\theta_e\right)$ can still be treated as the utility cost of entrepreneurial effort, where the entrepreneurs use their knowledge to manage the factor inputs (more generally, one can take $\phi_K\left(K,L_w,\theta_e\right)$).

optimal taxation formula is modified to take performance pay into consideration. To see this, notice that the tax wedges now satisfy $\tau_s(\cdot) = t_s$, $\tau_w(\theta_w) = T_w'(y_w(\theta_w))$ and $\tau_e(\theta_e) = 1 - (1-t_s)\left[1 - T_e'(y_e(\theta_e))\right] \cdot s$. Therefore, introducing s won't change the effective tax rate on the effort of the entrepreneur but proportionally increase $1 - T_e'(y_e(\theta_e))$. Our main results therefore still hold.

(iii) Monopolistic Competition with Kimball Aggregation and Endogenous Markups. In our benchmark model, we consider a technology with constant elasticity of substitution. Even though markups are endogenous, we find that tax policies do not alter the equilibrium markup. In this section, we consider a technology with non-constant elasticity of substitution, i.e., using Kimball aggregation. We show taxes can now affect markups. For tractability, we consider monopolistic competition and the second-best allocation.

The technology is described below:

$$1 = \int_{\theta_e} \chi(\theta_e)\, \Upsilon\left(\frac{Q(\theta_e)}{Q/N_e}\right) dF_e(\theta_e), \tag{A68}$$

where $Q(\theta_e) = x_e(\theta_e) l_e(\theta_e) L_w(\theta_e)^{\xi}$, $\Upsilon(\cdot)$ is a twice differentiable function, and Q is the quantity of final goods. Under the above technology

$$P(\theta_e) = \frac{\chi(\theta_e)\, \Upsilon'\left(\frac{Q(\theta_e)}{Q/N_e}\right)}{\int_{\theta_e} \chi(\theta_e)\, \Upsilon'\left(\frac{Q(\theta_e)}{Q/N_e}\right) \frac{Q(\theta_e)}{Q/N_e} dF_{\theta_e}(\theta_e)} \tag{A69}$$

$$\text{and} \quad \mu(\theta_e) = \frac{\varepsilon(\theta_e)}{\varepsilon(\theta_e) - 1} \text{ with } \varepsilon(\theta_e) = -\frac{\Upsilon''\left(\frac{Q(\theta_e)}{Q}\right)\frac{Q(\theta_e)}{Q}}{\Upsilon'\left(\frac{Q(\theta_e)}{Q}\right)}. \tag{A70}$$

The markup $\mu(\theta_e)$ is a function of $\frac{Q(\theta_e)}{Q}$. According to Lemma 1, the incentive compatible condition of the entrepreneur is:

$$V_e'(\theta_e) = \phi_e'(l_e(\theta_e))\, l_e(\theta_e)\left[\mu(\theta_e)\frac{\chi'(\theta_e)}{\chi(\theta_e)} + \frac{x_e'(\theta_e)}{x_e(\theta_e)}\right], \forall \theta_e \in \Theta_e. \tag{A71}$$

The planner chooses $\{l_e(\theta_e), L_w(\theta_e), V_e(\theta_e), l_w(\theta_w), V_w(\theta_w), Q\}_{\theta_e \in \Theta_e, \theta_w \in \Theta_w}$ to maximize (5) subject to the resource constraints (A68) and (13), the labor market clear condition (14), and the incentive conditions (A71) and (26). As a comparison to the optimal profit tax under monopolistic competition in the benchmark model, we now have the following proposition:

Proposition 7 *Under monopolistic competition with Kimball aggregation, the effective tax rate on entrepreneurial effort satisfies:*

$$\frac{1 - \frac{1-\tau_e(\theta_e)}{\mu(\theta_e)}}{\frac{1-\tau_e(\theta_e)}{\mu(\theta_e)}} = \left[1 - \bar{g}_e(\theta_e)\right]\frac{1 - F_e(\theta_e)}{f_e(\theta_e)}\left[\begin{array}{c} \frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu(\theta_e)\frac{\chi'(\theta_e)}{\chi(\theta_e)} + \frac{x_e'(\theta_e)}{x_e(\theta_e)}\right] \\ +\mu(\theta_e)\frac{\partial \ln \mu(\theta_e)}{\partial \ln Q(\theta_e)}\frac{\chi'(\theta_e)}{\chi(\theta_e)} \end{array}\right], \tag{A72}$$

for any $\theta_e \in \Theta_e$.

Proof. See Online Appendix OC.7. ■

Compared to the monopoly competitive case in the benchmark model (see e.g., equation (42)), the optimal tax here takes the endogeneity of markups into consideration. When $\frac{\partial \ln \mu(\theta_e)}{\partial \ln Q(\theta_e)} > 0$, the additional term is positive, which suggests that higher markups generally require a higher tax rate. Then, our main finding – the top profit tax rate should increase with the rise of markup – will remain unchanged with varying elasticities of substitution.

(iv) Free Entry. In the current paper, we assume the number of incumbents in a market is exogenous. Thus, our setup misses the impact of taxes on markups and efficiency through the extensive margin of entry, which is important for policy in the real world. Nonetheless, we consider that fixing the number of incumbents is a reasonable place to start the analysis, because introducing the extensive margin adds a complicated entry game to the analysis, both analytically and computationally. In related work, De Loecker et al. (2019) analyzes a model with endogenous markups à la Atkeson and Burstein (2008) and with entry. The entry game there builds on Berry (1992) and involves a loop to compare all possible alternative entry configurations. In conjunction with incomplete information and optimal taxation, the solution would become intractable. There is scope for future research to introduce the entry of firms and consider tax's effect on both margins, so as to give more detailed policy recommendations. There are several ways to introduce the extensive margin. For example, one can introduce occupational choice as in Scheuer (2014) and Rothschild and Scheuer (2013). Edmond et al. (2023) consider an economy with free entry and Kimball demand. Last, one can consider entry by introducing a fringe of small businesses to each market.

(v) Uniform Income Tax. We currently consider different tax policies for labor income and profit. However, the government may not be able to perfectly distinguish labor income from profit. Consider therefore a uniform tax on profit and labor income. The policy constraint leads to a rather complicated problem. As an illustration, denote by $\theta_w(\theta_e)$ the ability of a worker whose income is $y_e(\theta_e)$ (i.e., $y_e(\theta_e) = y_w(\theta_w(\theta_e))$). Then the first-order conditions of the agents' optimization imply:

$$1 - T'_w(y_e(\theta_e)) = \frac{\phi'_w\left(\frac{y_e(\theta_e)}{W\varkappa_w(\theta_w(\theta_e))}\right)}{W\varkappa_w(\theta_w(\theta_e))}, \quad \text{and} \quad 1 - T'_e(y_e(\theta_e)) = \frac{\phi'_e(l_e(\theta_e))}{\frac{P(\theta_e)}{\mu(\theta_e)}\frac{\partial Q_{ij}(\theta_e)}{\partial l_e(\theta_e)}}. \tag{A73}$$

Therefore, under the uniform tax on labor income and profit, the following policy constraint should be enforced:

$$\frac{\phi'_w\left(\frac{y_e(\theta_e)}{W\varkappa_w(\theta_w(\theta_e))}\right)}{W\varkappa_w(\theta_w(\theta_e))} = \frac{\phi'_e(l_e(\theta_e))}{\frac{P(\theta_e)}{\mu(\theta_e)}\frac{\partial Q_{ij}(\theta_e)}{\partial l_e(\theta_e)}},$$

where $y_e(\theta_e) = P\left(Q_{ij}(\theta_e), \theta_e\right) Q_{ij}(\theta_e) - WL_w(\theta_e)$. Meanwhile, $\theta_w(\theta_e)$ should be treated as an additional variable chosen by the planner (see Fu et al. (2021) for a solution to this problem).

In addition to the complexity of the problem with uniform taxes, there is a second, conceptual reason, why we solve differentiated taxes in our benchmark problem. Depending on how we interpret entrepreneurship, part of the income of entrepreneurs is subject to corporate taxes which is clearly distinct from income taxation. Of course, even with corporate taxation, entrepreneurs eventually need to declare profits as income as well, but even then it is distinct from labor income. Therefore, the real world makes the case for considering differentiated taxation. In fact, both differentiated and uniform taxation are contemplated in the literature (see Scheuer (2014) and Rothschild and Scheuer (2013)).[51] From our perspective, it is a reasonable starting point to consider differentiated taxation to analyze the role of rising market power.

[51] Scheuer (2014) considers differential taxation on labor and profit income while Rothschild and Scheuer (2013) considers uniform taxation.

References

AKCIGIT, U., D. HANLEY, AND S. STANTCHEVA (2016): “Optimal Taxation and R&D Policies,” Working Paper 22908, National Bureau of Economic Research, series: Working Paper Series.

ALES, L., A. A. BELLOFATTO, AND J. J. WANG (2017): “Taxing Atlas: Executive compensation, firm size, and their impact on optimal top income tax rates,” *Review of Economic Dynamics*, 26, 62–90.

ALES, L., M. KURNAZ, AND C. SLEET (2015): “Technical Change, Wage Inequality, and Taxes,” *The American Economic Review*, 105, 3061–3101, publisher: American Economic Association.

ALES, L. AND C. SLEET (2016): “Taxing Top CEO Incomes,” *American Economic Review*, 106, 3331–3366, publisher: American Economic Association.

ANDERSON, S. P., A. D. PALMA, AND B. KREIDER (2001): “Tax incidence in differentiated product oligopoly,” *Journal of Public Economics*, 81, 173–192.

ATESAGAOGLU, O. E. AND H. YAZICI (2021): “Optimal Taxation of Capital in the Presence of Declining Labor Share,” Tech. rep., Bristol University mimeo.

ATKESON, A. AND A. BURSTEIN (2008): “Pricing-to-Market, Trade Costs, and International Relative Prices,” *American Economic Review*, 98, 1998–2031.

ATKINSON, A. B. (2015): *Inequality: What can be done?*, Harvard University Press.

ATKINSON, A. B. AND J. E. STIGLITZ (1976): “The design of tax structure: direct versus indirect taxation,” *Journal of Public Economics*, 6, 55–75.

AUTOR, D., D. DORN, L. F. KATZ, C. PATTERSON, AND J. VAN REENEN (2020): “The Fall of the Labor Share and the Rise of Superstar Firms,” *Quarterly Journal of Economics*, 135, 645–709.

BAKER, J. B. AND S. C. SALOP (2015): “Antitrust, competition policy, and inequality,” *Geo. LJ Online*, 104, 1.

BERRY, S., J. LEVINSOHN, AND A. PAKES (1995): “Automobile prices in market equilibrium,” *Econometrica: Journal of the Econometric Society*, 841–890.

BERRY, S. T. (1992): “Estimation of a Model of Entry in the Airline Industry,” *Econometrica*, 889–917.

BOAR, C. AND V. MIDRIGAN (2021): “Markups and inequality,” Tech. rep., NYU mimeo.

BOVENBERG, A. L. AND F. VAN DER PLOEG (1994): “Environmental policy, public finance and the labour market in a second-best world,” *Journal of Public Economics*, 55, 349–390.

BRODA, C. AND D. E. WEINSTEIN (2006): “Globalization and the Gains from Variety,” *The Quarterly journal of economics*, 121, 541–585.

CHARI, V. V. AND P. J. KEHOE (1999): “Optimal fiscal and monetary policy,” in *Handbook of Macroeconomics*, Elsevier, vol. 1, chap. 26, 1671–1745.

CHETTY, R. (2012): “Bounds on Elasticities With Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply,” *Econometrica*, 80, 969–1018.

CHETTY, R., A. GUREN, D. MANOLI, AND A. WEBER (2011): "Are micro and macro labor supply elasticities consistent? A review of evidence on the intensive and extensive margins," *American Economic Review*, 101, 471–75.

CHRISTIANSEN, V. (1984): "Which commodity taxes should supplement the income tax?" *Journal of Public Economics*, 24, 195–220.

COLCIAGO, A. (2016): "Endogenous market structures and optimal taxation," *The Economic Journal*, 126, 1441–148.

COSTINOT, A. AND I. WERNING (2023): "Robots, Trade, and Luddism: A Sufficient Statistic Approach to Optimal Technology Regulation," *The Review of Economic Studies*, 90, 2261–2291.

CREMER, H. AND J.-F. THISSE (1994): "Commodity taxation in a differentiated oligopoly," *International Economic Review*, 35, 613–633.

CUI, X., L. GONG, AND W. LI (2021): "Supply-side optimal capital taxation with endogenous wage inequality," *Journal of Public Economics*, 198, 104421.

DE LOECKER, J., J. EECKHOUT, AND S. MONGEY (2019): "Quantifying Market Power," Tech. rep., Working Paper.

DE LOECKER, J., J. EECKHOUT, AND G. UNGER (2020): "The Rise of Market Power and the Macroeconomic Implications," *Quarterly Journal of Economics*, 135, 561–644.

DE LOECKER, J. AND F. M. P. WARZYNSKI (2012): "Markups and Firm-level Export Status," *American Economic Review*, 102, 2437–2471.

DIAMOND, P. AND E. SAEZ (2011): "The Case for a Progressive Tax: From Basic Research to Policy Recommendations," *Journal of Economic Perspectives*, 25, 165–190.

DIAMOND, P. A. (1998): "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates," *American Economic Review*, 88, 83–95.

DIAMOND, P. A. AND J. A. MIRRLEES (1971a): "Optimal Taxation and Public Production I: Production Efficiency," *The American Economic Review*, 61, 8–27, publisher: American Economic Association.

——— (1971b): "Optimal Taxation and Public Production II: Tax Rules," *The American Economic Review*, 61, 261–278, publisher: American Economic Association.

DIXIT, A. (1985): "Chapter 6 Tax policy in open economies," in *Handbook of Public Economics*, Elsevier, vol. 1, 313–374.

EDMOND, C., V. MIDRIGAN, AND D. Y. XU (2019): "How costly are markups?" Tech. rep., National Bureau of Economic Research.

——— (2023): "How Costly Are Markups?" *Journal of Political Economy*, 131, 1619–1675, publisher: The University of Chicago Press.

EECKHOUT, J. AND P. KIRCHER (2018): "Assortative matching with large firms," *Econometrica*, 86, 85–132.

FARHI, E. AND X. GABAIX (2020): "Optimal Taxation with Behavioral Agents," *American Economic Review*, 110, 298–336.

FINDEISEN, S. AND D. SACHS (2017): "Redistribution and insurance with simple tax instruments," *Journal of Public Economics*, 146, 58–78, publisher: Elsevier.

FU, C., L. GONG, AND W. LI (2021): "Optimal Central Redistribution with Decentralized Redistribution," SSRN Working paper.

GOLOSOV, M., N. KOCHERLAKOTA, AND A. TSYVINSKI (2003): "Optimal Indirect and Capital Taxation," *The Review of Economic Studies*, 70, 569–587.

GOLOSOV, M., A. TSYVINSKI, AND N. WERQUIN (2016): "Chapter 10 - Recursive Contracts and Endogenously Incomplete Markets," in *Handbook of Macroeconomics*, ed. by J. B. Taylor and H. Uhlig, Elsevier, vol. 2, 725–841.

GÜRER, E. (2021): "Rising Markups and Optimal Redistributive Taxation," Tech. rep., Frankfurt University mimeo.

HEATHCOTE, J., K. STORESLETTEN, AND G. L. VIOLANTE (2017): "Optimal Tax Progressivity: An Analytical Framework*," *The Quarterly Journal of Economics*, 132, 1693–1754.

HENDEL, I. AND A. NEVO (2006): "Measuring the implications of sales and consumer inventory behavior," *Econometrica*, 74, 1637–1673.

JACQUET, L., E. LEHMANN, AND B. VAN DER LINDEN (2013): "Optimal redistributive taxation with both extensive and intensive responses," *Journal of Economic Theory*, 148, 1770–1805.

JARAVEL, X. AND A. OLIVI (2019): "Optimal Taxation and Demand-Led Productivity," SSRN Scholarly Paper ID 3383693, Social Science Research Network, Rochester, NY.

KAPLOW, L. (2019): "Market Power and Income Taxation," Tech. Rep. w25578, National Bureau of Economic Research.

KARABARBOUNIS, L. AND B. NEIMAN (2014): "The Global Decline of the Labor Share*." *Quarterly Journal of Economics*, 129.

KATZ, L. F. AND K. M. MURPHY (1992): "Changes in relative wages, 1963–1987: supply and demand factors," *The Quarterly Journal of Economics*, 107, 35–78.

KHAN, L. M. AND S. VAHEESAN (2017): "Antitrust, competition policy, and inequality," *Harvard Law and Policy Review*, 11, 235.

KOPCZUK, W. (2003): "A note on optimal taxation in the presence of externalities," *Economics Letters*, 80, 81–86.

KUSHNIR, A. AND R. ZUBRICKAS (2019): "Optimal Income Taxation with Endogenous Prices," SSRN Scholarly Paper ID 3445132, Social Science Research Network, Rochester, NY.

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

LUENBERGER, D. G. (1997): *Optimization by Vector Space Methods*, New York: Wiley-Interscience, 1969th edition ed.

MELITZ, M. J. (2003): "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 71, 1695–1725.

MIRRLEES, J. A. (1971): "An Exploration in the Theory of Optimum Income Taxation," *The Review of Economic Studies*, 38, 175–208.

——— (1976): "Optimal tax theory: A synthesis," *Journal of Public Economics*, 6, 327–358.

MYLES, G. D. (1989): "Ramsey tax rules for economies with imperfect competition," *Journal of Public Economics*, 38, 95–115.

——— (2008): *Public Economics*, Cambridge University Press, illustrated edition ed.

NAITO, H. (1999): “Re-examination of uniform commodity taxes under a nonlinear income tax system and its implication for production efficiency,” *Journal of Public Economics*, 71, 165–188.

——— (2004): “Endogenous human capital accumulation, comparative advantage and direct vs. indirect redistribution,” *Journal of Public Economics*, 88, 2685–2710.

NG, Y.-K. (1980): “Optimal Corrective Taxes or Subsidies When Revenue Raising Imposes an Excess Burden,” *The American Economic Review*, 70, 744–751, publisher: American Economic Association.

PIKETTY, T. (1997): “La redistribution fiscale face au chômage,” *Revue française d'économie*, 12, 157–201, publisher: Persée - Portail des revues scientifiques en SHS.

PRESCOTT, E. C. AND M. VISSCHER (1980): “Organization capital,” *Journal of political Economy*, 88, 446–461.

ROSEN, S. (1981): “The Economics of Superstars,” *American Economic Review*, 71, 845–858.

ROTHSCHILD, C. AND F. SCHEUER (2013): “Redistributive Taxation in the Roy Model*,” *Quarterly Journal of Economics*, 128, 623–668.

——— (2016): “Optimal Taxation with Rent-Seeking,” *The Review of Economic Studies*, 83, 1225–1262.

SACHS, D., A. TSYVINSKI, AND N. WERQUIN (2016): “Nonlinear Tax Incidence and Optimal Taxation in General Equilibrium,” Working Paper 6089, CESifo Working Paper.

——— (2020): “Nonlinear tax incidence and optimal taxation in general equilibrium,” *Econometrica*, 88, 469–493.

SAEZ, E. (2001): “Using Elasticities to Derive Optimal Income Tax Rates,” *The Review of Economic Studies*, 68, 205–229.

——— (2004): “Direct or indirect tax instruments for redistribution: Short-run versus long-run,” *Journal of Public Economics*, 88, 503–518.

SAEZ, E. AND S. STANTCHEVA (2016): “Generalized Social Marginal Welfare Weights for Optimal Tax Theory,” *The American Economic Review*, 106, 24–45.

——— (2018): “A simpler theory of optimal capital taxation,” *Journal of Public Economics*, 162, 120–142, publisher: Elsevier.

SALANIÉ, B. (2003): *The economics of taxation*, The MIT Press.

SANDMO, A. (1975): “Optimal Taxation in the Presence of Externalities,” *The Swedish Journal of Economics*, 77, 86–98, publisher: [Scandinavian Journal of Economics, Wiley].

SATTINGER, M. (1975a): “Comparative advantage and the distributions of earnings and abilities,” *Econometrica: Journal of the Econometric Society*, 455–468.

——— (1975b): “Comparative advantage and the distributions of earnings and abilities,” *Econometrica: Journal of the Econometric Society*, 455–468.

——— (1993): “Assignment Models and the Distribution of Earnings,” *Journal of Economic Literature*, 31, 831–880.

SCHEUER, F. (2014): “Entrepreneurial taxation with endogenous entry,” *American Economic Journal: Economic Policy*, 6, 126–63.

SCHEUER, F. AND I. WERNING (2017): "The taxation of superstars," *The Quarterly Journal of Economics*, 132, 211–270.

STERN, N. (1987): "The effects of taxation, price control and government contracts in oligopoly and monopolistic competition," *Journal of Public Economics*, 32, 133–158.

STIGLITZ, J. E. (1982): "Self-selection and Pareto efficient taxation," *Journal of Public Economics*, 17, 213–240.

——— (2012): *The price of inequality: How today's divided society endangers our future*, W. W. Norton & Company.

——— (2018): "Pareto efficient taxation and expenditures: Pre- and re-distribution," *Journal of Public Economics*, 162, 101–119.

SUTTON, J. (1991): *Sunk costs and market structure: Price competition, advertising, and the evolution of concentration*, MIT press.

——— (2001): *Technology and market structure: theory and history*, MIT press.

TERVIO, M. (2008): "The Difference that CEOs Make: An Assignment Model Approach," *American Economic Review*, 98(3), \mbox642–668.

Online Appendix

OA Supplements to Environment

OA.1 Elasticity of Profit to the Skill

In this section, we provide elasticity of profit with respect to the skill. Remember that entrepreneur's FOCs imply:

$$\phi_e'\left(l_e\left(\theta_e\right)\right)=\left[1-T_e'\left(\left(\frac{\mu\left(\theta_e\right)}{\xi}-1\right)WL_w\left(\theta_e\right)\right)\right]\frac{WL_w\left(\theta_e\right)}{\xi}\frac{1}{l_e\left(\theta_e\right)},$$

and

$$WL_w\left(\theta_e\right)=\left(1-t_s\right)P\left(\theta_e\right)Q_{ij}\left(\theta_e\right)\frac{\xi}{\mu\left(\theta_e\right)}.$$

Take the derivative of both sides of the above equations with respect to θ:

$$\left(1+\frac{1}{\varepsilon_e}\right)\frac{l_e'\left(\theta_e\right)}{l_e\left(\theta_e\right)}=\frac{L_w'\left(\theta_e\right)}{L_w\left(\theta_e\right)}\left(1-\frac{T_e''(y_e\left(\theta_e\right))y_e\left(\theta_e\right)}{1-T_e'\left(y_e\left(\theta\right)\right)}\right)-\frac{T_e''(y_e\left(\theta_e\right))y_e\left(\theta_e\right)}{1-T_e'\left(y_e\left(\theta_e\right)\right)}\frac{\frac{\mu'(\theta_e)}{\xi}}{\frac{\mu(\theta_e)}{\xi}-1}$$

and

$$\frac{L_w'\left(\theta_e\right)}{L_w\left(\theta_e\right)}\left(1-\xi\varepsilon^S_{Q_{ij}}\left(\theta_e\right)\right)=\left[\frac{x_e'(\theta_e)}{x_e(\theta_e)}+\frac{l_e'\left(\theta_e\right)}{l_e\left(\theta_e\right)}\right]\varepsilon^S_{Q_{ij}}\left(\theta_e\right)+\frac{\chi'(\theta_e)}{\chi(\theta_e)}-\frac{\mu'\left(\theta_e\right)}{\mu\left(\theta_e\right)},\tag{OA1}$$

where we use

$$\frac{Q_{ij}'(\theta_e)}{Q_{ij}(\theta_e)}=\frac{x_e'(\theta_e)}{x_e(\theta_e)}+\frac{l_e'\left(\theta_e\right)}{l_e\left(\theta_e\right)}+\xi\frac{L_w'\left(\theta_e\right)}{L_w\left(\theta_e\right)}.\tag{OA2}$$

Combine the above the equations:

$$\frac{L_w'\left(\theta_e\right)}{L_w\left(\theta_e\right)}=\frac{\frac{\frac{1+\varepsilon_e}{\varepsilon_e}}{\varepsilon^S_{Q_{ij}}(\theta_e)}\left[\frac{X_e{}'(\theta_e)}{X_e(\theta_e)}-\frac{\mu'(\theta_e)}{\mu(\theta_e)}\right]-\frac{T_e''(y_e(\theta))y_e(\theta)}{1-T_e'(y_e(\theta))}\frac{\frac{\mu'(\theta_e)}{\xi}}{\frac{\mu(\theta_e)}{\xi}-1}}{\frac{1+\varepsilon_e}{\varepsilon_e}\left[\frac{1}{\varepsilon^S_{Q_{ij}}(\theta_e)}-\xi\right]-\left[1-\frac{T_e''(y_e(\theta))y_e(\theta)}{1-T_e'(y_e(\theta))}\right]}.$$

Last, notice that $y_e\left(\theta_e\right)=WL_w\left(\theta_e\right)\left(\frac{\mu(\theta_e)}{\xi}-1\right)$. We have:

$$\frac{y_e'\left(\theta_e\right)}{y_e\left(\theta_e\right)}=\frac{\mu'\left(\theta_e\right)}{\mu\left(\theta_e\right)-\xi}+\frac{L_w'\left(\theta\right)}{L_w\left(\theta\right)},\forall\theta_e\in\Theta_e.\tag{OA3}$$

Combine the above two equations:

$$\frac{y_e'\left(\theta_e\right)}{y_e\left(\theta_e\right)}=\frac{\frac{\frac{1+\varepsilon_e}{\varepsilon_e}}{\varepsilon^S_{Q_{ij}}(\theta_e)}\frac{d}{d\theta_e}\left[\ln\frac{X_e(\theta_e)}{\mu(\theta_e)}\right]-\left[1-\left(\frac{1}{\varepsilon^S_{Q_{ij}}(\theta_e)}-\xi\right)\frac{1+\varepsilon_e}{\varepsilon_e}\right]\frac{\frac{\mu'(\theta_e)}{\xi}}{\frac{\mu(\theta_e)}{\xi}-1}}{\frac{1+\varepsilon_e}{\varepsilon_e}\left[\frac{1}{\varepsilon^S_{Q_{ij}}(\theta_e)}-\xi\right]-\left[1-\frac{T_e''(y_e(\theta))y_e(\theta)}{1-T_e'(y_e(\theta))}\right]}.$$

Use $\widetilde{\varepsilon}^{y_e}_{1-\tau_e}(\theta_e) = \frac{1}{\frac{1+\varepsilon_e}{\varepsilon_e}\frac{1-\frac{\sigma-1}{\sigma}\xi}{\frac{\sigma-1}{\sigma}} - \left[1-\frac{\pi T''_e(y_e)}{1-T'_e(y_e)}\right]}$:

$$\frac{y'_e(\theta_e)}{y_e(\theta_e)} = \widetilde{\varepsilon}^{y_e}_{1-\tau_e}(\theta_e)\left[\frac{\frac{1+\varepsilon_e}{\varepsilon_e}}{\frac{\sigma-1}{\sigma}}\frac{d}{d\theta_e}\left[\ln\frac{X_e(\theta_e)}{\mu(\theta_e)}\right] - \left[1-\left(\frac{1}{\frac{\sigma-1}{\sigma}}-\xi\right)\frac{1+\varepsilon_e}{\varepsilon_e}\right]\frac{\frac{\mu'(\theta_e)}{\xi}}{\frac{\mu(\theta_e)}{\xi}-1}\right].$$

OA.2 Solution to the Equilibrium

OA.2.1 Allocations and Prices

To simplify the expression, in the following analysis, we consider $t_s = 0$. Besides, we use $X_e(\theta_e) = A^{\frac{\sigma-1}{\sigma}}x_e(\theta_e)^{\frac{\sigma-1}{\sigma}}\chi(\theta_e)$ and $f_w(\theta_w) = f_e(\theta_e) = 1$ to shorten the expressions. Combination of (A4), (A5), and (A1) gives:

$$l_e(\theta_e) = \left[\frac{X_e(\theta_e)}{\mu(\theta_e)}\left(\frac{Q}{N_e}\right)^{\frac{1}{\sigma}}\right]^{\frac{\sigma\varepsilon_e}{\varepsilon_e+\sigma}} L_w(\theta_e)^{\frac{\xi(\sigma-1)\varepsilon_e}{\varepsilon_e+\sigma}}\left[1-\tau_e(\theta_e)\right]^{\frac{\varepsilon_e\sigma}{\sigma+\varepsilon_e}}. \tag{OA4}$$

Substituting $P(\theta_e)$ and $Q_{ij}(\theta_e)$ in (A4) with (A1) and $Q_{ij}(\theta_e) = x_e(\theta_e)\, l_e(\theta_e)\, L_w(\theta_e)^{\xi}$, respectively, we have:

$$\begin{aligned} L_w(\theta_e) &= \frac{\xi}{W\mu(\theta_e)}\chi(\theta_e)\left[x_e(\theta_e)l_e(\theta_e)\, L_w(\theta_e)^{\xi}\right]^{\frac{\sigma-1}{\sigma}}\left(\frac{Q}{N_e}\right)^{\frac{1}{\sigma}}\left[1-\tau_e(\theta_e)\right]^{\frac{\varepsilon_e(\sigma-1)}{\sigma+\varepsilon_e}} \\ &= \frac{\xi}{W}\left[\frac{X_e(\theta_e)}{\mu(\theta_e)}\left(\frac{Q}{N_e}\right)^{\frac{1}{\sigma}}\right]^{\frac{\sigma(\varepsilon_e+1)}{\varepsilon_e+\sigma}} L_w(\theta_e)^{\frac{\xi(\sigma-1)(\varepsilon_e+1)}{\sigma+\varepsilon_e}}\left[1-\tau_e(\theta_e)\right]^{\frac{\varepsilon_e(\sigma-1)}{\sigma+\varepsilon_e}}, \end{aligned} \tag{OA5}$$

where we substitute $l_e(\theta_e)$ with (OA4) in the second equation.

Rearranging the above equation gives:

$$L_w(\theta_e) = \left(\frac{\xi}{W}\right)^{\frac{\sigma+\varepsilon_e}{\sigma+\varepsilon_e-\xi(\sigma-1)(\varepsilon_e+1)}}\left\{\frac{X_e(\theta_e)\left[1-\tau_e(\theta_e)\right]^{\frac{(\sigma-1)\varepsilon_e}{\sigma(\varepsilon_e+1)}}}{\mu(\theta_e)}\left(\frac{Q}{N_e}\right)^{\frac{1}{\sigma}}\right\}^{\frac{\sigma(\varepsilon_e+1)}{\sigma+\varepsilon_e-\xi(\sigma-1)(\varepsilon_e+1)}}. \tag{OA6}$$

Substituting the above equation into (OA4), we have:

$$P(\theta_e)Q_{ij}(\theta_e) = \mu(\theta_e)\left(\frac{\xi}{W}\right)^{\frac{\xi(\sigma-1)(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}}\left\{\frac{X_e(\theta_e)\left[1-\tau_e(\theta_e)\right]^{\frac{\varepsilon_e(\sigma-1)}{\sigma(\varepsilon_e+1)}}}{\mu(\theta_e)}\left(\frac{Q}{N_e}\right)^{\frac{1}{\sigma}}\right\}^{\frac{\sigma(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}} \tag{OA7}$$

and

$$l_e(\theta_e) = \left(\frac{\xi}{W}\right)^{\frac{\xi(\sigma-1)\varepsilon_e}{\sigma+\varepsilon_e-\xi(\sigma-1)(\varepsilon_e+1)}}\left\{\frac{X_e(\theta_e)\left[1-\tau_e(\theta_e)\right]^{\frac{\sigma-\xi(\sigma-1)}{\sigma}}}{\mu(\theta_e)}\left(\frac{Q}{N_e}\right)^{\frac{1}{\sigma}}\right\}^{\frac{\sigma\varepsilon_e}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}}. \tag{OA8}$$

Equation (A3) gives:

$$l_w(\theta_w) = \left\{W x_w(\theta_w)\left[1-\tau_w(\theta_w)\right]\right\}^{\varepsilon_w}. \tag{OA9}$$

The three equations above together with (A7) and (A8) solve the symmetric equilibrium allocation $\{L_w(\theta_e), l_e(\theta_e), l_w(\theta_w)\}$, price system $\{P(\theta_e), W\}$, and total output Q. Lastly, one can derive other allocations with individuals' budget constraints. See below for details.

For later use, we define:

$$
\begin{aligned}
A_1 &= \int_{\theta_e} N\mu(\theta_e)\left[\frac{X_e(\theta_e)\left[1-\tau_e(\theta_e)\right]^{\frac{(\sigma-1)\varepsilon_e}{\sigma(\varepsilon_e+1)}}}{\mu(\theta_e)\, N_e^{\frac{1}{\sigma}}}\right]^{\frac{\sigma(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}} d\theta_e, && \text{(OA10)}\\
A_2 &= \int N\left[\frac{X_e(\theta_e)\left[1-\tau_e(\theta_e)\right]^{\frac{(\sigma-1)\varepsilon_e}{\sigma(\varepsilon_e+1)}}}{\mu(\theta_e)\, N_e^{\frac{1}{\sigma}}}\right]^{\frac{\sigma(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}} d\theta_e,\\
A_3 &= N_w \xi^{\varepsilon_w}\int_{\theta_w} x(\theta_w)^{\varepsilon_w+1}\left[1-\tau_w(\theta_w)\right]^{\varepsilon_w} d\theta_w.
\end{aligned}
$$

Substituting $L_w(\theta_e)$ in (A4) with (OA6), we have:

$$
\begin{aligned}
P(\theta_e)Q_{ij}(\theta_e) &= \mu(\theta_e)\left(\frac{\xi}{W}\right)^{\frac{\xi(\sigma-1)(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}}\left[\frac{X_e(\theta_e)}{\mu(\theta_e)}\left(\frac{Q}{N_e}\right)^{\frac{1}{\sigma}}\right]^{\frac{\sigma(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}} && \text{(OA11)}\\
&\quad\times\left[1-\tau_e(\theta_e)\right]^{\frac{(\sigma-1)\varepsilon_e}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}}.
\end{aligned}
$$

Substituting $P(\theta_e)Q_{ij}(\theta_e)$ in (A8) with (OA11), we have

$$
Q=\int_{\theta_e} N_e\mu(\theta_e)\left(\frac{\xi}{W}\right)^{\frac{\xi(\sigma-1)(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}}\left[\frac{X_e(\theta_e)}{\mu(\theta_e)}\left(\frac{Q}{N_e}\right)^{\frac{1}{\sigma}}\right]^{\frac{\sigma(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}}\left[1-\tau_e(\theta_e)\right]^{\frac{(\sigma-1)\varepsilon_e}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}} d\theta_e,
$$

which gives the following equation by the definition of A_1:

$$
Q=\left(\frac{\xi}{W}\right)^{\frac{\xi(\varepsilon_e+1)}{1-\xi(\varepsilon_e+1)}} A_1^{\frac{\sigma+\varepsilon_e-\xi(\sigma-1)(\varepsilon_e+1)}{1-\xi(\varepsilon_e+1)}\frac{1}{\sigma-1}}. \tag{OA12}
$$

Similarly, substituting $L_w(\theta_e)$ in (A7) with (OA6), we have the aggregate labor demand

$$
L^D \equiv \left(\frac{\xi}{W}\right)^{\frac{1}{1-\xi(\varepsilon_e+1)}}(A_1)^{\frac{\varepsilon_e+1}{(\sigma-1)[1-\xi(\varepsilon_e+1)]}} A_2. \tag{OA13}
$$

On the other hand, according to (A7) and (OA9), we have the aggregate labor supply

$$
L^S \equiv N_w\left[W\right]^{\varepsilon_w}\int_{\theta_w} x(\theta_w)^{\varepsilon_w+1} f_w(\theta_w)d\theta_w=\left[\frac{W}{\xi}\right]^{\varepsilon_w} A_3. \tag{OA14}
$$

Combining (OA13) and (OA14) gives

$$
\left[\frac{W}{\xi}\right]^{\varepsilon_w+\frac{1}{1-\xi(\varepsilon_e+1)}} = (A_1)^{\frac{\varepsilon_e+1}{(\sigma-1)[1-\xi(\varepsilon_e+1)]}}\frac{A_2}{A_3},
$$

that is

$$
W=\xi\left[(A_1)^{\frac{\varepsilon_e+1}{(\sigma-1)[1-\xi(\varepsilon_e+1)]}}\frac{A_2}{A_3}\right]^{\frac{1}{\varepsilon_w+\frac{1}{1-\xi(\varepsilon_e+1)}}}. \tag{OA15}
$$

Lastly, substituting W in (OA12) with (OA15), we have

$$Q=\left[\frac{A_3}{A_2 A_1^{\frac{\varepsilon_e+1}{(\sigma-1)[1-\xi(\varepsilon_e+1)]}}}\right]^{\frac{1}{\varepsilon_w+\frac{1}{1-\xi(\varepsilon_e+1)}}\frac{\xi(\varepsilon_e+1)}{1-\xi(\varepsilon_e+1)}} A_1^{\frac{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}{(\sigma-1)[1-\xi(\varepsilon_e+1)]}}. \tag{OA16}$$

Then we can derive $l_w(\theta_w)$, $L_w(\theta_e)$, and $l_e(\theta_e)$ by substituting Q and W into (OA6), (OA8), and (OA9). Moreover, by definition, we have

$$Q_{ij}(\theta_e)=x_e(\theta_e)\left(\frac{\xi}{W}\right)^{\frac{\xi\sigma(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}}\left[\frac{X_e(\theta_e)}{\mu(\theta_e)}\left(\frac{Q}{N_e}\right)^{\frac{1}{\sigma}}\right]^{\frac{\sigma(\varepsilon_e+1)\xi+\sigma\varepsilon_e}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}}\left[1-\tau_e(\theta_e)\right]^{\frac{\sigma\varepsilon_e}{\sigma+\varepsilon_e-\xi(\sigma-1)(\varepsilon_e+1)}}, \tag{OA17}$$

and

$$P(\theta_e)=\frac{\mu(\theta_e)}{x_e(\theta_e)}\left(\frac{\xi}{W}\right)^{\frac{-\xi(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}}\left[\frac{X_e(\theta_e)}{\mu(\theta_e)}\left(\frac{Q}{N_e}\right)^{\frac{1}{\sigma}}\right]^{\frac{\sigma-\sigma(\varepsilon_e+1)\xi}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}}\left[1-\tau_e(\theta_e)\right]^{\frac{\sigma-\xi(\sigma-1)(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}}. \tag{OA18}$$

According to the above results, we have:

$$\begin{aligned}
\frac{d\ln l_e(\theta_e)}{d\theta_e} &= \frac{\frac{d\ln X_e(\theta_e)/\mu(\theta_e)}{d\theta_e}+\frac{\sigma-\xi(\sigma-1)}{\sigma}\frac{d\ln[1-\tau_e(\theta_e)]}{d\theta_e}}{\frac{1+\varepsilon_e}{\varepsilon_e}-\frac{\sigma-1}{\sigma}\left(1+\frac{1+\varepsilon_e}{\varepsilon_e}\xi\right)},\\
\frac{d\ln L_w(\theta_e)}{d\theta_e} &= \frac{\frac{1+\varepsilon_e}{\varepsilon_e}\frac{d\ln X_e(\theta_e)/\mu(\theta_e)}{d\theta_e}+\frac{\sigma-1}{\sigma}\frac{d\ln[1-\tau_e(\theta_e)]}{d\theta_e}}{\frac{1+\varepsilon_e}{\varepsilon_e}-\frac{\sigma-1}{\sigma}\left(1+\frac{1+\varepsilon_e}{\varepsilon_e}\xi\right)},\\
\frac{d\ln Q_{ij}(\theta_e)}{d\theta_e} &= \frac{d\ln x_e(\theta_e)}{d\theta_e}+\frac{\left(\xi+\frac{\varepsilon_e}{1+\varepsilon_e}\right)\frac{d\ln X_e(\theta_e)/\mu(\theta_e)}{d\theta_e}+\frac{d\ln[1-\tau_e(\theta_e)]}{d\theta_e}}{\frac{1}{\sigma}+\frac{\sigma-1}{\sigma}\left(\frac{1}{1+\varepsilon_e}-\xi\right)},\\
\frac{d\ln\left[Q_{ij}(\theta_e)P(\theta_e)\right]}{d\theta_e} &= \frac{\frac{1-\sigma}{\sigma}\left(\xi+\frac{\varepsilon_e}{1+\varepsilon_e}\right)\frac{d\ln\mu(\theta_e)}{d\theta_e}+\frac{d\ln X_e(\theta_e)}{d\theta_e}}{\frac{1}{\sigma}+\frac{\sigma-1}{\sigma}\frac{1}{1+\varepsilon_e}-\xi}+\frac{(\sigma-1)\frac{\varepsilon_e}{1+\varepsilon_e}\frac{d\ln[1-\tau_e(\theta_e)]}{d\theta_e}}{1+(\sigma-1)\frac{1}{1+\varepsilon_e}-\xi}.
\end{aligned}$$

Moreover, we have:

$$\begin{aligned}
\frac{d\ln y_e(\theta_e)}{d\theta_e} &= \frac{d\ln\frac{\mu(\theta_e)-\xi}{\xi}WL_w(\theta_e)}{d\theta_e}=\frac{d\ln[\mu(\theta_e)-\xi]}{d\theta_e}+\frac{d\ln L_w(\theta_e)}{d\theta_e}\\
&= \frac{d\ln[\mu(\theta_e)-\xi]}{d\theta_e}+\frac{\frac{d\ln X_e(\theta_e)/\mu(\theta_e)}{d\theta_e}+\frac{\sigma-1}{\sigma}\frac{\varepsilon_e}{1+\varepsilon_e}\frac{d\ln[1-\tau_e(\theta_e)]}{d\theta_e}}{1-\frac{\sigma-1}{\sigma}\left(\frac{\varepsilon_e}{1+\varepsilon_e}+\xi\right)}
\end{aligned}$$

and

$$
\begin{aligned}
& \frac{x_e'(\theta_e)}{x_e(\theta_e)} + \mu(\theta_e) \frac{d \ln P_{ij}(Q_{ij}, Q_{-ij}(\theta_e), \theta_e)}{d\theta_e}|_{Q_{ij}=Q_{ij}(\theta_e)} \\
= \; & \frac{x_e'(\theta_e)}{x_e(\theta_e)} + \mu(\theta_e) \left\{ \frac{\chi'(\theta_e)}{\chi(\theta_e)} + \left[\frac{\sigma-1}{\sigma} - \frac{1}{\mu(\theta_e)} \right] \frac{d \ln Q_{ij}(\theta_e)}{d\theta_e} \right\} \\
= \; & \frac{x_e'(\theta_e)}{x_e(\theta_e)} + \mu(\theta_e) \left\{ \frac{\chi'(\theta_e)}{\chi(\theta_e)} + \left[\frac{\sigma-1}{\sigma} - \frac{1}{\mu(\theta_e)} \right] \left[\frac{d \ln x_e(\theta_e)}{d\theta_e} + \frac{d \ln l_e(\theta_e)}{d\theta_e} + \xi \frac{d \ln L_w(\theta_e)}{d\theta_e} \right] \right\} \\
= \; & \mu(\theta_e) \frac{X'(\theta_e)}{X(\theta_e)} + \left[\frac{\sigma-1}{\sigma} \mu(\theta_e) - 1 \right] \left[\frac{d \ln l_e(\theta_e)}{d\theta_e} + \xi \frac{d \ln L_w(\theta_e)}{d\theta_e} \right] \\
= \; & \mu(\theta_e) \frac{X'(\theta_e)}{X(\theta_e)} + \left[\frac{\sigma-1}{\sigma} \mu(\theta_e) - 1 \right] \left[\frac{\left(1 + \xi \frac{1+\varepsilon_e}{\varepsilon_e}\right) \frac{d \ln X_e(\theta_e)/\mu(\theta_e)}{d\theta_e} + \frac{d \ln[1-\tau_e(\theta_e)]}{d\theta_e}}{\frac{1+\varepsilon_e}{\varepsilon_e} - \frac{\sigma-1}{\sigma}\left(1 + \frac{1+\varepsilon_e}{\varepsilon_e}\xi\right)} \right] \\
= \; & [\mu(\theta_e) - \xi] \frac{d \ln [\mu(\theta_e) - \xi]}{d\theta_e} + \frac{\left[\begin{array}{c} \left[\mu(\theta_e) - \left(\frac{\varepsilon_e}{1+\varepsilon_e} + \xi\right)\right] \frac{d \ln X_e(\theta_e)/\mu(\theta_e)}{d\theta_e} \\ + \left[\frac{\sigma-1}{\sigma}\mu(\theta_e) - 1\right] \frac{\varepsilon_e}{1+\varepsilon_e} \frac{d \ln[1-\tau_e(\theta_e)]}{d\theta_e} \end{array} \right]}{1 - \frac{\sigma-1}{\sigma}\left(\frac{\varepsilon_e}{1+\varepsilon_e} + \xi\right)}
\end{aligned}
$$

OA.2.2 Skill Gap

To derive optimal profit tax formula in termes of parameters, we need $\frac{d \ln L_w(\theta_e)}{d\theta_e}$, $\frac{d \ln l_e(\theta_e)}{d\theta_e}$ and $\frac{d \ln y_e(\theta_e)}{d\theta_e}$ in terms of θ_e and profit tax rate. The entrepreneur's choice of labor inputs satisfies (A4), i.e., $WL_w(\theta_e) = \frac{\xi}{\mu(\theta_e)} P(\theta_e) Q_{ij}(\theta_e)$. Substitute $P(\theta_e)$ in (A4) by the inverse demand function (A6), and take the total differential on both sides of the derived equation:

$$
\frac{d \ln L_w(\theta_e)}{d\theta_e} = \frac{d \ln \chi(\theta_e)/\mu(\theta_e)}{d\theta_e} + \left[\frac{1}{\mu(\theta_e)} + \varepsilon_{Q_{-ij}}^{P,cross}(\theta_e) \right] \left[\frac{x_e'(\theta_e)}{x_e(\theta_e)} + \frac{d \ln l_e(\theta_e)}{d\theta_e} + \xi \frac{d \ln L_w(\theta_e)}{d\theta_e} \right]. \quad \text{(OA19)}
$$

Rearrange the above equation:

$$
\begin{aligned}
\frac{d \ln L_w(\theta_e)}{d\theta_e} &= \frac{\frac{d \ln \chi(\theta_e)/\mu(\theta_e)}{d\theta_e} + \left[\frac{1}{\mu(\theta_e)} + \varepsilon_{Q_{-ij}}^{P,cross}(\theta_e)\right] \frac{x_e'(\theta_e)}{x_e(\theta_e)}}{1 - \left[\frac{1}{\mu(\theta_e)} + \varepsilon_{Q_{-ij}}^{P,cross}(\theta_e)\right] \xi} \\
&\quad + \frac{\frac{1}{\mu(\theta_e)} + \varepsilon_{Q_{-ij}}^{P,cross}(\theta_e)}{1 - \left[\frac{1}{\mu(\theta_e)} + \varepsilon_{Q_{-ij}}^{P,cross}(\theta_e)\right] \xi} \frac{d \ln l_e(\theta_e)}{d\theta_e} \\
&= \frac{1}{1 - \xi \varepsilon_{Q_{ij}}^S(\theta_e)} \frac{d \ln X_e(\theta_e)/\mu(\theta_e)}{d\theta_e} + \frac{\varepsilon_{Q_{ij}}^S(\theta_e)}{1 - \xi \varepsilon_{Q_{ij}}^S(\theta_e)} \frac{l_e'(\theta_e)}{l_e(\theta_e)},
\end{aligned} \quad \text{(OA20)}
$$

where $\varepsilon_{Q_{ij}}^S(\theta_e) = \frac{1}{\mu(\theta_e)} + \varepsilon_{Q_{-ij}}^{P,cross}(\theta_e) = \frac{\sigma-1}{\sigma}$.

The entrepreneurial effort $l_e(\theta_e)$ satisfies the first-order condition (A5), i.e., $\frac{P(\theta_e)Q_{ij}(\theta_e)}{\mu(\theta_e)}[1 - \tau_e(\theta_e)] = l_e(\theta_e)^{1+\frac{1}{\varepsilon_e}}$, where $\tau_e(\theta_e) = T'(y_e(\theta_e))$. Notice that (A5) and (A4) imply:

$$
\frac{WL_w(\theta_e)}{\xi}[1 - \tau_e(\theta_e)] = l_e(\theta_e)^{1+\frac{1}{\varepsilon_e}}.
$$

We have

$$\left(1+\frac{1}{\varepsilon_e}\right)\frac{d\ln l_e\left(\theta_e\right)}{d\theta_e}=\frac{d\ln L_w\left(\theta_e\right)}{d\theta_e}+\frac{d\ln\left[1-\tau_e\left(\theta_e\right)\right]}{d\theta_e}. \tag{OA21}$$

Combination of (OA20) and (OA21) delivers:

$$\frac{d\ln l_e\left(\theta_e\right)}{d\theta_e}=\frac{\frac{d\ln X_e(\theta_e)/\mu(\theta_e)}{d\theta_e}+\left[1-\xi\frac{\sigma-1}{\sigma}\right]\frac{d\ln[1-\tau_e(\theta_e)]}{d\theta_e}}{\frac{1+\varepsilon_e}{\varepsilon_e}\left(1-\xi\frac{\sigma-1}{\sigma}\right)-\frac{\sigma-1}{\sigma}} \tag{OA22}$$

Combination of firm's first-order condition (A4) and $y_e\left(\theta_e\right)=P_{ij}\left(\theta_e\right)Q_{ij}\left(\theta_e\right)\left(1-t_s\right)-WL_{w,ij}\left(\theta_e\right)$ gives:

$$y_e\left(\theta_e\right)=\frac{\mu\left(\theta_e\right)-\xi}{\xi}WL_w(\theta_e).$$

Take the total differential on both sides of the above equation:

$$\begin{aligned}\frac{d\ln y_e\left(\theta_e\right)}{d\theta_e} &= \frac{d\ln\left[\mu\left(\theta_e\right)-\xi\right]}{d\theta_e}+\frac{d\ln L_w\left(\theta_e\right)}{d\theta_e} \\ &= \frac{d\ln\left[\mu\left(\theta_e\right)-\xi\right]}{d\theta_e}+\left(1+\frac{1}{\varepsilon_e}\right)\frac{d\ln l_e\left(\theta_e\right)}{d\theta_e}-\frac{d\ln\left[1-\tau_e\left(\theta_e\right)\right]}{d\theta_e} \\ &= \frac{\frac{1+\varepsilon_e}{\varepsilon_e}\frac{1}{\varepsilon^S_{Q_{ij}}(\theta_e)}\frac{d\ln[X_e(\theta_e)/\mu(\theta_e)]}{d\theta_e}+\frac{d\ln[1-\tau_e(\theta_e)]}{d\theta_e}}{\frac{1+\varepsilon_e}{\varepsilon_e}\left(\frac{1}{\varepsilon^S_{Q_{ij}}(\theta_e)}-\xi\right)-1}+\frac{d\ln\left[\mu\left(\theta_e\right)-\xi\right]}{d\theta_e},\end{aligned} \tag{OA23}$$

where the second and third equations are derived by (OA21) and (OA22).

Therefore,

$$\frac{d\ln y^o_e\left(\theta_e\right)}{d\theta_e}=\frac{\frac{1+\varepsilon_e}{\varepsilon_e}\frac{1}{\varepsilon^S_{Q_{ij}}(\theta_e)}\frac{d\ln[X_e(\theta_e)/\mu(\theta_e)]}{d\theta_e}}{\frac{1+\varepsilon_e}{\varepsilon_e}\left(\frac{1}{\varepsilon^S_{Q_{ij}}(\theta_e)}-\xi\right)-1}+\frac{d\ln\left[\mu\left(\theta_e\right)-\xi\right]}{d\theta_e},$$

and

$$\begin{aligned}\frac{\gamma_e'\left(\theta_e\right)}{\gamma_e\left(\theta_e\right)} &= \frac{d\ln y^o_e\left(\theta_e\right)}{d\theta_e}\frac{1}{\varepsilon^{y^o_e}_{1-\tau_e}\left(\theta_e\right)}\frac{\varepsilon_e}{1+\varepsilon_e} \\ &= \left[\frac{\frac{1+\varepsilon_e}{\varepsilon_e}\frac{1}{\varepsilon^S_{Q_{ij}}(\theta_e)}\frac{d\ln[X_e(\theta_e)/\mu(\theta_e)]}{d\theta_e}}{\frac{1+\varepsilon_e}{\varepsilon_e}\left(\frac{1}{\varepsilon^S_{Q_{ij}}(\theta_e)}-\xi\right)-1}+\frac{d\ln\left[\mu\left(\theta_e\right)-\xi\right]}{d\theta_e}\right]\frac{1}{\varepsilon^{y^o_e}_{1-\tau_e}\left(\theta_e\right)}\frac{\varepsilon_e}{1+\varepsilon_e}\end{aligned} \tag{OA24}$$

and

$$H(\theta_e)=\frac{1-F_e(\theta_e)}{f_e\left(\theta_e\right)}\left[\frac{\frac{1+\varepsilon_e}{\varepsilon_e}\frac{1}{\varepsilon^S_{Q_{ij}}(\theta_e)}\frac{d\ln[X_e(\theta_e)/\mu(\theta_e)]}{d\theta_e}}{\frac{1+\varepsilon_e}{\varepsilon_e}\left(\frac{1}{\varepsilon^S_{Q_{ij}}(\theta_e)}-\xi\right)-1}+\frac{d\ln\left[\mu\left(\theta_e\right)-\xi\right]}{d\theta_e}\right]=\frac{1-F_e(\theta_e)}{f_e\left(\theta_e\right)}\frac{\gamma_e'\left(\theta_e\right)}{\gamma_e\left(\theta_e\right)}\varepsilon^{y^o_e}_{1-\tau_e}\left(\theta_e\right)\frac{1+\varepsilon_e}{\varepsilon_e} \tag{OA25}$$

Under the Atkeson-Burstein economy, equation (OA24) is equivalent to (33).

$H(\theta_e)$ is the hazard ratio of profit ($\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{d\ln y_e(\theta_e)}{d\theta_e}$) when the profit tax rate is constant, i.e., $\tau_e'(\theta_e)=0$. Combination of (OA25) and (OA23) implies:

$$H(\theta_e)=\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\left[\frac{d\ln y_e(\theta_e)}{d\theta_e}-\frac{\frac{d\ln[1-\tau_e(\theta_e)]}{d\theta_e}}{\frac{1+\varepsilon_e}{\varepsilon_e}\left(\frac{1}{\varepsilon^S_{Q_{ij}}(\theta_e)}-\xi\right)-1}\right]. \tag{OA26}$$

We now derive a more explicit expression of $\mu(\theta_e)\frac{d\ln P_{ij}\left(Q_{ij},Q_{ij}(\theta_e),\theta_e\right)}{d\theta_e}|_{Q_{ij}=Q_{ij}(\theta_e)}+\frac{x_e'(\theta_e)}{x_e(\theta_e)}$. Remind that $\frac{d\ln P_{ij}\left(Q_{ij},Q_{ij}(\theta_e),\theta_e\right)}{d\theta_e}|_{Q_{ij}=Q_{ij}(\theta_e)}=\frac{\chi'(\theta_e)}{\chi(\theta_e)}+\varepsilon^{P,cross}_{Q_{-ij}}(\theta_e)\frac{d\ln Q_{ij}(\theta_e)}{d\theta_e}$ (see e.g., (A12)), where:

$$\begin{aligned}\frac{d\ln Q_{ij}(\theta_e)}{d\theta_e} &= \frac{x_e'(\theta_e)}{x_e(\theta_e)}+\frac{d\ln l_e(\theta_e)}{d\theta_e}+\xi\frac{d\ln L_w(\theta_e)}{d\theta_e} \\ &= \frac{x_e'(\theta_e)}{x_e(\theta_e)}+\left(1+\xi\frac{1+\varepsilon_e}{\varepsilon_e}\right)\frac{d\ln l_e(\theta_e)}{d\theta_e}-\xi\frac{d\ln[1-\tau_e(\theta_e)]}{d\theta_e} \\ &= \frac{x_e'(\theta_e)}{x_e(\theta_e)}+\frac{1}{1-\xi\varepsilon^S_{Q_{ij}}(\theta_e)}\frac{l_e'(\theta_e)}{l_e(\theta_e)}+\frac{\xi}{1-\xi\varepsilon^S_{Q_{ij}}(\theta_e)}\frac{d}{d\theta_e}\ln\frac{X_e(\theta_e)}{\mu(\theta_e)}.\end{aligned} \tag{OA27}$$

The second and third equations are derived by (OA21) and (OA20). Substitute $\frac{l_e'(\theta_e)}{l_e(\theta_e)}$ in (OA27) by (OA22):

$$\frac{d\ln Q_{ij}(\theta_e)}{d\theta_e}=\frac{x_e'(\theta_e)}{x_e(\theta_e)}+\frac{\left(1+\xi\frac{1+\varepsilon_e}{\varepsilon_e}\right)\frac{d}{d\theta_e}\ln\frac{X_e(\theta_e)}{\mu(\theta_e)}+\frac{d\ln[1-\tau_e(\theta_e)]}{d\theta_e}}{\frac{1+\varepsilon_e}{\varepsilon_e}-\varepsilon^S_{Q_{ij}}(\theta_e)\left(1+\frac{1+\varepsilon_e}{\varepsilon_e}\xi\right)}. \tag{OA28}$$

We have:

$$\begin{aligned}&\frac{d\ln P_{ij}\left(Q_{ij},Q_{ij}(\theta_e),\theta_e\right)}{d\theta_e}|_{Q_{ij}=Q_{ij}(\theta_e)}=\frac{\chi'(\theta_e)}{\chi(\theta_e)}+\varepsilon^{P,cross}_{Q_{-ij}}(\theta_e)\frac{d\ln Q_{ij}(\theta_e)}{d\theta_e} \\ =\;& \frac{\chi'(\theta_e)}{\chi(\theta_e)}+\varepsilon^{P,cross}_{Q_{-ij}}(\theta_e)\left[\frac{x_e'(\theta_e)}{x_e(\theta_e)}+\left[1+\xi\frac{1+\varepsilon_e}{\varepsilon_e}\right]\frac{l_e'(\theta_e)}{l_e(\theta_e)}-\xi\frac{d\ln[1-\tau_e(\theta_e)]}{d\theta_e}\right] \\ =\;& \frac{\chi'(\theta_e)}{\chi(\theta_e)}+\varepsilon^{P,cross}_{Q_{-ij}}(\theta_e)\left[\frac{x_e'(\theta_e)}{x_e(\theta_e)}+\frac{1}{1-\xi\varepsilon^S_{Q_{ij}}(\theta_e)}\frac{l_e'(\theta_e)}{l_e(\theta_e)}+\frac{\xi}{1-\xi\varepsilon^S_{Q_{ij}}(\theta_e)}\frac{d\ln\frac{X_e(\theta_e)}{\mu(\theta_e)}}{d\theta_e}\right].\end{aligned}$$

Substitute $\frac{d\ln Q_{ij}(\theta_e)}{d\theta_e}$ by (OA28):

$$\begin{aligned}
& \mu\left(\theta_e\right) \frac{d \ln P_{ij}\left(Q_{ij}, Q_{ij}\left(\theta_e\right), \theta_e\right)}{d\theta_e}|_{Q_{ij}=Q_{ij}(\theta_e)} + \frac{x_e'\left(\theta_e\right)}{x_e\left(\theta_e\right)} \\
= & \mu\left(\theta_e\right)\left[\frac{\chi'(\theta_e)}{\chi(\theta_e)} + \varepsilon_{Q_{-ij}}^{P,cross}\left(\theta_e\right) \frac{d\ln Q_{ij}\left(\theta_e\right)}{d\theta_e}\right] + \frac{x_e'\left(\theta_e\right)}{x_e\left(\theta_e\right)} \\
= & \mu\left(\theta_e\right) \frac{\chi'(\theta_e)}{\chi(\theta_e)} + \left[\mu\left(\theta_e\right) \varepsilon_{Q_{-ij}}^{P,cross}\left(\theta_e\right) + 1\right] \frac{x_e'\left(\theta_e\right)}{x_e\left(\theta_e\right)} + \\
& \mu\left(\theta_e\right) \varepsilon_{Q_{-ij}}^{P,cross}\left(\theta_e\right) \frac{\left(1+\xi\frac{1+\varepsilon_e}{\varepsilon_e}\right)\frac{d}{d\theta_e}\ln\frac{X_e(\theta_e)}{\mu(\theta_e)} + \frac{d\ln[1-\tau_e(\theta_e)]}{d\theta_e}}{\frac{1+\varepsilon_e}{\varepsilon_e} - \varepsilon_{Q_{ij}}^{S}\left(\theta_e\right)\left(1+\frac{1+\varepsilon_e}{\varepsilon_e}\xi\right)},
\end{aligned}$$

where

$$\begin{aligned}
& \mu\left(\theta_e\right) \frac{\chi'(\theta_e)}{\chi(\theta_e)} + \left[\mu\left(\theta_e\right) \varepsilon_{Q_{-ij}}^{P,cross}\left(\theta_e\right) + 1\right] \frac{x_e'\left(\theta_e\right)}{x_e\left(\theta_e\right)} \\
= & \mu\left(\theta_e\right) \frac{\chi'(\theta_e)}{\chi(\theta_e)} + \mu\left(\theta_e\right) \varepsilon_{Q_{ij}}^{S}\left(\theta_e\right) \frac{x_e'\left(\theta_e\right)}{x_e\left(\theta_e\right)} \\
= & \mu\left(\theta_e\right) \frac{d}{d\theta_e}\left[\ln\frac{X_e(\theta_e)}{\mu\left(\theta_e\right)}\right] + \mu\left(\theta_e\right) \frac{d\ln\mu\left(\theta_e\right)}{d\theta_e} \\
= & \frac{\mu\left(\theta_e\right) \varepsilon_{Q_{-ij}}^{P,cross}\left(\theta_e\right)+1}{\varepsilon_{Q_{ij}}^{S}\left(\theta_e\right)} \frac{d}{d\theta_e}\left[\ln\frac{X_e(\theta_e)}{\mu\left(\theta_e\right)}\right] + \mu\left(\theta_e\right) \frac{d\ln\mu\left(\theta_e\right)}{d\theta_e}.
\end{aligned}$$

The first and third equations of the above equations are derived by $\varepsilon_{Q_{-ij}}^{P,cross}\left(\theta_e\right) = -\frac{1}{\mu(\theta_e)} + \frac{\sigma-1}{\sigma}$, and the second equation is derived by $X_e(\theta_e) = A^{\frac{\sigma-1}{\sigma}}\chi\left(\theta_e\right) x_e\left(\theta_e\right)^{\frac{\sigma-1}{\sigma}}$. In conclusion,

$$\begin{aligned}
& \mu\left(\theta_e\right) \frac{d \ln P_{ij}\left(Q_{ij}, Q_{ij}\left(\theta_e\right), \theta_e\right)}{d\theta_e}|_{Q_{ij}=Q_{ij}(\theta_e)} + \frac{x_e'\left(\theta_e\right)}{x_e\left(\theta_e\right)} \\
= & \frac{\mu\left(\theta_e\right) \varepsilon_{Q_{-ij}}^{P,cross}\left(\theta_e\right)+1}{\varepsilon_{Q_{ij}}^{S}\left(\theta_e\right)} \frac{d}{d\theta_e}\left[\ln\frac{X_e(\theta_e)}{\mu\left(\theta_e\right)}\right] + \mu\left(\theta_e\right) \frac{d\ln\mu\left(\theta_e\right)}{d\theta_e} \\
& +\mu\left(\theta_e\right) \varepsilon_{Q_{-ij}}^{P,cross}\left(\theta_e\right) \frac{\left(1+\xi\frac{1+\varepsilon_e}{\varepsilon_e}\right)\frac{d}{d\theta_e}\left[\ln\frac{X_e(\theta_e)}{\mu(\theta_e)}\right] + \frac{d\ln[1-\tau_e(\theta_e)]}{d\theta_e}}{\frac{1+\varepsilon_e}{\varepsilon_e} - \varepsilon_{Q_{ij}}^{S}\left(\theta_e\right)\left(1+\frac{1+\varepsilon_e}{\varepsilon_e}\xi\right)} \\
= & \mu\left(\theta_e\right) \frac{d\ln X_e(\theta_e)}{d\theta_e} + \left[\mu\left(\theta_e\right) \varepsilon_{Q_{ij}}^{S}\left(\theta_e\right) - 1\right] \frac{\left(1+\xi\frac{1+\varepsilon_e}{\varepsilon_e}\right)\frac{d}{d\theta_e}\ln\frac{X_e(\theta_e)}{\mu(\theta_e)} + \frac{d\ln[1-\tau_e(\theta_e)]}{d\theta_e}}{\frac{1+\varepsilon_e}{\varepsilon_e} - \varepsilon_{Q_{ij}}^{S}\left(\theta_e\right)\left(1+\frac{1+\varepsilon_e}{\varepsilon_e}\xi\right)}.
\end{aligned} \tag{OA29}$$

When $\tau_e'\left(\theta_e\right) = 0$,

$$\begin{aligned}
& \mu\left(\theta_e\right) \frac{d \ln P_{ij}\left(Q_{ij}, Q_{ij}\left(\theta_e\right), \theta_e\right)}{d\theta_e}|_{Q_{ij}=Q_{ij}(\theta_e)} + \frac{x_e'\left(\theta_e\right)}{x_e\left(\theta_e\right)} \\
= & \mu\left(\theta_e\right) \frac{d\ln X_e(\theta_e)}{d\theta_e} + \left[\mu\left(\theta_e\right) \varepsilon_{Q_{ij}}^{S}\left(\theta_e\right) - 1\right] \frac{\left(1+\xi\frac{1+\varepsilon_e}{\varepsilon_e}\right)\frac{d}{d\theta_e}\left[\ln\frac{X_e(\theta_e)}{\mu(\theta_e)}\right]}{\frac{1+\varepsilon_e}{\varepsilon_e} - \varepsilon_{Q_{ij}}^{S}\left(\theta_e\right)\left(1+\frac{1+\varepsilon_e}{\varepsilon_e}\xi\right)}.
\end{aligned}$$

By (A6), we have:

$$\frac{d\ln P_{ij}(Q_{ij},Q_{ij}\left(\theta_{e}\right),\theta_{e})}{d\theta_{e}}|_{Q_{ij}=Q_{ij}(\theta_{e})}=\frac{\chi'(\theta_{e})}{\chi(\theta_{e})}+\left[\frac{\sigma-1}{\sigma}-\frac{1}{\mu\left(\theta_{e}\right)}\right]\frac{d\ln Q_{ij}\left(\theta_{e}\right)}{d\theta_{e}},$$

where, by definition,

$$\begin{aligned}\frac{d\ln Q_{ij}\left(\theta_{e}\right)}{d\theta_{e}} &= \frac{d\ln x_{e}(\theta_{e})}{d\theta_{e}}+\frac{d\ln l_{e}(\theta_{e})}{d\theta_{e}}+\xi\frac{d\ln L_{w}(\theta_{e})}{d\theta_{e}}\\ &= \frac{\sigma}{\sigma-1}\left[\frac{d\ln\mu\left(\theta_{e}\right)}{d\theta_{e}}-\frac{d\ln x_{e}(\theta_{e})}{d\theta_{e}}+\frac{\sigma\left(\varepsilon_{e}+1\right)\left[\frac{d\ln X_{e}(\theta_{e})}{d\theta_{e}}-\frac{d\ln\mu(\theta_{e})}{d\theta_{e}}\right]}{\varepsilon_{e}+\sigma-\xi\left(\sigma-1\right)\left(\varepsilon_{e}+1\right)}\right]\\ &\quad+\frac{\sigma\varepsilon_{e}\frac{d\ln[1-\tau_{e}(\theta_{e})]}{d\theta_{e}}}{\sigma+\varepsilon_{e}-\xi\left(\sigma-1\right)\left(\varepsilon_{e}+1\right)}.\end{aligned}$$

Thus,

$$\begin{aligned}&\frac{x_{e}'(\theta_{e})}{x_{e}(\theta_{e})}+\mu\left(\theta_{e}\right)\frac{d\ln P_{ij}(Q_{ij},Q_{-ij}\left(\theta_{e}\right),\theta_{e})}{d\theta_{e}}|_{Q_{ij}=Q_{ij}(\theta_{e})}\\ =&\frac{\sigma\left(\varepsilon_{e}+1\right)\left[\mu\left(\theta_{e}\right)-\xi\right]-\sigma\varepsilon_{e}}{\varepsilon_{e}+\sigma-\xi\left(\sigma-1\right)\left(\varepsilon_{e}+1\right)}\frac{d\ln X(\theta_{e})}{d\theta_{e}}\\ &-\frac{\left(\sigma-1\right)\left[\varepsilon_{e}+\xi\left(\varepsilon_{e}+1\right)\right]\left[\mu\left(\theta_{e}\right)-\frac{\sigma}{\sigma-1}\right]}{\varepsilon_{e}+\sigma-\xi\left(\sigma-1\right)\left(\varepsilon_{e}+1\right)}\frac{d\ln\mu\left(\theta_{e}\right)}{d\theta_{e}}\\ &+\frac{\varepsilon_{e}\left[\left(\sigma-1\right)\mu\left(\theta_{e}\right)-\sigma\right]}{\sigma+\varepsilon_{e}-\xi\left(\sigma-1\right)\left(\varepsilon_{e}+1\right)}\frac{d\ln\left[1-\tau_{e}\left(\theta_{e}\right)\right]}{d\theta_{e}}.\end{aligned} \tag{OA30}$$

Substitute $\left[\mu\left(\theta_{e}\right)\frac{\partial\ln P(Q_{ij}(\theta_{e}),\theta_{e})}{\partial\theta_{e}}+\frac{x_{e}'(\theta_{e})}{x_{e}(\theta_{e})}\right]$ in $V_{e}'\left(\theta_{e}\right)=l_{e}(\theta_{e})^{\frac{1+\varepsilon_{e}}{\varepsilon_{e}}}\left[\mu\left(\theta_{e}\right)\frac{\partial\ln P(Q_{ij}(\theta_{e}),\theta_{e})}{\partial\theta_{e}}+\frac{x_{e}'(\theta_{e})}{x_{e}(\theta_{e})}\right]$:

$$V'(\theta_{e})=l_{e}(\theta_{e})^{\frac{1+\varepsilon_{e}}{\varepsilon_{e}}}\left[\begin{array}{c}\frac{\sigma(\varepsilon_{e}+1)\mu(\theta_{e})-\sigma\varepsilon_{e}-\xi\sigma(\varepsilon_{e}+1)}{\varepsilon_{e}+\sigma-\xi(\sigma-1)(\varepsilon_{e}+1)}\frac{d\ln X(\theta_{e})}{d\theta_{e}}\\ -\frac{(\sigma-1)[\varepsilon_{e}+\xi(\varepsilon_{e}+1)]\left[\mu(\theta_{e})-\frac{\sigma}{\sigma-1}\right]}{\varepsilon_{e}+\sigma-\xi(\sigma-1)(\varepsilon_{e}+1)}\frac{d\ln\mu(\theta_{e})}{d\theta_{e}}\\ +\frac{\varepsilon_{e}[(\sigma-1)\mu(\theta_{e})-\sigma]}{\sigma+\varepsilon_{e}-\xi(\sigma-1)(\varepsilon_{e}+1)}\frac{d\ln[1-\tau_{e}(\theta_{e})]}{d\theta_{e}}\end{array}\right],\forall\theta_{e}\in\Theta_{e}.$$

OA.2.3 The Laissez-faire Economy

Gross Utility. Note that $\frac{1+\varepsilon_{e}}{\varepsilon_{e}}-\frac{\sigma-1}{\sigma}\left(1+\frac{1+\varepsilon_{e}}{\varepsilon_{e}}\xi\right)$, $\varepsilon_{e}+\sigma-\xi\left(\sigma-1\right)\left(\varepsilon_{e}+1\right)$ and $\frac{\sigma}{\sigma-1}-\left(\frac{\varepsilon_{e}}{1+\varepsilon_{e}}+\xi\right)$ are positive under condition (24). Under such a condition, whether $l_{e}\left(\theta_{e}\right)$, $L_{w}\left(\theta_{e}\right)$, $Q_{ij}\left(\theta_{e}\right)$, $Q_{ij}\left(\theta_{e}\right)P(\theta_{e})$ and $V(\theta_{e})$ increases with θ_{e} is determined by the relative change of $X_{e}(\theta_{e})$ to $\mu\left(\theta_{e}\right)$. In particular, we have

$$\begin{aligned}\frac{d\ln V(\theta_{e})}{d\theta_{e}} &= \left[\begin{array}{c}\frac{\mu(\theta_{e})}{\mu(\theta_{e})-\left(\xi+\frac{\varepsilon_{e}}{1+\varepsilon_{e}}\right)}\\ -\frac{\frac{\sigma}{\sigma-1}}{\frac{\sigma}{\sigma-1}-\left(\xi+\frac{\varepsilon_{e}}{1+\varepsilon_{e}}\right)}\end{array}\right]\frac{\mu'\left(\theta_{e}\right)}{\mu\left(\theta_{e}\right)}+\frac{\frac{\sigma}{\sigma-1}\frac{d\ln X_{e}(\theta_{e})}{d\theta_{e}}}{\frac{\sigma}{\sigma-1}-\left(\frac{\varepsilon_{e}}{1+\varepsilon_{e}}+\xi\right)}\\ &\geq \frac{\frac{\sigma}{\sigma-1}\frac{d\ln X_{e}(\theta_{e})}{d\theta_{e}}}{\frac{\sigma}{\sigma-1}-\left(\frac{\varepsilon_{e}}{1+\varepsilon_{e}}+\xi\right)},\end{aligned}$$

where the second inequality is derived by $\mu(\theta_e) \leq \frac{\sigma}{\sigma-1}$. Thus, $\frac{d\ln V(\theta_e)}{d\theta_e}$ increases with $\frac{d\ln X_e(\theta_e)}{d\theta_e}$ and introducing market power inequality rises $\frac{d\ln V(\theta_e)}{d\theta_e}$.

We also have:

$$V_e'(\theta_e) = l_e(\theta_e)\,\phi'(l_e(\theta_e))\left[\mu(\theta_e)\frac{\partial \ln P(Q_{ij}(\theta_e),\theta_e)}{\partial\theta_e} + \frac{x_e'(\theta_e)}{x_e(\theta_e)}\right], \tag{OA31}$$

with

$$\begin{aligned}
& \mu(\theta_e)\frac{d\ln P_{ij}(Q_{ij},Q_{ij}(\theta_e),\theta_e)}{d\theta_e}\Big|_{Q_{ij}=Q_{ij}(\theta_e)} + \frac{x_e'(\theta_e)}{x_e(\theta_e)} \\
= \; & \mu(\theta_e)\left\{\frac{\chi'(\theta_e)}{\chi(\theta_e)} + \left[\frac{\sigma-1}{\sigma} - \frac{1}{\mu(\theta_e)}\right]\frac{d\ln Q_{ij}(\theta_e)}{d\theta_e}\right\} + \frac{x_e'(\theta_e)}{x_e(\theta_e)} \\
= \; & \frac{\sigma(\varepsilon_e+1)\left[\mu(\theta_e)-\xi\right]-\sigma\varepsilon_e}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}\frac{d\ln X_e(\theta_e)}{d\theta_e} \\
& + \frac{(\sigma-1)\left[\varepsilon_e+\xi(\varepsilon_e+1)\right]\left[\frac{\sigma}{\sigma-1}-\mu(\theta_e)\right]}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}\frac{d\ln\mu(\theta_e)}{d\theta_e}.
\end{aligned} \tag{OA32}$$

Notice that $\varepsilon_e + \sigma - \xi(\sigma-1)(\varepsilon_e+1)$ is positive under condition (24). The entrepreneurial skill gap increases with $\frac{d\ln X_e(\theta_e)}{d\theta_e}$. Moreover, since $\mu(\theta_e) \leq \frac{\sigma}{\sigma-1}$, the second term on the right side of (OA32) is positive under condition (24), which suggests that the θ_e-type entrepreneur's skill gap increases in $\frac{d\ln\mu(\theta_e)}{d\theta_e}$.

Technology and Equilibrium. $x_e(\cdot)$ and $\chi(\cdot)$ have different economic meanings. They can refer to quantity-augmenting and quality-augmenting (Rosen (1981)), ability and talent (Sattinger (1975b)), and effort-augmenting and total-productivity-augmenting (non-effort-augmenting) elements (Ales et al. (2017)), all of which catch the difference between an entrepreneur and a worker.

The expressions for allocations and prices in Appendix A.1 show that $Q_{ij}(\theta_e)$ and $P(\theta_e)$ are generally dependent on the specific values of $x_e(\theta_e)$ and $\chi(\theta_e)$ instead of only depending on the value of $X_e(\theta_e) = A^{\frac{\sigma-1}{\sigma}}x_e(\theta_e)^{\frac{\sigma-1}{\sigma}}\chi(\theta_e)$.

Consider the economy without taxes. According to (OA17) and (OA18), we have:

$$Q_{ij}(\theta_e) = x_e(\theta_e)\left(\frac{\xi}{W}\right)^{\frac{\xi\sigma(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}}\left[\frac{X_e(\theta_e)}{\mu(\theta_e)}\left(\frac{Q}{N_e}\right)^{\frac{1}{\sigma}}\right]^{\frac{\sigma(\varepsilon_e+1)\xi+\sigma\varepsilon_e}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}},$$

and

$$P(\theta_e) = \frac{\mu(\theta_e)}{x_e(\theta_e)}\left(\frac{\xi}{W}\right)^{\frac{-\xi(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}}\left[\frac{X_e(\theta_e)}{\mu(\theta_e)}\left(\frac{Q}{N_e}\right)^{\frac{1}{\sigma}}\right]^{\frac{\sigma-\sigma(\varepsilon_e+1)\xi}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}}, \forall\theta_e\in\Theta.$$

On the other hand, given $X_e(\theta_e)$, we see that $P(\theta_e)Q_{ij}(\theta_e)$, $L_w(\theta_e)$, $l_e(\theta_e)$, and $V_e(\theta_e)$ are independent of the specific values of $\chi(\theta_e)$ and $x_e(\theta_e)$. According to (OA6) to (OA8), we have the following results for any

$\theta_e \in \Theta_e$:

$$
\begin{aligned}
L_w(\theta_e) &= \left(\frac{\xi}{W}\right)^{\frac{\sigma+\varepsilon_e}{\sigma+\varepsilon_e-\xi(\sigma-1)(\varepsilon_e+1)}} \left[\frac{X_e(\theta_e)}{\mu\left(\theta_e\right)}\left(\frac{Q}{N_e}\right)^{\frac{1}{\sigma}}\right]^{\frac{\sigma(\varepsilon_e+1)}{\sigma+\varepsilon_e-\xi(\sigma-1)(\varepsilon_e+1)}}, \\
P(\theta_e)Q_{ij}(\theta_e) &= \mu\left(\theta_e\right)\left(\frac{\xi}{W}\right)^{\frac{\xi(\sigma-1)(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}} \left[\frac{X_e(\theta_e)}{\mu\left(\theta_e\right)}\left(\frac{Q}{N_e}\right)^{\frac{1}{\sigma}}\right]^{\frac{\sigma(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}}, \\
l_e\left(\theta_e\right) &= \left(\frac{\xi}{W}\right)^{\frac{\xi(\sigma-1)\varepsilon_e}{\sigma+\varepsilon_e-\xi(\sigma-1)(\varepsilon_e+1)}} \left[\frac{X_e(\theta_e)}{\mu\left(\theta_e\right)}\left(\frac{Q}{N_e}\right)^{\frac{1}{\sigma}}\right]^{\frac{\sigma\varepsilon_e}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}},
\end{aligned}
$$

In addition, by $c_e(\theta_e) = P(\theta_e)Q_{ij}(\theta_e) - WL_{w,ij}(\theta_e)$ and $V(\theta_e) = c_e(\theta_e) - \phi\left(l_e(\theta_e)\right)$, we have:

$$
\begin{aligned}
V(\theta_e) &= \left[\mu\left(\theta_e\right) - \xi - \frac{\varepsilon_e}{1+\varepsilon_e}\right]\phi'\left(l_e(\theta_e)\right)l_e(\theta_e) \qquad \text{(OA33)} \\
&= \left[\mu\left(\theta_e\right) - \xi - \frac{\varepsilon_e}{1+\varepsilon_e}\right]\left(\frac{\xi}{W}\right)^{\frac{\xi(\sigma-1)(\varepsilon_e+1)}{\sigma+\varepsilon_e-\xi(\sigma-1)(\varepsilon_e+1)}} \left[\frac{X_e(\theta_e)}{\mu\left(\theta_e\right)}\left(\frac{Q}{N_e}\right)^{\frac{1}{\sigma}}\right]^{\frac{\sigma(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}}
\end{aligned}
$$

and

$$
c_e(\theta_e) = \left[\mu\left(\theta_e\right) - \xi\right]\left(\frac{\xi}{W}\right)^{\frac{\xi(\sigma-1)(\varepsilon_e+1)}{\sigma+\varepsilon_e-\xi(\sigma-1)(\varepsilon_e+1)}} \left[\frac{X_e(\theta_e)}{\mu\left(\theta_e\right)}\left(\frac{Q}{N_e}\right)^{\frac{1}{\sigma}}\right]^{\frac{\sigma(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}}. \qquad \text{(OA34)}
$$

Similarly, one can see that W, $l_w\left(\theta_w\right)$, and $V_w\left(\theta_w\right)$ are also only dependent on $X_e(\theta_e)$.

Lastly, we find that given $\frac{d\ln X_e(\theta_e)}{d\theta_e}$, $\frac{V'(\theta_e)}{V(\theta_e)}$ is independent of the specific values of $\chi(\theta_e)$ and $x_e(\theta_e)$. Combining (OA31) and (OA32) gives:

$$
V'(\theta_e) = \left[l_e(\theta_e)\right]^{\frac{1+\varepsilon_e}{\varepsilon_e}}\left[\begin{array}{c} \frac{\sigma(\varepsilon_e+1)\mu(\theta_e)-\sigma\varepsilon_e-\xi\sigma(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}\frac{d\ln X_e(\theta_e)}{d\theta_e} \\ -\frac{(\sigma-1)[\varepsilon_e+\xi(\varepsilon_e+1)]\left[\mu(\theta_e)-\frac{\sigma}{\sigma-1}\right]}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}\frac{d\ln\mu(\theta_e)}{d\theta_e}\end{array}\right], \forall\theta_e \in \Theta_e.
$$

Combining (OA33) and (OA31) gives

$$
\begin{aligned}
\frac{V'(\theta_e)}{V(\theta_e)} &= \frac{\mu\left(\theta_e\right)\frac{d\ln P_{ij}(Q_{ij},Q_{ij}(\theta_e),\theta_e)}{d\theta_e}|_{Q_{ij}=Q_{ij}(\theta_e)} + \frac{x_e'(\theta_e)}{x_e(\theta_e)}}{\mu\left(\theta_e\right) - \xi - \frac{\varepsilon_e}{1+\varepsilon_e}} \qquad \text{(OA35)} \\
&= \frac{\frac{\sigma(\varepsilon_e+1)[\mu(\theta_e)-\xi]-\sigma\varepsilon_e}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}\frac{d\ln X_e(\theta_e)}{d\theta_e} + \frac{(\sigma-1)[\varepsilon_e+\xi(\varepsilon_e+1)]\left[\frac{\sigma}{\sigma-1}-\mu(\theta_e)\right]}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}\frac{d\ln\mu(\theta_e)}{d\theta_e}}{\mu\left(\theta_e\right) - \xi - \frac{\varepsilon_e}{1+\varepsilon_e}},
\end{aligned}
$$

where the second equation is derived by (OA32) and (OA31), $\theta_e \in \Theta_e$. Specially, when markup is constant, we have

$$
\frac{V'(\theta_e)}{V(\theta_e)} = \frac{\sigma\left(\varepsilon_e+1\right)}{\varepsilon_e+\sigma-\xi\left(\sigma-1\right)\left(\varepsilon_e+1\right)}\frac{d\ln X_e(\theta_e)}{d\theta_e}, \forall\theta_e \in \Theta_e.
$$

OA.3 Proof of Proposition 1

Part 1 of Proposition 1 can be derived by (16) and (20). We now prove part 2. By (OA16), (OA15), and (OA14), we have

$$\nu(I) \triangleq \frac{WL}{\xi Q} = \frac{A_2}{A_1}, \tag{OA36}$$

where L is the aggregate labor input. Substituting A_1 and A_2 by (OA10), we have

$$\nu(I) = \frac{\int f_e(\theta_e)\left[\frac{x_e(\theta_e)^{\frac{\sigma-1}{\sigma}}\chi(\theta_e)}{\mu(\theta_e)}\right]^{\frac{\sigma(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}}d\theta_e}{\int_{\theta_e} f_e\left(\theta_e\right)\mu\left(\theta_e\right)\left[\frac{x_e(\theta_e)^{\frac{\sigma-1}{\sigma}}\chi(\theta_e)}{\mu(\theta_e)}\right]^{\frac{\sigma(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}}d\theta_e}. \tag{OA37}$$

For the convenience of analysis, define

$$\begin{aligned}
m\left(\theta_e\right) &\equiv f_e\left(\theta_e\right)\left[x_e(\theta_e)^{\frac{\sigma-1}{\sigma}}\chi(\theta_e)\right]^{\frac{\sigma(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}}, \\
g(\theta_e, I) &\equiv \left[\frac{1}{\mu\left(\theta_e\right)}\right]^{(\sigma-1)\frac{\varepsilon_e+\xi(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}}, \\
f_s(\theta_e, I) &\equiv \frac{g(\theta_e, I)m\left(\theta_e\right)}{\int_{\theta_e} g(\theta_e, I)m\left(\theta_e\right)d\theta_e}, \forall\theta_e \in \Theta_e.
\end{aligned}$$

Then we have:

$$\nu\left(I\right) = \int_{\theta_e}\frac{f_s(\theta_e, I)}{\mu\left(\theta_e\right)}d\theta_e. \tag{OA38}$$

In addition,

$$\begin{aligned}
\frac{\frac{d\nu(I)}{d\ln I}}{\frac{(\sigma-1)(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}} &= \int_{\theta_e} f_s(\theta_e, I)\left[\left[\left(1+\frac{1}{\sigma-1}\right)\frac{1}{\mu\left(\theta_e\right)} - \nu\left(I\right)\left(1-\frac{1}{1+\varepsilon_e}+\xi\right)\right]\frac{d\ln\left[\frac{1}{\mu(\theta_e)}\right]}{d\ln I}\right]d\theta \\
&= \left(\frac{1}{\sigma-1}+\frac{1}{1+\varepsilon_e}-\xi\right)\int_{\theta_e} f_s(\theta_e, I)\frac{1}{\mu\left(\theta_e\right)}\frac{d\ln\left[\frac{1}{\mu(\theta_e)}\right]}{d\ln I} \\
&\quad + \left(\frac{\varepsilon_e}{1+\varepsilon_e}+\xi\right)\int_{\theta_e} f_s(\theta_e, I)\left[\frac{1}{\mu\left(\theta_e\right)}-\nu\left(I\right)\right]\frac{d\ln\left[\frac{1}{\mu(\theta_e)}\right]}{d\ln I}d\theta.
\end{aligned}$$

Since $\varepsilon_e+\sigma-\xi\left(\sigma-1\right)\left(\varepsilon_e+1\right) > 0$ the sign of $\frac{d\nu(I)}{d\ln I}$ is same with the right equations, and $\left(\frac{1}{\sigma-1}+\frac{1}{1+\varepsilon_e}-\xi\right) > 0$ and $\frac{d\ln\left[\frac{1}{\mu(\theta_e)}\right]}{d\ln I} > 0$, we have:

$$\left(\frac{1}{\sigma-1}+\frac{1}{1+\varepsilon_e}-\xi\right)\int_{\theta_e} f_s(\theta_e, I)\frac{1}{\mu\left(\theta_e\right)}\frac{d\ln\left[\frac{1}{\mu(\theta_e)}\right]}{d\ln I}d\theta_e > 0.$$

Notice that $\frac{d\ln\left[\frac{1}{\mu(\theta_e)}\right]}{d\ln I} = \left[1 - \frac{\sigma-1}{\sigma}\mu\left(\theta_e\right)\right]\frac{I}{I-1}$ decrease in $\mu\left(\theta_e\right)$. We now try to prove that:

$$\int_{\theta_e} f_s(\theta_e, I)\left[\frac{1}{\mu\left(\theta_e\right)} - \nu\left(I\right)\right]\frac{d\ln\left[\frac{1}{\mu(\theta_e)}\right]}{d\ln I}d\theta_e \geq 0.$$

To do this, note that by (OA38), we use

$$\int_{\theta_e} f_s(\theta_e, I)\left[\frac{1}{\mu\left(\theta_e\right)} - \nu\left(I\right)\right]d\theta_e = 0,$$

where $f_s(\theta_e, I)\left[\frac{1}{\mu(\theta_e)} - \nu\left(I\right)\right]$ is positive if and only if $\frac{1}{\mu(\theta_e)} - \nu\left(I\right)$ is positive. Set $\Omega \equiv \left\{\theta_e | \mu\left(\theta_e\right) < \frac{1}{\nu(I)}\right\}$. $f_s(\theta_e, I)\left[\frac{1}{\mu(\theta_e)} - \nu\left(I\right)\right] > 0$ if and only if $\theta_e \in \Omega$.

Notice that:

$$\int_{\theta_e \in \Omega_*} f_s(\theta_e, I)\left[\frac{1}{\mu\left(\theta_e\right)} - \nu\left(I\right)\right]d\theta_e + \int_{\theta_e \notin \Omega_*} f_s(\theta_e, I)\left[\frac{1}{\mu\left(\theta_e\right)} - \nu\left(I\right)\right]d\theta_e = 0,$$

and

$$\int_{\theta_e \in \Omega_*} f_s(\theta_e, I)\left[\frac{1}{\mu\left(\theta_e\right)} - \nu\left(I\right)\right]d\theta_e > 0$$

and $\frac{d\ln\left[\frac{1}{\mu(\theta_e)}\right]}{d\ln I} < 0$. One can see that for any $\theta_e \in \Omega$,

$$\frac{d\ln\left[\frac{1}{\mu(\theta_e)}\right]}{d\ln I} \geq \frac{d\ln\left[\frac{1}{\mu(\theta_e)}\right]}{d\ln I}\,|_{\mu(\theta_e)=\nu(I)} = \left[1 - \frac{\sigma-1}{\sigma}\nu\left(I\right)\right]\frac{I}{I-1}.$$

Therefore, we have:

$$\begin{aligned} & \int_{\theta_e \in \Omega_*} f_s(\theta_e, I)\left[\frac{1}{\mu\left(\theta_e\right)} - \nu\left(I\right)\right]\frac{d\ln\left[\frac{1}{\mu(\theta_e)}\right]}{d\ln I}d\theta_e \\ \geq \quad & \left[1 - \frac{\sigma-1}{\sigma}\nu\left(I\right)\right]\frac{I}{I-1}\int_{\theta_e \in \Omega_*} f_s(\theta_e, I)\left[\frac{1}{\mu\left(\theta_e\right)} - \nu\left(I\right)\right]d\theta_e. \end{aligned} \tag{OA39}$$

On the other hand, for any $\theta_e \notin \Omega$:

$$\frac{d\ln\left[\frac{1}{\mu(\theta_e)}\right]}{d\ln I} \leq \left[1 - \frac{\sigma-1}{\sigma}\nu\left(I\right)\right]\frac{I}{I-1}, f_s(\theta_e, I)\left[\frac{1}{\mu\left(\theta_e\right)} - \nu\left(I\right)\right] \leq 0.$$

Therefore, we have:

$$\begin{aligned} & \int_{\theta_e \notin \Omega_*} f_s(\theta_e, I)\left[\frac{1}{\mu\left(\theta_e\right)} - \nu\left(I\right)\right]\frac{d\ln\left[\frac{1}{\mu(\theta_e)}\right]}{d\ln I}d\theta_e \\ \geq \quad & \left[1 - \frac{\sigma-1}{\sigma}\nu\left(I\right)\right]\frac{I}{I-1}\int_{\theta_e \notin \Omega_*} f_s(\theta_e, I)\left[\frac{1}{\mu\left(\theta_e\right)} - \nu\left(I\right)\right]d\theta_e. \end{aligned} \tag{OA40}$$

Combination of (OA39) and (OA40) gives:

$$\int_{\theta_e} f_s(\theta_e, I) \left[\frac{1}{\mu(\theta_e)} - \nu(I)\right] \frac{d \ln\left[\frac{1}{\mu(\theta_e)}\right]}{d \ln I} d\theta_e \geq 0,$$

which suggests $\frac{d\nu(I)}{d \ln I} \geq 0$.■

OA.4 Proof of Proposition 2

Proposition 8 •

(i) At the individual level, the labor share $\nu_{ij}(\theta_e)$, the quantity $Q_{ij}(\theta_e)$, sales $P_{ij}(\theta_e)Q_{ij}(\theta_e)$, entrepreneurial effort $l_{e,ij}(\theta_e)$, worker effort $l_w(\theta_w)$, income $y_w(\theta_w)$ and utility $V_w(\theta_w)$ decrease; The price $P_{ij}(\theta_e)$ remains unchanged; The effects on entrepreneur utility $V_{ij,e}(\theta_e)$ and entrepreneur profits $y_{e,ij}(\theta_e)$ are ambiguous;

(ii) At the aggregate level, the wage rate W, the aggregate labor share ν and output Q decrease. The effects on aggregate entrepreneur profits is ambiguous.

(iii) Individual and aggregate entrepreneur profits increase if and only if $\mu \leq \frac{\xi}{\frac{\varepsilon_e}{1+\varepsilon_e} + \frac{\varepsilon_w}{\varepsilon_w+1}\xi}$, and individual and aggregate entrepreneur utility increase if and only if $\mu \leq \frac{\xi + \frac{\varepsilon_e}{1+\varepsilon_e}}{\frac{\varepsilon_e}{1+\varepsilon_e} + \frac{\varepsilon_w}{1+\varepsilon_w}\xi}$.

The proof for part one and two are below. Remember that markups are constant. Equation (OA15) and (OA16) give:

$$\frac{W}{\xi} = \left[(A_1)^{\frac{\varepsilon_e+1}{(\sigma-1)[1-\xi(\varepsilon_e+1)]}} \frac{A_2}{A_3}\right]^{\frac{1-\xi(\varepsilon_e+1)}{\varepsilon_w[1-\xi(\varepsilon_e+1)]+1}} \propto \left(\frac{1}{\mu}\right)^{\frac{(\varepsilon_e+1)}{\varepsilon_w[1-\xi(\varepsilon_e+1)]+1}}$$

and

$$Q \propto \left(\frac{1}{\mu}\right)^{-\frac{\xi(\varepsilon_e+1)}{\varepsilon_w[1-\xi(\varepsilon_e+1)]+1}\frac{(\varepsilon_e+1)}{1-\xi(\varepsilon_e+1)}} \mu \left(\frac{1}{\mu}\right)^{\frac{(\varepsilon_e+1)}{1-\xi(\varepsilon_e+1)}} = \left(\frac{1}{\mu}\right)^{\frac{(\varepsilon_w+1)\varepsilon_e+\varepsilon_w\xi(\varepsilon_e+1)}{\varepsilon_w[1-\xi(\varepsilon_e+1)]+1}}.$$

Substituting W and Q in (OA6) with the above equations, we have

$$L_w(\theta_e) \propto \left[\frac{1}{\mu}\right]^{\frac{(\sigma-1)(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}} \left[\frac{1}{\mu}\right]^{\frac{\varepsilon_w(\varepsilon_e+1)-\sigma+1}{\varepsilon_w[1-\xi(\varepsilon_e+1)]+1}\frac{\varepsilon_e+1}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}} = \left[\frac{1}{\mu}\right]^{\frac{(\varepsilon_e+1)\varepsilon_w}{\varepsilon_w[1-\xi(\varepsilon_e+1)]+1}}, \forall\theta_e \in \Theta_e.$$

Similarly, we have:

$$
\begin{aligned}
S(\theta_e) &\propto \left[\frac{1}{\mu}\right]^{\frac{\varepsilon_e(\varepsilon_w+1)+\varepsilon_w\xi(\varepsilon_e+1)}{\varepsilon_w[1-\xi(\varepsilon_e+1)]+1}}, \\
l_e(\theta_e) &\propto \left[\frac{1}{\mu}\right]^{\frac{(\varepsilon_w+1)\varepsilon_e}{\varepsilon_w[1-\xi(\varepsilon_e+1)]+1}}, \\
Q_{ij}(\theta_e) &\propto \left[\frac{1}{\mu}\right]^{\frac{(\varepsilon_e+1)\varepsilon_w\xi+(\varepsilon_w+1)\varepsilon_e}{\varepsilon_w[1-\xi(\varepsilon_e+1)]+1}}, \\
P(\theta_e) &\propto \left[\frac{1}{\mu}\right]^{0}, \\
l_w(\theta_w) &\propto \left(\frac{1}{\mu}\right)^{\frac{\varepsilon_w(\varepsilon_e+1)}{\varepsilon_w[1-\xi(\varepsilon_e+1)]+1}}, \\
c_w(\theta_w) &\propto \left(\frac{1}{\mu}\right)^{\frac{(\varepsilon_w+1)(\varepsilon_e+1)}{\varepsilon_w[1-\xi(\varepsilon_e+1)]+1}}, \\
V_w(\theta_w) &\propto \left(\frac{1}{\mu}\right)^{\frac{(\varepsilon_w+1)(\varepsilon_e+1)}{\varepsilon_w[1-\xi(\varepsilon_e+1)]+1}}, \forall\theta_w \in \Theta_w.
\end{aligned}
$$

It's easy to see that under the conditions (25) and (24), $L_w(\theta_e)$, $S(\theta_e)$, $l_e(\theta_e)$, $Q_{ij}(\theta_e)$, and $P(\theta_e)$ go down with the decrease of I. Since the markup is uniform, firm-level labor shares must go down too.

We now prove the part three. By (OA34) and (**??**) we have:

$$
\begin{aligned}
c_e(\theta_e) &\propto [\mu-\xi]\left[\frac{1}{\mu}\right]^{\frac{(\varepsilon_w+1)(\varepsilon_e+1)}{\varepsilon_w[1-\xi(\varepsilon_e+1)]+1}}, \\
V_e(\theta_e) &\propto \left[\mu-\xi-\frac{\varepsilon_e}{1+\varepsilon_e}\right]\left[\frac{1}{\mu}\right]^{\frac{(\varepsilon_w+1)(\varepsilon_e+1)}{\varepsilon_w[1-\xi(\varepsilon_e+1)]+1}}, \forall\theta_e \in \Theta_e.
\end{aligned}
$$

Notice that

$$
\frac{d\ln c(\theta_e)}{d\ln\mu} \geq 0 \Leftrightarrow \frac{\mu-\xi}{\mu} \leq \frac{\varepsilon_w+1-\varepsilon_w\xi\left(\varepsilon_e+1\right)}{\left(\varepsilon_w+1\right)\left(\varepsilon_e+1\right)}.
$$

One can see that the condition for $\frac{d\ln c_{ij}(\theta_e)}{d\ln\mu} \geq 0$ is

$$
\mu \leq \frac{\xi}{\frac{\varepsilon_e}{1+\varepsilon_e}+\frac{\varepsilon_w}{\varepsilon_w+1}\xi}
$$

On the other hand,

$$
\frac{dV_e(\theta_e)}{d\ln\mu} \propto \left[\xi+\frac{\varepsilon_e}{1+\varepsilon_e}\right]-\left[\frac{\varepsilon_w}{1+\varepsilon_w}\xi+\frac{\varepsilon_e}{1+\varepsilon_e}\right]\mu.
$$

Therefore, $\mu \leq \frac{\xi+\frac{\varepsilon_e}{1+\varepsilon_e}}{\frac{\varepsilon_e}{1+\varepsilon_e}+\frac{\varepsilon_w}{1+\varepsilon_w}\xi}$ is a condition for $\frac{dV_e(\theta_e)}{d\ln\mu} \geq 0$.■

OB Supplements to Solution

OB.1 Validity of FOA

The solution to the relaxed planner's problem might not be a solution to the original planner's problem, because the first-order conditions are only necessary conditions. In a perfectly competitive economy, the validity of the FOA is guaranteed by the standard Spence-Mirrlees condition on preference and a monotonicity condition on the allocation. However, under the Cournot competition, the conditions are more involved. Lemma 2 provides an insight for when the first-order approach is valid. For the quantitative analysis in Section 7, we verify incentive compatibility as well as any omitted non-negativity constraints numerically.

Lemma 2 *Under Assumption 1, the first-order incentive condition* $\frac{\partial V_e(\theta_e'|\theta_e)}{\partial\theta_e'}|_{\theta_e'=\theta_e} = 0$ *is not only a necessary but also a sufficient condition for the agent's problem if for any* $(\theta_e, \theta_e') \in \Theta^2$, $\frac{1+\varepsilon_e}{\varepsilon_e}\frac{\partial \ln l_e(\theta_e'|\theta_e)}{\partial\theta_e}\frac{\partial \ln l_e(\theta_e'|\theta_e)}{\partial\theta_e'} + \frac{\partial^2 \ln l_e(\theta_e'|\theta_e)}{\partial\theta_e'\partial\theta_e} < 0$.

Proof. See Online Appendix OB.2. ■

Notice that $\frac{1+\varepsilon_e}{\varepsilon_e} > 0$, a stronger condition for the validity of the first-order approach is $\frac{\partial \ln l_e(\theta_e'|\theta_e)}{\partial\theta_e} < 0$, $\frac{\partial \ln l_e(\theta_e'|\theta_e)}{\partial\theta_e'} > 0$ and $\frac{\partial^2 \ln l_e(\theta_e'|\theta_e)}{\partial\theta_e\partial\theta_e'} \le 0$. $\frac{\partial \ln l_e(\theta_e'|\theta_e)}{\partial\theta_e} < 0$ means the higher the ability of the entrepreneur, the lower the effort needed to finish the task. $\frac{\partial \ln l_e(\theta_e'|\theta_e)}{\partial\theta_e'} > 0$ means the higher the reported type, the more effort needed to finish the task. Last, $\frac{\partial^2 \ln l_e(\theta_e'|\theta_e)}{\partial\theta_e'\partial\theta_e} \le 0$ means $\frac{\partial \ln l_e(\theta_e'|\theta_e)}{\partial\theta_e'}$ non-increase in θ_e. All of the above conditions hold in the monopoly competitive case with $X_e'(\cdot) > 0$. Incentive compatibility in more general case needs to be discussed with specific examples.

Here we establish implementability of the constrainted optimal allocation.[52,53]

Lemma 3 *Suppose that the FOCs of the agents and the final good producers are both necessary and sufficient. Symmetric Cournot competitive tax equilibrium* $\{\mathcal{A}, \mathcal{T}, \mathcal{P}\}$ *satisfies the following conditions jointly:*

(i) Incentive conditions (26) and (29) are satisfied;

(ii) Prices and wage satisfy (15) and (16);

(iii) Market clearing conditions (12) to (14) are satisfied.

Conversely, suppose the allocation $\mathcal{A}$ *and price* $\mathcal{P}$ *satisfy the above parts (i) to (iii). Then there exists a tax system* $\mathcal{T}$ *with* $t_s = 0$ *such that the allocation* $\mathcal{A}$ *can be implemented at the prices* $\mathcal{P}$ *by the tax system* $\mathcal{T}$.

Proof. See Online Appendix OB.3. ■

[52] To see why the sales tax is redundant, suppose that $\{T_w(y), T_e(\pi), t_s\}$ is the optimal tax that implements the second-best allocation and that there exists another optimal tax system $\{T_w^{\#}(y), T_e^{\#}(\pi), t_s^{\#}\}$ that can implement the second-best allocation with $t_s^{\#} = 0$. Then the tax system can be constructed such that $1 - T_o^{\#}(x) = [1 - T_o'(x)](1 - t_s), x \in \mathbb{R}_+$.

[53] In our model, we allow profit and labor income tax to be different, which governs the wage rate, so that there is no need to use the sales tax to manipulate W to achieve indirect redistribution between the entrepreneur and worker. The sales tax is not redundant if income taxes on labor income and profit are restricted to be uniform (see e.g., Scheuer (2014)).

OB.2 Proof of Lemma 2

(i) We first show that there is a unique solution to problem (28). Remember that we have set $Q_{ij}\left(\theta'_e|\theta_e\right) = Q_{ij}\left(x_e\left(\theta_e\right) l_e\left(\theta'_e|\theta_e\right), L_w\left(\theta'_e|\theta_e\right)\right)$ and $P_{ij}\left(\theta'_e|\theta_e\right) = P_{ij}\left(Q_{ij}\left(\theta'_e|\theta_e\right), Q_{-ij}\left(\theta_e\right), \theta_e\right)$. The first-order condition of problem (28),

$$\frac{\partial\left[P_{ij}\left(\theta'_e|\theta_e\right) Q_{ij}\left(\theta'_e|\theta_e\right)\left(1-t_s\right) - WL_w\left(\theta'_e|\theta_e\right)\right]}{\partial L_w\left(\theta'_e|\theta_e\right)} = 0, \tag{OB2}$$

is equivalent to

$$\left[1+\frac{\partial \ln P_{ij}\left(\theta'_e|\theta_e\right)}{\partial \ln Q_{ij}\left(\theta'_e|\theta_e\right)}\right]\frac{P_{ij}\left(\theta'_e|\theta_e\right) Q_{ij}\left(\theta'_e|\theta_e\right)}{L_w\left(\theta'_e|\theta_e\right)}\xi\left(1-t_s\right) - W = 0.$$

The second-order condition of problem (28),

$$\begin{aligned}
0 \quad > \quad & \frac{\partial\left[1+\frac{\partial \ln P_{ij}\left(\theta'_e|\theta_e\right)}{\partial \ln Q_{ij}\left(\theta'_e|\theta_e\right)}\right]}{\partial L_w\left(\theta'_e|\theta_e\right)}\frac{P_{ij}\left(\theta'_e|\theta_e\right) Q_{ij}\left(\theta'_e|\theta_e\right)}{L_w\left(\theta'_e|\theta_e\right)}\xi\left(1-t_s\right) - \\
& \left[1+\frac{\partial \ln P_{ij}\left(\theta'_e|\theta_e\right)}{\partial \ln Q_{ij}\left(\theta'_e|\theta_e\right)}\right]\frac{P_{ij}\left(\theta'_e|\theta_e\right) Q_{ij}\left(\theta'_e|\theta_e\right)}{L_w\left(\theta'_e|\theta_e\right)^2}\xi\left(1-t_s\right) + \\
& \left[1+\frac{\partial \ln P_{ij}\left(\theta'_e|\theta_e\right)}{\partial \ln Q_{ij}\left(\theta'_e|\theta_e\right)}\right]\frac{1}{L_w\left(\theta'_e|\theta_e\right)}\frac{\partial\left[P_{ij}\left(\theta'_e|\theta_e\right) Q_{ij}\left(\theta'_e|\theta_e\right)\right]}{\partial L_w\left(\theta'_e|\theta_e\right)}\xi\left(1-t_s\right),
\end{aligned}$$

holds because

$$\frac{\partial\left[1+\frac{\partial \ln P_{ij}\left(\theta'_e|\theta_e\right)}{\partial \ln Q_{ij}\left(\theta'_e|\theta_e\right)}\right]}{\partial L_w\left(\theta'_e|\theta_e\right)}\frac{P_{ij}\left(\theta'_e|\theta_e\right) Q_{ij}\left(\theta'_e|\theta_e\right)}{L_w\left(\theta'_e|\theta_e\right)}\xi\left(1-t_s\right) < 0$$

and

$$\begin{aligned}
& \frac{1}{L_w\left(\theta'_e|\theta_e\right)}\frac{\partial\left[P_{ij}\left(\theta'_e|\theta_e\right) Q_{ij}\left(\theta'_e|\theta_e\right)\right]}{\partial L_w\left(\theta'_e|\theta_e\right)} - \frac{P_{ij}\left(\theta'_e|\theta_e\right) Q_{ij}\left(\theta'_e|\theta_e\right)}{L_w\left(\theta'_e|\theta_e\right)^2} \\
= \; & \frac{P_{ij}\left(\theta'_e|\theta_e\right) Q_{ij}\left(\theta'_e|\theta_e\right)}{L_w\left(\theta'_e|\theta_e\right)^2}\left[\frac{\partial \ln\left[P_{ij}\left(\theta'_e|\theta_e\right) Q_{ij}\left(\theta'_e|\theta_e\right)\right]}{\partial \ln L_w\left(\theta'_e|\theta_e\right)} - 1\right] \\
= \; & \frac{P_{ij}\left(\theta'_e|\theta_e\right) Q_{ij}\left(\theta'_e|\theta_e\right)}{L_w\left(\theta'_e|\theta_e\right)^2}\left[\frac{\partial \ln\left[P_{ij}\left(\theta'_e|\theta_e\right) Q_{ij}\left(\theta'_e|\theta_e\right)\right]}{\partial \ln Q_{ij}\left(\theta'_e|\theta_e\right)}\xi - 1\right] < 0
\end{aligned}$$

hold for any $L_w\left(\theta'_e|\theta_e\right)$ and $l_e\left(\theta'_e|\theta_e\right)$. Therefore, there must be a unique solution to the problem (28).

(ii) According to the definition of $V_e(\theta'_e|\theta_e)$, we have

$$\frac{\partial V_e(\theta'_e|\theta_e)}{\partial \theta'_e} = c'_e\left(\theta'_e\right) - \phi'_e\left(l_e\left(\theta'_e|\theta_e\right)\right)\frac{\partial l_e\left(\theta'_e|\theta_e\right)}{\partial \theta'_e}. \tag{OB3}$$

The first-order necessary condition $\frac{\partial V_e(\theta'_e|\theta_e)}{\partial \theta'_e}|_{\theta'_e=\theta_e} = 0$ implies $\lim_{\theta_e \to \theta'_e}\frac{\partial V_e(\theta'_e|\theta_e)}{\partial \theta'_e} = 0$, i.e.,

$$0 = \left[c'_e\left(\theta'_e\right) - \phi'_e\left(l_e\left(\theta'_e|\theta_e\right)\right)\frac{\partial l_e\left(\theta'_e|\theta_e\right)}{\partial \theta'_e}\right]|_{\theta_e=\theta'_e}, \tag{OB4}$$

Adding (OB3) into (OB4), we have

$$\frac{\partial V_e(\theta_e'|\theta_e)}{\partial \theta_e'} = \left[\phi_e'\left(l_e\left(\theta_e'|\theta_e\right)\right)\frac{\partial l_e\left(\theta_e'|\theta_e\right)}{\partial \theta_e'}\right]|_{\theta_e=\theta_e'} - \phi_e'\left(l_e\left(\theta_e'|\theta_e\right)\right)\frac{\partial l_e\left(\theta_e'|\theta_e\right)}{\partial \theta_e'}$$

Using the mean value theorem, the sign of the right-hand side is given by

$$\frac{d\left[\phi_e'\left(l_e\left(\theta_e'|\theta_e^*\right)\right)\frac{\partial l_e(\theta_e'|\theta_e^*)}{\partial \theta_e'}\right]}{d\theta^*}\left(\theta_e'-\theta_e\right)$$

for some θ_e^* that lies between θ_e' and θ_e. If one has $\frac{d\left[\phi_e'(l_e(\theta_e'|\theta_e^*))\frac{\partial l_e(\theta_e'|\theta_e^*)}{\partial \theta_e'}\right]}{d\theta^*} < 0$ for any $(\theta_e^*, \theta_e') \in \Theta^2$, the function $V_e(\theta_e'|\theta_e)$ will increase with θ_e' until $\theta_e' = \theta_e$ and then decreases with θ_e'. In conclusion, there is a unique the local maximum point that is also the global maximizer of $V_e(\theta_e'|\theta_e)$. Then under Assumption 1, the first-order incentive condition is not only necessary but also sufficient for the agent's problem.

Notice that

$$\frac{\partial^2 \ln l_e\left(\theta_e'|\theta_e^*\right)}{\partial \theta_e^* \partial \theta_e'} = \frac{1}{l_e\left(\theta_e'|\theta_e^*\right)}\frac{\partial^2 l_e\left(\theta_e'|\theta_e^*\right)}{\partial \theta_e' \partial \theta_e^*} - \frac{1}{l_e\left(\theta_e'|\theta_e^*\right)^2}\frac{\partial l_e\left(\theta_e'|\theta_e^*\right)}{\partial \theta_e^*}\frac{\partial l_e\left(\theta_e'|\theta_e^*\right)}{\partial \theta_e'}$$

and

$$\begin{aligned}\frac{d\left[\phi_e'\left(l_e\left(\theta_e'|\theta_e^*\right)\right)\frac{\partial l_e(\theta_e'|\theta_e^*)}{\partial \theta_e'}\right]}{d\theta^*} &= \phi_e''\left(l_e\left(\theta_e'|\theta_e^*\right)\right)\frac{\partial l_e\left(\theta_e'|\theta_e^*\right)}{\partial \theta_e^*}\frac{\partial l_e\left(\theta_e'|\theta_e^*\right)}{\partial \theta_e'} + \phi_e'\left(l_e\left(\theta_e'|\theta_e^*\right)\right)\frac{\partial^2 l_e\left(\theta_e'|\theta_e^*\right)}{\partial \theta_e^* \partial \theta_e'} \\ &= \phi_e'\left(l_e\left(\theta_e'|\theta_e^*\right)\right) l_e\left(\theta_e'|\theta_e^*\right)\left[\begin{array}{c}\frac{\phi_e''(l_e(\theta_e'|\theta_e^*))l_e(\theta_e'|\theta_e^*)}{\phi_e'(l_e(\theta_e'|\theta_e^*))}\frac{\partial \ln l_e(\theta_e'|\theta_e^*)}{\partial \theta_e^*}\frac{\partial \ln l_e(\theta_e'|\theta_e^*)}{\partial \theta_e'} \\ +\frac{1}{l_e(\theta_e'|\theta_e^*)}\frac{\partial^2 l_e(\theta_e'|\theta_e^*)}{\partial \theta_e^* \partial \theta_e'}\end{array}\right].\end{aligned}$$

We have

$$\frac{d\left[\phi_e'\left(l_e\left(\theta_e'|\theta_e^*\right)\right)\frac{\partial l_e(\theta_e'|\theta_e^*)}{\partial \theta_e'}\right]}{d\theta^*} = \phi_e'\left(l_e\left(\theta_e'|\theta_e^*\right)\right) l_e\left(\theta_e'|\theta_e^*\right)\left[\begin{array}{c}\left[\frac{\phi_e''(l_e(\theta_e'|\theta_e^*))l_e(\theta_e'|\theta_e^*)}{\phi_e'(l_e(\theta_e'|\theta_e^*))}+1\right]\frac{\partial \ln l_e(\theta_e'|\theta_e^*)}{\partial \theta_e^*}\frac{\partial \ln l_e(\theta_e'|\theta_e^*)}{\partial \theta_e'} \\ +\frac{\partial^2 \ln l_e(\theta_e'|\theta_e^*)}{\partial \theta_e^* \partial \theta_e'}\end{array}\right].$$

It can be seen that $\frac{d\left[\phi_e'(l_e(\theta_e'|\theta_e^*))\frac{\partial l_e(\theta_e'|\theta_e^*)}{\partial \theta_e'}\right]}{d\theta^*} < 0$ just means that the second-order necessary condition is satisfied. To see this, notice that the first-order necessary condition implies $\left[\frac{\partial^2 V_e(\theta_e'|\theta_e)}{\partial (\theta_e')^2}d\theta_e' + \frac{\partial^2 V_e(\theta_e'|\theta_e)}{\partial \theta_e' \partial \theta_e}d\theta_e\right]|_{|_{\theta_e'=\theta_e}} = 0$. The second-order necessary condition $\frac{\partial^2 V_e(\theta_e'|\theta_e)}{\partial (\theta_e')^2} \leq 0$ is equivalent to $\frac{\partial^2 V_e(\theta_e'|\theta_e)}{\partial \theta_e' \partial \theta_e} \geq 0$ under the first-order necessary condition, i.e.,

$$-\left[\frac{\phi_e''\left(l_e\left(\theta_e'|\theta_e\right)\right) l_e\left(\theta_e'|\theta_e\right)}{\phi_e'\left(l_e\left(\theta_e'|\theta_e\right)\right)}\frac{1}{l_e\left(\theta_e'|\theta_e\right)}\frac{\partial l_e\left(\theta_e'|\theta_e\right)}{\partial \theta_e}\frac{\partial l_e\left(\theta_e'|\theta_e\right)}{\partial \theta_e'} + \frac{\partial^2 l_e\left(\theta_e'|\theta_e\right)}{\partial \theta_e' \partial \theta_e}\right]\phi_e'\left(l_e\left(\theta_e'|\theta_e\right)\right) \geq 0.$$

Notice that $\frac{\partial^2 l_e(\theta_e'|\theta_e)}{\partial \theta_e \partial \theta_e'} = \frac{\partial^2 \ln l_e(\theta_e'|\theta_e)}{\partial \theta_e \partial \theta_e'} l_e\left(\theta_e'|\theta_e\right) + \frac{\partial l_e(\theta_e'|\theta_e)}{\partial \theta_e}\frac{\partial l_e(\theta_e'|\theta_e)}{\partial \theta_e'}\frac{1}{l_e(\theta_e'|\theta_e)}$. The above inequality is equivalent to

$$-\left[\left(1+\frac{1}{\varepsilon_e}\right)\frac{\partial \ln l_e\left(\theta_e'|\theta_e\right)}{\partial \theta_e}\frac{\partial \ln l_e\left(\theta_e'|\theta_e\right)}{\partial \theta_e'} + \frac{\partial^2 \ln l_e\left(\theta_e'|\theta_e\right)}{\partial \theta_e \partial \theta_e'}\right] l_e\left(\theta_e'|\theta_e\right)\phi_e'\left(l_e\left(\theta_e'|\theta_e\right)\right) \geq 0.$$

Now that

$$\text{sign}\left(\frac{d\left[\phi_e'\left(l_e\left(\theta_e'|\theta_e^*\right)\right)\frac{\partial l_e(\theta_e'|\theta_e^*)}{\partial\theta_e'}\right]}{d\theta^*}\right)=\text{sign}\left(\begin{array}{c}\left(1+\frac{1}{\varepsilon_e}\right)\frac{\partial\ln l_e(\theta_e'|\theta_e^*)}{\partial\theta_e^*}\frac{\partial\ln l_e(\theta_e'|\theta_e^*)}{\partial\theta_e'}\\+\frac{\partial^2\ln l_e(\theta_e'|\theta_e^*)}{\partial\theta_e^*\partial\theta_e'}\end{array}\right),$$

one can see that $\left(1+\frac{1}{\varepsilon_e}\right)\frac{\partial\ln l_e(\theta_e'|\theta_e^*)}{\partial\theta_e^*}\frac{\partial\ln l_e(\theta_e'|\theta_e^*)}{\partial\theta_e'}+\frac{\partial^2\ln l_e(\theta_e'|\theta_e^*)}{\partial\theta_e^*\partial\theta_e'}<0$ for any $\left(\theta_e^*,\theta_e'\right)\in\Theta^2$ is a sufficient condition for the validity of FOA.■

OB.3 Proof of Lemma 3

We first show that a symmetric Cournot competitive tax equilibrium must satisfy parts 1 to 3. First, by the definition of SCCTE, (15) to (16) and (12) to (14) must be satisfied. Second, by the definition of SCCTE, agents maximize their utilities, which, by envelop theory, means (26) and (29) must be satisfied (see e.g., Lemma 1).

Next, suppose that we are given an allocation $\mathcal{A}$ and price $\mathcal{P}$ to satisfy the properties in parts 1 to 3. We now construct the tax system $\mathcal{T}$ (with $t_s=0$), which together with the given allocation $\mathcal{A}$ and price $\mathcal{P}$ constructs a SCCTE. We first construct a policy system with the given allocation $\mathcal{A}$ and price $\mathcal{P}$. We then show that this constructed policy system together with $\mathcal{A}$ and $\mathcal{P}$ constructs a SCCTE.

First, we construct the policy system. By the definition of tax wedges and $t_s=0$, the marginal tax rates are constructed as follows:

$$T_w'\left(y_w\left(\theta_w\right)\right)=1-\frac{\phi_w'\left(l_w\left(\theta_w\right)\right)}{\frac{P(\theta_e)}{\mu(\theta_e)}\frac{\partial Q_{ij}(\theta_e)}{\partial L_w(\theta_e)}x_w\left(\theta_w\right)}$$

and

$$T_e'\left(y_e\left(\theta_e\right)\right)=1-\frac{\phi_e'\left(l_e\left(\theta_e\right)\right)}{\frac{P(\theta_e)}{\mu(\theta_e)}\frac{\partial Q_{ij}(\theta_e)}{\partial l_e(\theta_e)}}.$$

We use agents' budget constraints to fix the labor income taxes. We first construct $T_w\left(\cdot\right)$. To do this, we substitute

$$T_w\left(y_w(\theta_w)\right)=T_w\left(y_w(\underline{\theta}_w)\right)+\int_{y_w(\underline{\theta}_w)}^{y_w(\theta_w)}T_w'\left(y_w\right)dy$$

into

$$y_w(\underline{\theta}_w)-T_w\left(y_w(\underline{\theta}_w)\right)-c_w(\underline{\theta}_w)=0.$$

We now show that the constructed labor income tax is consistent with the agent's problems. By construction, we have shown that the constructed $T_w\left(\cdot\right)$ is consistent with the $\underline{\theta}_w$-type agent's action. We should show the constructed labor income tax is also consistent with other agents' choices. In particular, since by construction the first-order conditions are already satisfied, we should show that the budget constraints are also satisfied. This is equivalent to say that for any $\theta_w\in\Theta_w$,

$$y_w'(\theta_w)\left[1-T_w'\left(y_w(\theta_w)\right)\right]-c_w'(\theta_w)=0.$$

Substituting $1-T_w'(y_w(\theta_w))$ with the FOC (17), the above equation is equivalent to

$$c_w'(\theta_w) - \frac{\phi_w'(l_w(\theta_w))}{Wx_w(\theta_w)} y_w'(\theta_w) = 0.$$

The above equations holds because for any $\theta_w \in \Theta_w$ the allocation satisfies:

$$\begin{aligned} V_w'(\theta_w) &= \frac{\phi_w'(l_w(\theta_w))\, l_w(\theta_w) x_w'(\theta_w)}{x_w(\theta_w)} \\ &= c_w'(\theta_w) - \frac{y_w'(\theta_w)}{Wx_w(\theta_w)} \phi_w'(l_w(\theta_w)) + \frac{\phi_w'(l_w(\theta_w))\, l_w(\theta_w) x_w'(\theta_w)}{x_w(\theta_w)}. \end{aligned} \tag{OB5}$$

The first equation of (OB5) is the incentive condition, and the second equation is derived through the definition of $V_w(\theta_w)$. In conclusion, given the allocation, we can construct a unique labor income tax that is consistent with the allocation in the equilibrium.

The construction of $T_e(\cdot)$ is similar to the construction of $T_w(\cdot)$. Note that $T_w(y_w(\underline{\theta}_w))$ can be different from $T_e(y_e(\underline{\theta}_e))$. We substitute

$$T_e(y_e(\theta_e)) = T_e(y_e(\underline{\theta}_e)) + \int_{y_e(\underline{\theta}_e)}^{y_e(\theta_e)} T_e'(y_e)\, d\pi$$

into

$$y_e(\theta_e) - T_e(y_e(\theta_e)) - c_e(\theta_e) = 0$$

and show that there exists $T_e(\cdot)$ such that given allocation $\mathcal{A}$, price $\mathcal{P}$ and the constructed marginal profit income tax rates, the above equation is satisfied for any $\theta_e \in \Theta_e$. To be consistent with the $\underline{\theta}_e$-type agent's budget constraint, $T_e(y_e(\underline{\theta}_e))$ must satisfy

$$y_e(\underline{\theta}_e) - T_e(y_e(\underline{\theta}_e)) - c_e(\underline{\theta}_e) = 0.$$

We should show this constructed profit tax on $y_e(\underline{\theta}_e)$, i.e., $T_e(y_e(\underline{\theta}_e))$, is also consistent with other agents' budget constraints. This is equivalent to say that

$$y_e(\theta_e)\left[1 - T_e'(y_e(\theta_e))\right] - c_e'(\theta_e) = 0.$$

Substituting $1 - T_e'(y_e(\theta_e))$ with the FOC (18), the above equation is equivalent to

$$c_e'(\theta_e) - \frac{\phi_e'(l_e(\theta_e))}{\frac{P(\theta_e)}{\mu(\theta_e)} \frac{\partial Q_{ij}(\theta_e)}{\partial l_e(\theta_e)}} y_e(\theta_e) = 0,$$

which is further equivalent to

$$c_e'(\theta_e) - \mu(\theta_e) \frac{\phi_e'(l_e(\theta_e))\, l_e(\theta_e)}{P(\theta_e)\, Q_{ij}(\theta_e)} y_e(\theta_e) = 0. \tag{OB6}$$

The above equations are true since we have

$$
\begin{aligned}
V_e'(\theta_e) &= \phi_e'\left(l_e\left(\theta_e\right)\right) l_e\left(\theta_e\right)\left[\mu\left(\theta_e\right)\frac{d\ln P_{ij}\left(Q_{ij},Q_{ij}\left(\theta_e\right),\theta_e\right)}{d\theta_e}|_{Q_{ij}=Q_{ij}(\theta_e)}+\frac{\zeta_e'\left(\theta_e\right)}{\zeta_e\left(\theta_e\right)}\right] \\
&= c_e'(\theta_e)-\phi_e'\left(l_e\left(\theta_e\right)\right) l_e\left(\theta_e\right)\frac{l_e'\left(\theta_e\right)}{l_e\left(\theta_e\right)},
\end{aligned}
\tag{OB7}
$$

and

$$
y_e(\theta_e)=Q_{ij}(\theta_e)P\left(\theta_e\right)\left[\begin{array}{c}\frac{d\ln P_{ij}\left(Q_{ij},Q_{ij}(\theta_e),\theta_e\right)}{d\theta_e}|_{Q_{ij}=Q_{ij}(\theta_e)}+ \\ \left[1+\frac{\partial\ln P_{ij}\left(Q_{ij}(\theta_e),Q_{-ij},\theta_e\right)}{\partial\ln Q_{ij}(\theta_e)}|_{Q_{-ij}=Q_{ij}(\theta_e)}\right]\left[\frac{x_e'(\theta_e)}{x_e(\theta_e)}+\frac{l_e'(\theta_e)}{l_e(\theta_e)}\right]\end{array}\right]. \tag{OB8}
$$

Substituting $\frac{l_e'(\theta_e)}{l_e(\theta_e)}$ in (OB7) by (OB8) and using $1+\frac{\partial\ln P_{ij}\left(Q_{ij}(\theta_e),Q_{-ij},\theta_e\right)}{\partial\ln Q_{ij}(\theta_e)}|_{Q_{-ij}=Q_{ij}(\theta_e)}=\frac{1}{\mu(\theta_e)}$ delivers (OB6) immediately. The first equation of (OB7) is the incentive condition, and the second equation is derived through the definition of $V_e(\theta_e)$. (OB8) is derived from the definition of $y_e(\theta_e)$ (i.e., $y_e(\theta_e)=P(\theta_e)Q_{ij}(\theta_e)-WL_w(\theta_e)$) and the fact that the derivative of $y_e(\theta_e)$ with respect to $L_w(\theta_e)$ is zero.

In conclusion, given the allocation, we can construct a unique combination of labor income tax and profit tax that is consistent with the allocation in the equilibrium.

We now show that the allocation $\mathcal{A}$ and price $\mathcal{P}$ satisfying parts 1 to 3 and the constructed tax system $\mathcal{T}$ construct an SCCTE. First, the constructed allocation and taxation satisfy the first-order conditions. Since we have assumed that the first-order conditions are both necessary and sufficient, the allocation solves agents' problems. Second, the price $\mathcal{P}$ satisfies (15) and (16). Third, the market clear conditions (12) to (14) are satisfied. Lastly, agents' budget constraints (8) and (11) are embedded in the definitions of gross utilities and the construction of income taxes. In conclusion, the constructed tax system $\mathcal{T}$ together with the given allocation $\mathcal{A}$ and price $\mathcal{P}$ constructs an SCCTE.■

OC Supplements to Main Results

OC.1 An Explicit Formula of Tax Rate

Theorem 1 fully describes the optimal tax wedges, but the tax rate for the entrepreneurs $\tau_e(\theta_e)$ cannot be written explicitly because the weights $\omega(\theta_e)$ in equation (39) and hence the average markup μ is a function of the tax rate $\tau_e(\theta_e)$. In what follows, we can write the tax rate explicitly under a particular parameter configuration. Than we can solve explicitly for the weights $\omega(\theta_e)$ and therefore we can write the average markup μ explicitly in the following Corollary:

Corollary 2 *When $\frac{1+\varepsilon_e}{\varepsilon_e}\left(\frac{\sigma}{\sigma-1}-\xi\right)=2$, we have*

$$\mu=\frac{\left(\frac{\sigma}{\sigma-1}+\xi\right)\int_{\theta_e}\overline{\gamma}(\theta_e)\mu\left(\theta_e\right)d\theta_e}{2\xi}-\frac{\sqrt{\Delta}}{2\xi},$$

where

$$\begin{aligned}
\overline{\gamma}(\theta_e) &= \frac{\gamma(\theta_e)}{\int_{\theta_e}\gamma(\theta_e)d\theta_e},\\
\gamma(\theta_e) &= \frac{\left(\frac{X_e(\theta_e)}{\mu(\theta_e)}\right)^{\frac{1+\varepsilon_e}{\varepsilon_e}\frac{\sigma}{\sigma-1}}f_e\left(\theta_e\right)}{1+\left[1-\bar{g}_e(\theta_e)\right]H(\theta_e)\left[\frac{1}{\varepsilon_e}\mu\left(\theta_e\right)-\frac{1+\varepsilon_e}{\varepsilon_e}\xi+\frac{1}{\sigma-1}\right]+\left[\mu\left(\theta_e\right)-\frac{\sigma}{\sigma-1}\right]\left[1-g_e(\theta_e)\right]},\\
\Delta &= \left(\frac{\sigma}{\sigma-1}-\xi\right)^2\left[\int_{\theta_e}\overline{\gamma}(\theta_e)\mu\left(\theta_e\right)d\theta_e\right]^2-4\frac{\sigma}{\sigma-1}\xi\left[\int_{\theta_e}\mu\left(\theta_e\right)^2\overline{\gamma}(\theta_e)d\theta_e-\left(\int_{\theta_e}\overline{\gamma}(\theta_e)\mu\left(\theta_e\right)d\theta_e\right)^2\right].
\end{aligned}$$

Proof. See Online Appendix OC.2. ■

Corollary 2 provides a special case where we can provide explicit optimal tax formulas with given social welfare weights. Note that $\frac{1+\varepsilon_e}{\varepsilon_e}\left(\frac{\sigma}{\sigma-1}-\xi\right)=2$ is consistent with condition (24). Corollary 2 thus suggests that under some reasonable parameter values, there is a unique and well-defined solution to the equation system (36) and (39).

OC.2 Proof of Corollary 2

Substituting elasticities, $RE\left(\theta_e\right)$ and $IRE\left(\theta_e\right)$ in (36) by (37) and (38), we have:

$$1-\tau_e\left(\theta_e\right)=\frac{\frac{\frac{\sigma}{\sigma-1}}{\frac{\sigma}{\sigma-1}-\xi}-\frac{\xi}{\frac{\sigma}{\sigma-1}-\xi}\frac{\mu}{\mu(\theta_e)}}{\frac{1+\left[1-\bar{g}_e(\theta_e)\right]H(\theta_e)\left[\frac{1}{\varepsilon_e}\mu(\theta_e)-\frac{1+\varepsilon_e}{\varepsilon_e}\xi+\frac{1}{\sigma-1}\right]+\left[\mu(\theta_e)-\frac{\sigma}{\sigma-1}\right]\left[1-g_e(\theta_e)\right]}{\mu(\theta_e)}}$$

Notice that $1<\sigma<\eta(\theta_e)$, $\frac{\sigma}{\sigma-1}\geq\mu\left(\theta_e\right)=\frac{1}{1-\left[\frac{1}{\eta(\theta_e)}\frac{I-1}{I}+\frac{1}{\sigma}\frac{1}{I}\right]}>1$, $0\leq\xi<1$. We have

$$1-\frac{\xi}{\frac{\sigma}{\sigma-1}-\xi}RE\left(\theta_e\right)=\frac{1}{\frac{\sigma}{\sigma-1}-\xi}\left[\frac{\sigma}{\sigma-1}\frac{\mu\left(\theta_e\right)}{\mu}-\xi\right]>0.$$

Then, we have

$$1+\left[1-\bar{g}_e(\theta_e)\right]H(\theta_e)\left[\frac{1}{\varepsilon_e}\mu\left(\theta_e\right)-\frac{1+\varepsilon_e}{\varepsilon_e}\xi+\frac{1}{\sigma-1}\right]+\left[\mu\left(\theta_e\right)-\frac{\sigma}{\sigma-1}\right]\left[1-g_e(\theta_e)\right]\geq 0,$$

since $1-\tau_e\left(\theta_e\right)\geq 0$. According to the above inequality, we have $\gamma(\theta_e)\geq 0$ and $\overline{\gamma}(\theta_e)\geq 0$. When

$\frac{1+\varepsilon_e}{\varepsilon_e}\left(\frac{\sigma}{\sigma-1}-\xi\right)=2$, we have

$$\begin{aligned}\mu &= \frac{\int_{\theta_e}\mu\left(\theta_e\right)\left[1-\tau_e\left(\theta_e\right)\right]\left(\frac{X_e(\theta_e)}{\mu(\theta_e)}\right)^{\frac{1+\varepsilon_e}{\varepsilon_e}\frac{\sigma}{\sigma-1}}f_e\left(\theta_e\right)d\theta_e}{\int_{\theta_e}\left[1-\tau_e\left(\theta_e\right)\right]\left(\frac{X_e(\theta_e)}{\mu(\theta_e)}\right)^{\frac{1+\varepsilon_e}{\varepsilon_e}\frac{\sigma}{\sigma-1}}f_e\left(\theta_e\right)d\theta_e} \\ &= \frac{\int_{\theta_e}\mu\left(\theta_e\right)\gamma(\theta_e)\left[\frac{\sigma}{\sigma-1}\mu\left(\theta_e\right)-\xi\mu\right]d\theta_e}{\int_{\theta_e}\gamma(\theta_e)\left[\frac{\sigma}{\sigma-1}\mu\left(\theta_e\right)-\xi\mu\right]d\theta_e},\end{aligned}$$

where the second equation is derived by the relationship between $\gamma(\theta_e)$ and $\left[1-\tau_e\left(\theta_e\right)\right]$. Accordingly, we have

$$\begin{aligned}&\frac{\sigma}{\sigma-1}\mu\int_{\theta_e}\gamma(\theta_e)\mu\left(\theta_e\right)d\theta_e-\xi\mu^2\int_{\theta_e}\gamma(\theta_e)d\theta_e \\ =\;&\frac{\sigma}{\sigma-1}\int_{\theta_e}\mu\left(\theta_e\right)^2\gamma(\theta_e)d\theta_e-\xi\mu\int_{\theta_e}\mu\left(\theta_e\right)\gamma(\theta_e)d\theta_e,\end{aligned}$$

or equivalently

$$0=\xi\mu^2-\left[\left(\frac{\sigma}{\sigma-1}+\xi\right)\int_{\theta_e}\overline{\gamma}(\theta_e)\mu\left(\theta_e\right)d\theta_e\right]\mu+\frac{\sigma}{\sigma-1}\int_{\theta_e}\mu\left(\theta_e\right)^2\overline{\gamma}(\theta_e)d\theta_e, \tag{OC2}$$

which is a quadratic equation of μ.

We define

$$\Delta=\left(\frac{\sigma}{\sigma-1}-\xi\right)^2\left[\int_{\theta_e}\overline{\gamma}(\theta_e)\mu\left(\theta_e\right)d\theta_e\right]^2-4\frac{\sigma}{\sigma-1}\xi\left[\int_{\theta_e}\mu^2\left(\theta_e\right)\overline{\gamma}(\theta_e)d\theta_e-\left(\int_{\theta_e}\overline{\gamma}(\theta_e)\mu\left(\theta_e\right)d\theta_e\right)^2\right]$$

as the discriminant of (OC2). Set

$$\mathbf{E}_{\overline{\gamma}}\left(\mu\right)=\int_{\theta_e}\overline{\gamma}(\theta_e)\mu\left(\theta_e\right)d\theta_e,\ \mathbf{Var}_{\overline{\gamma}}\left(\mu\right)=\left[\int_{\theta_e}\mu^2\left(\theta_e\right)\overline{\gamma}(\theta_e)d\theta_e-\left(\int_{\theta_e}\overline{\gamma}(\theta_e)\mu\left(\theta_e\right)d\theta_e\right)^2\right]$$

and

$$\mu^*\left(\theta_e\right)=(\sigma-1)\left[\mu\left(\theta_e\right)-1\right].$$

We have $\mu^*(\theta_e) \in (0,1]$, because $\mu(\theta_e) \in (1, \frac{\sigma}{\sigma-1}]$. Set

$$\begin{aligned}
\mathbf{E}_{\overline{\gamma}}(\mu^*) &= \int_{\theta_e} \overline{\gamma}(\theta_e)\mu^*(\theta_e)\, d\theta_e, \mathbf{E}_{\overline{\gamma}}\left(\mu^{*2}\right) = \int_{\theta_e} \overline{\gamma}(\theta_e)\mu^*(\theta_e)^2\, d\theta_e, \\
\mathbf{Var}_{\overline{\gamma}}(\mu^*) &= \left[\int_{\theta_e} \mu^*(\theta_e)^2\, \overline{\gamma}(\theta_e) d\theta_e - \left(\int_{\theta_e} \overline{\gamma}(\theta_e)\mu^*(\theta_e)\, d\theta_e\right)^2\right].
\end{aligned}$$

Then, we have

$$\Delta = \left(\frac{\sigma}{\sigma-1} - \xi\right)^2 \left[1 + \frac{1}{\sigma-1}\mathbf{E}_{\overline{\gamma}}(\mu^*)\right]^2 - 4\frac{\sigma}{\sigma-1}\frac{\xi}{(\sigma-1)^2}\mathbf{Var}_{\overline{\gamma}}(\mu).$$

One necessary condition for the exist of solution to (OC2) is $\Delta \geq 0$. Notice that $\mathbf{E}_{\overline{\gamma}}\left(\mu^{*2}\right) \leq \mathbf{E}_{\overline{\gamma}}(\mu^*)$. We have

$$\begin{aligned}
\Delta &= \left(\frac{\sigma}{\sigma-1} - \xi\right)^2 + \frac{2}{\sigma-1}\left(\frac{\sigma}{\sigma-1} - \xi\right)^2 \mathbf{E}_{\overline{\gamma}}(\mu^*) + \frac{\left(\frac{\sigma}{\sigma-1} + \xi\right)^2}{(\sigma-1)^2}\left[\mathbf{E}_{\overline{\gamma}}(\mu^*)\right]^2 \\
&\quad - \frac{2}{\sigma-1}\left(\frac{\sigma}{\sigma-1}\right)^2 \frac{2\xi}{\sigma}\mathbf{E}_{\overline{\gamma}}\left(\mu^{*2}\right) \\
&\geq \Delta_H\left(\mathbf{E}_{\overline{\gamma}}(\mu^*)\right),
\end{aligned} \tag{OC3}$$

where we set

$$\begin{aligned}
\Delta_H\left(\mathbf{E}_{\overline{\gamma}}(\mu^*)\right) &= \left(\frac{\sigma}{\sigma-1} - \xi\right)^2 + \frac{\left(\frac{\sigma}{\sigma-1} + \xi\right)^2}{(\sigma-1)^2}\left[\mathbf{E}_{\overline{\gamma}}(\mu^*)\right]^2 \\
&\quad - \frac{2}{\sigma-1}\left(\frac{\sigma}{\sigma-1}\right)^2 \left[\frac{2\xi}{\sigma} - \left(1 - \frac{\sigma-1}{\sigma}\xi\right)^2\right]\mathbf{E}_{\overline{\gamma}}(\mu^*)
\end{aligned}$$

as a quadratic function of $\mathbf{E}_{\overline{\gamma}}(\mu^*)$.

The minimium value of Δ_H is derived at $\mu^* \equiv \frac{(\sigma-1)\left[\frac{2\sigma-1}{\sigma^2}\xi^2 - (1-\xi)^2\right]}{\left(1+\frac{\sigma-1}{\sigma}\xi\right)^2} < 1$. However, $\mathbf{E}_{\overline{\gamma}}(\mu^*) \in (0,1]$, and thus μ^* may not belong to the domain of $\mathbf{E}_{\overline{\gamma}}(\mu^*)$.

If $\mu^* \leq 0$, to prove that $\Delta \geq 0$, we only need to prove that $\Delta_H \geq 0$ when $\mathbf{E}_{\overline{\gamma}}(\mu^*) = 1$ and $\mathbf{E}_{\overline{\gamma}}(\mu^*) = 0$. This is true. According to (OC3), when $\mathbf{E}_{\overline{\gamma}}(\mu^*) = 1$, we have

$$\begin{aligned}
\Delta_H &= \left(\frac{\sigma}{\sigma-1} - \xi\right)^2 - \frac{2}{\sigma-1}\left(\frac{\sigma}{\sigma-1}\right)^2 \left[\frac{2\xi}{\sigma} - \left(1 - \frac{\sigma-1}{\sigma}\xi\right)^2\right] + \frac{\left(\frac{\sigma}{\sigma-1} + \xi\right)^2}{(\sigma-1)^2} \\
&= \left(\frac{\sigma}{\sigma-1}\right)^2 \left(\frac{\sigma}{\sigma-1} - \xi\right)^2 - \left(\frac{\sigma}{\sigma-1}\right)^2 \frac{2}{\sigma-1}\frac{2\xi}{\sigma} + \left(\frac{\sigma}{\sigma-1}\right)^2 \frac{2}{\sigma-1}\frac{2\xi}{\sigma} \\
&= \left(\frac{\sigma}{\sigma-1}\right)^2 \left(\frac{\sigma}{\sigma-1} - \xi\right)^2 > 0.
\end{aligned}$$

When $\mathbf{E}_{\overline{\gamma}}(\mu^*) = 0$, we have

$$\Delta_H = \left(\frac{\sigma}{\sigma-1} - \xi\right)^2 > 0.$$

If $\mu^* > 0$ (note that μ^* must be lower than one), to prove $\Delta \geq 0$, we only need to prove that $\Delta_H(\mu^*) \geq 0$. To see this, first note that when $\mu^* \in (0,1)$, we have

$$\frac{2\sigma - 1}{\sigma^2} > \left(\frac{1}{\xi} - 1\right)^2$$

and

$$\frac{2}{\sigma-1}\left(\frac{\sigma}{\sigma-1}\right)^2\left[\frac{2\xi}{\sigma} - \left(1 - \frac{\sigma-1}{\sigma}\xi\right)^2\right] > 0.$$

We set

$$\widetilde{\Delta} = \left\{\frac{2}{\sigma-1}\left(\frac{\sigma}{\sigma-1}\right)^2\left[\frac{2\xi}{\sigma} - \left(1 - \frac{\sigma-1}{\sigma}\xi\right)^2\right]\right\}^2 - 4\left(\frac{\sigma}{\sigma-1} - \xi\right)^2\frac{\left(\frac{\sigma}{\sigma-1}+\xi\right)^2}{(\sigma-1)^2}$$

as the discriminant of $\Delta_H\left(\mathbf{E}_{\overline{\gamma}}(\mu^*)\right) > 0$. $\widetilde{\Delta} < 0$ is a sufficient condition for $\Delta > 0$. To prove $\widetilde{\Delta} < 0$, we only need to show that

$$\frac{1}{\sigma-1}\left(\frac{\sigma}{\sigma-1}\right)^2\left[\frac{2\xi}{\sigma} - \left(1 - \frac{\sigma-1}{\sigma}\xi\right)^2\right] < \frac{\left(\frac{\sigma}{\sigma-1}-\xi\right)\left(\frac{\sigma}{\sigma-1}+\xi\right)}{(\sigma-1)},$$

which is equivalent to

$$2\left(\frac{\sigma}{\sigma-1}\right)^2(1-\xi) > 0.$$

Notice that the above inequality must hold. Thus, we must have $\Delta > 0$. In conclusion, there are two solutions to the quadratic equation (OC2). However, $\Delta > 0$ does not necessarily mean that the solutions are all in the domain of μ (i.e., $(1, \frac{\sigma}{\sigma-1}]$).

In the following analysis, we prove that there exists unique solution in the domain of μ. The two potential solutions are μ_1 and μ_2:

$$\begin{aligned}
\mu_1 &= \frac{\left(\frac{\sigma}{\sigma-1}+\xi\right)\int_{\theta_e}\overline{\gamma}(\theta_e)\mu(\theta_e)\,d\theta_e}{2\xi} - \frac{\sqrt{\Delta}}{2\xi} \geq \int_{\theta_e}\overline{\gamma}(\theta_e)\mu(\theta_e)\,d\theta_e, \\
\mu_2 &= \frac{\left(\frac{\sigma}{\sigma-1}+\xi\right)\int_{\theta_e}\overline{\gamma}(\theta_e)\mu(\theta_e)\,d\theta_e}{2\xi} + \frac{\sqrt{\Delta}}{2\xi} \leq \frac{\sigma}{\sigma-1}\frac{1}{\xi}\int_{\theta_e}\overline{\gamma}(\theta_e)\mu(\theta_e)\,d\theta_e.
\end{aligned}$$

In the following analysis, we prove that $\mu_2 > \frac{\sigma}{\sigma-1}$ and $\mu_1 \in (1, \frac{\sigma}{\sigma-1}]$. To prove this, we only need to show that

$$\sqrt{\Delta} > \left|2\xi\frac{\sigma}{\sigma-1} - \left(\frac{\sigma}{\sigma-1}+\xi\right)\int_{\theta_e}\overline{\gamma}(\theta_e)\mu(\theta_e)\,d\theta_e\right|.$$

In particular, $\mu_2 > \frac{\sigma}{\sigma-1}$ is equivalent to

$$\mu_2 = \frac{\left(\frac{\sigma}{\sigma-1}+\xi\right)\int_{\theta_e}\overline{\gamma}(\theta_e)\mu(\theta_e)\,d\theta_e}{2\xi} + \frac{\sqrt{\Delta}}{2\xi} > \frac{\sigma}{\sigma-1},$$

i.e.,

$$\sqrt{\Delta} > 2\xi\frac{\sigma}{\sigma-1} - \left(\frac{\sigma}{\sigma-1}+\xi\right)\int_{\theta_e}\overline{\gamma}(\theta_e)\mu(\theta_e)\,d\theta_e.$$

Substituting $\mu(\theta_e)$ in the above inequality by $\mu(\theta_e) = 1 + \frac{1}{\sigma-1}\mu^*(\theta_e)$, we have

$$\begin{aligned}\sqrt{\Delta} &> 2\xi\frac{\sigma}{\sigma-1} - \left(\frac{\sigma}{\sigma-1} + \xi\right)\left[1 + \frac{1}{\sigma-1}\mathbf{E}_{\overline{\gamma}}(\mu^*)\right] \\ &= \left(\frac{\sigma+1}{\sigma-1}\xi - \frac{\sigma}{\sigma-1}\right) - \left(\frac{\sigma}{\sigma-1} + \xi\right)\frac{1}{\sigma-1}\mathbf{E}_{\overline{\gamma}}(\mu^*).\end{aligned}$$

When $\xi \le \frac{\sigma}{\sigma+1}$, the above inequality must holds. When $\xi > \frac{\sigma}{\sigma+1}$, to prove the above inequality, we only need to show

$$\Delta - \left[\left(\frac{\sigma+1}{\sigma-1}\xi - \frac{\sigma}{\sigma-1}\right) - \left(\frac{\sigma}{\sigma-1} + \xi\right)\frac{1}{\sigma-1}\mathbf{E}_{\overline{\gamma}}(\mu^*)\right]^2 > 0.$$

To see this, notice that

$$\begin{aligned}\Delta_M &< \Delta - \left[\left(\frac{\sigma+1}{\sigma-1}\xi - \frac{\sigma}{\sigma-1}\right) - \left(\frac{\sigma}{\sigma-1} + \xi\right)\frac{1}{\sigma-1}E_{\overline{\gamma}}(\mu^*)\right]^2 \\ &= \left(\frac{\sigma}{\sigma-1} - \xi\right)^2 + \frac{2}{\sigma-1}\left(\frac{\sigma}{\sigma-1} - \xi\right)^2 \mathbf{E}_{\overline{\gamma}}(\mu^*) \\ &\quad + \frac{\left(\frac{\sigma}{\sigma-1} + \xi\right)^2}{(\sigma-1)^2}\left[\mathbf{E}_{\overline{\gamma}}(\mu^*)\right]^2 - \frac{2}{\sigma-1}\left(\frac{\sigma}{\sigma-1}\right)^2 \frac{2\xi}{\sigma}\mathbf{E}_{\overline{\gamma}}(\mu^{*2}) \\ &\quad - \left[\left(\frac{\sigma+1}{\sigma-1}\xi - \frac{\sigma}{\sigma-1}\right) - \left(\frac{\sigma}{\sigma-1} + \xi\right)\frac{1}{\sigma-1}\mathbf{E}_{\overline{\gamma}}(\mu^*)\right]^2,\end{aligned}$$

where

$$\Delta_M = \Delta_H\left(\mathbf{E}_{\overline{\gamma}}(\mu^*)\right) - \left[\left(\frac{\sigma+1}{\sigma-1}\xi - \frac{\sigma}{\sigma-1}\right) - \left(\frac{\sigma}{\sigma-1} + \xi\right)\frac{1}{\sigma-1}\mathbf{E}_{\overline{\gamma}}(\mu^*)\right]^2.$$

Rearranging the right side the the above equation, we have

$$\Delta_M = \frac{2}{\sigma-1}\frac{2\sigma}{\sigma-1}\xi(1-\xi)\left[1 - \mathbf{E}_{\overline{\gamma}}(\mu^*)\right] \ge 0.$$

Since

$$\Delta - \left[\left(\frac{\sigma+1}{\sigma-1}\xi - \frac{\sigma}{\sigma-1}\right) - \left(\frac{\sigma}{\sigma-1} + \xi\right)\frac{1}{\sigma-1}\mathbf{E}_{\overline{\gamma}}(\mu^*)\right]^2 > \Delta_M > 0,$$

we have

$$\mu_2 = \frac{\sigma}{\sigma-1}\frac{1}{\xi}\int_{\theta_e}\overline{\gamma}(\theta_e)\mu(\theta_e)\,d\theta_e > \frac{\sigma}{\sigma-1}.$$

We now prove that

$$\frac{\sigma}{\sigma-1} \ge \mu_1 = \frac{\left(\frac{\sigma}{\sigma-1} + \xi\right)\int_{\theta_e}\overline{\gamma}(\theta_e)\mu(\theta_e)\,d\theta_e}{2\xi} - \frac{\sqrt{\Delta}}{2\xi} > 1$$

First, $\mu_1 \ge \int_{\theta_e}\overline{\gamma}(\theta_e)\mu(\theta_e)\,d\theta_e$, where $\mu(\theta_e) \in (1, \frac{\sigma}{\sigma-1}]$ and $\overline{\gamma}(\theta_e)$ is a densify with support Θ_e. Thus, $\mu_1 > 1$. Second, notice that

$$\Delta - \left[\left(\frac{\sigma+1}{\sigma-1}\xi - \frac{\sigma}{\sigma-1}\right) - \left(\frac{\sigma}{\sigma-1} + \xi\right)\frac{1}{\sigma-1}\mathbf{E}_{\overline{\gamma}}(\mu^*)\right]^2 \ge 0,$$

(see e.g., (OC3)) we have

$$\begin{aligned}\sqrt{\Delta} &\geq \left(\frac{\sigma}{\sigma-1}+\xi\right)\left[1+\frac{1}{\sigma-1}\mathbf{E}_{\overline{\gamma}}\left(\mu^{*}\right)\right]\left(\frac{\sigma}{\sigma-1}+\xi\right)-\frac{2\xi\sigma}{\sigma-1} \\ &= \left(\frac{\sigma}{\sigma-1}+\xi\right)\frac{1}{\sigma-1}\mathbf{E}_{\overline{\gamma}}\left(\mu^{*}\right)-\left(\frac{\sigma+1}{\sigma-1}\xi-\frac{\sigma}{\sigma-1}\right)\end{aligned}$$

The above inequality implies

$$\sqrt{\Delta} \geq \left(\frac{\sigma}{\sigma-1}+\xi\right)\int_{\theta_e}\overline{\gamma}(\theta_e)\mu\left(\theta_e\right)d\theta_e-\frac{2\xi\sigma}{\sigma-1},$$

which gives

$$\mu_1=\frac{\left(\frac{\sigma}{\sigma-1}+\xi\right)\int_{\theta_e}\overline{\gamma}(\theta_e)\mu\left(\theta_e\right)d\theta_e}{2\xi}-\frac{\sqrt{\Delta}}{2\xi}\leq\frac{\sigma}{\sigma-1}.$$

In conclusion, we have

$$\mu=\frac{\left(\frac{\sigma}{\sigma-1}+\xi\right)\int_{\theta_e}\overline{\gamma}(\theta_e)\mu\left(\theta_e\right)d\theta_e}{2\xi}-\frac{\sqrt{\Delta}}{2\xi}.$$

■

OC.3 Statistic-based Optimal Tax Formulas

In the following Theorem OC1, we provide optimal tax rules in terms of social welfare weights and a small number of empirical observable statistics (see Appendix A.2 for the definitions of elasticities).

Theorem OC1 *Suppose the cross wage elasticities $\varepsilon^{\omega}_{L_w}(\theta_e',\theta_e)$ and $\varepsilon^{\omega}_{l_e}(\theta_e',\theta_e)$ are independent θ_e' for any $(\theta_e',\theta_e)\in\Theta_e^2$; the markup $\mu\left(\theta_e\right)$ is exogenous in the equilibrium; and the firm-level production technology is in the form of (1). The optimal tax wedges satisfy the following statistic-based tax formula for any $\theta_w\in\Theta_w$ and $\theta_e\in\Theta_e$:*

$$\frac{1}{1-\tau_w\left(\theta_w\right)}=\frac{1+\left[1-\bar{g}_w(\theta_w)\right]\frac{1-F_{y_w}(y_w(\theta_w))}{y_w(\theta_w)f_{y_w}(y_w(\theta_w))}\frac{1}{\varepsilon^{y_w}_{1-\tau_w}(\theta_e)}}{\mu}. \tag{OC4}$$

and

$$\frac{1}{1-\tau_e\left(\theta_e\right)} = \left[\begin{array}{c}\frac{1}{\mu(\theta_e)}\left[1+\left[1-\bar{g}_e(\theta_e)\right]\frac{1-F_{y_e}(y_e(\theta_e))}{y_e(\theta_e)f_{y_e}(y_e(\theta_e))}\frac{1}{\bar{\varepsilon}^{y_e}_{1-\tau_e}(\theta_e)}\frac{\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu(\theta_e)-\xi\right]-1}{\frac{1+\varepsilon_e}{\varepsilon_e}\left[1/\varepsilon^{S}_{Q_{ij}}(\theta_e)-\xi\right]-1}\right]+ \\ \varepsilon^{P,cross}_{Q-ij}\left(\theta_e\right)\left[1-\frac{\varepsilon^{\omega}_{l_e}(\theta_e)}{\varepsilon^{\omega}_{L_w}(\theta_e)}\right]\left[1-g_e(\theta_e)-\frac{\left[1-\bar{g}_e(\theta_e)\right]\frac{1-F_{y_e}(y_e(\theta_e))}{y_e(\theta_e)f_{y_e}(y_e(\theta_e))}\frac{1}{\bar{\varepsilon}^{y_e}_{1-\tau_e}(\theta_e)}}{\frac{1+\varepsilon_e}{\varepsilon_e}\left[1/\varepsilon^{S}_{Q_{ij}}(\theta_e)-\xi\right]-1}\right]\end{array}\right] \times\frac{1}{1+\left[\frac{\mu}{\mu(\theta_e)}-1\right]\xi\frac{\varepsilon^{\omega}_{l_e}(\theta_e)}{\varepsilon^{\omega}_{L_w}(\theta_e)}}. \tag{OC5}$$

Proof. See Online Appendix OC.4. ■

Equations (OC4) and (OC5) are statistic-based formulas in the the sense that the statistics on the right side can be observed at each income or profit level. Theorem OC1 shows in addition to the social welfare weights, income distributions, and elasticities of income, the statistic-based optimal tax formulas are related to the markups. When the market is competitive, such that $\mu(\theta_e) = \mu = 1$ and $\varepsilon_{Q_{-ij}}^{P,cross}(\theta_e) = 0$, (OC5) is reduced to the traditional optimal statistic-based tax formula:

$$\frac{1}{1-\tau_e(\theta_e)} = 1 + [1 - \bar{g}_e(\theta_e)] \frac{1 - F_{y_e}(y_e(\theta_e))}{y_e(\theta_e) f_{y_e}(y_e(\theta_e))} \frac{1}{\tilde{\varepsilon}_{1-\tau_e}^{y_e}(\theta_e)}$$

However, in general cases, the statistic-based optimal profit tax in an economy with market power is different from the traditional one (see e.g., Piketty (1997); Diamond (1998); Saez (2001)). Moreover, the statistic-based optimal profit tax formula is quite different from the statistic-based optimal income tax formular, unless the economy is competitive or monopoly competitive.

OC.4 Proof of Theorem OC1

Notice that (OA26) and $\frac{d\ln[1-\tau_e(\theta_e)]}{d\theta_e} \frac{d\ln y_e(\theta_e)}{d\theta_e} = -\frac{T_e''(y_e(\theta_e))y_e(\theta_e)}{1-T_e'(y_e(\theta_e))}$, we have:

$$\begin{aligned}
H(\theta_e) &= \frac{1-F_e(\theta_e)}{f_e(\theta_e)} \left[\frac{d\ln y_e(\theta_e)}{d\theta_e} - \frac{\varepsilon_{Q_{ij}}^S(\theta_e) \frac{d\ln[1-\tau_e(\theta_e)]}{d\theta_e}}{\frac{1+\varepsilon_e}{\varepsilon_e}\left[1 - \xi\varepsilon_{Q_{ij}}^S(\theta_e)\right] - \varepsilon_{Q_{ij}}^S(\theta_e)} \right] \\
&= \frac{1-F_e(\theta_e)}{f_e(\theta_e)} \frac{d\ln y_e(\theta_e)}{d\theta_e} \frac{\frac{1+\varepsilon_e}{\varepsilon_e}\left[\frac{1}{\varepsilon_{Q_{ij}}^S(\theta_e)} - \xi\right] - \left[1 - \frac{T_e''(y_e(\theta_e))y_e(\theta_e)}{1-T_e'(y_e(\theta_e))}\right]}{\frac{1+\varepsilon_e}{\varepsilon_e}\left[\frac{1}{\varepsilon_{Q_{ij}}^S(\theta_e)} - \xi\right] - 1} \\
&= \frac{1-F_e(\theta_e)}{f_e(\theta_e)} \frac{d\ln y_e(\theta_e)}{d\theta_e} \frac{1}{\tilde{\varepsilon}_{1-\tau_e}^{y_e}(\theta_e)} \frac{1}{\frac{1+\varepsilon_e}{\varepsilon_e}\left[\frac{1}{\varepsilon_{Q_{ij}}^S(\theta_e)} - \xi\right] - 1}.
\end{aligned}$$

Therefore, (36) is equivalent to:

$$\begin{aligned}
\frac{1}{1-\tau_e(\theta_e)} = & \frac{\frac{1}{\mu(\theta_e)}\left[1 + [1-\bar{g}_e(\theta_e)] \frac{1-F_{y_e}(y_e(\theta_e))}{y_e(\theta_e)f_{y_e}(y_e(\theta_e))} \frac{1}{\tilde{\varepsilon}_{1-\tau_e}^{y_e}(\theta_e)} \frac{\frac{1+\varepsilon_e}{\varepsilon_e}[\mu(\theta_e)-\xi]-1}{\frac{1+\varepsilon_e}{\varepsilon_e}\left[1/\varepsilon_{Q_{ij}}^S(\theta_e)-\xi\right]-1}\right]}{1 + RE(\theta_e)\,\xi \frac{\varepsilon_{l_e}^{\omega}(\theta_e)}{\varepsilon_{L_w}^{\omega}(\theta_e)}} \\
& + \frac{\left[1 - \xi\frac{\varepsilon_{l_e}^{\omega}(\theta_e)}{\varepsilon_{L_w}^{\omega}(\theta_e)}\right] \varepsilon_{Q_{-ij}}^{P,cross}(\theta_e) \left[\begin{array}{c} [1-g_e(\theta_e)] - \\ [1-\bar{g}_e(\theta_e)] \frac{\frac{1-F_{y_e}(y_e(\theta_e))}{y_e(\theta_e)f_{y_e}(y_e(\theta_e))} \frac{1}{\tilde{\varepsilon}_{1-\tau_e}^{y_e}(\theta_e)}}{\frac{1+\varepsilon_e}{\varepsilon_e}\left[1/\varepsilon_{Q_{ij}}^S(\theta_e)-\xi\right]-1} \end{array} \right]}{1 + RE(\theta_e)\,\xi\frac{\varepsilon_{l_e}^{\omega}(\theta_e)}{\varepsilon_{L_w}^{\omega}(\theta_e)}},
\end{aligned}$$

where we use $\varepsilon_{l_e}^{Q_{ij}}(\theta_e) = 1$ and $\varepsilon_{L_w}^{Q_{ij}}(\theta_e) = \xi$. Notice that $\frac{\frac{1+\varepsilon_e}{\varepsilon_e}[\mu(\theta_e)-\xi]-1}{\frac{1+\varepsilon_e}{\varepsilon_e}\left[1/\varepsilon_{Q_{ij}}^S(\theta_e)-\xi\right]-1} = 1$ when $I = 1$, and $\frac{\frac{1+\varepsilon_e}{\varepsilon_e}[\mu(\theta_e)-\xi]-1}{\frac{1+\varepsilon_e}{\varepsilon_e}\left[1/\varepsilon_{Q_{ij}}^S(\theta_e)-\xi\right]-1} < 1$ when $I > 1$. When tax is linear $\frac{1}{\tilde{\varepsilon}_{1-\tau_e}^{y_e}(\theta_e)} = \frac{1}{\frac{1+\varepsilon_e}{\varepsilon_e}\left[1/\varepsilon_{Q_{ij}}^S(\theta_e)-\xi\right]-1}$.

Last, (OC4) is equivalent to (35) bcause by the definition of $\varepsilon_{1-\tau_w}^{y_w}(\theta_e)$, one has $\frac{1-F_{y_w}(y_w(\theta_w))}{y_w(\theta_w)f_{y_w}(y_w(\theta_w))} \frac{1}{\varepsilon_{1-\tau_w}^{y_w}(\theta_e)} =$

$\frac{1-F_w(\theta_w)}{f_w(\theta_w)} \frac{x_w'(\theta_w)}{x_w(\theta_w)} \frac{1+\varepsilon_w}{\varepsilon_w}$.■

OC.5 Proof of Theorem 2

OC.5.1 Non-linear Sales Taxes

Before prove Theorem 2, we should present the second best problem formally. In our benchmark model, we consider an environment with uniform linear sales tax, which restricts $\tau_s(\theta_e)$ to be constant. In this section, we remove this policy constraint and allow for non-linear sales tax as considered by Ales et al. (2017). To do this, we allow the planner to contract with entrepreneurs on sales income $S(\theta_e) \equiv P(\theta_e) Q_{ij}(\theta_e)$ in addition to $y_e(\theta_e)$. An entrepreneur reporting θ_e' should obtain $y_e(\theta_e')$ in profit, $S(\theta_e')$ in sales income, and receive $c_e(\theta_e')$ in after-tax profit. Thus, a θ_e-type entrepreneur reporting θ_e' should choose L_w and l_e to satisfy the following two promise-keeping constraints:

$$P_{ij}\left(Q_{ij}, Q_{-ij}(\theta_e), \theta_e\right) Q_{ij} = S\left(\theta_e'\right) \quad \text{and} \quad P_{ij}\left(Q_{ij}, Q_{-ij}(\theta_e), \theta_e\right) Q_{ij} - W L_w = y_e\left(\theta_e'\right).$$

The two promise-keeping constraints determine the combination of L_w and l_e that are needed to complete the tasks of sales income and profit. Denote $L_w(\theta_e'|\theta_e)$ and $l_e(\theta_e'|\theta_e)$ as the labor input and effort needed to complete the tasks, respectively. Denote $Q_{ij}(x_e(\theta_e) l_e, L_w)$ as the firm-level production function. Combining these two constraints we immediately have

$$L_w\left(\theta_e'|\theta_e\right) = \frac{S\left(\theta_e'\right) - y_e\left(\theta_e'\right)}{W},$$

and $l_e(\theta_e'|\theta_e)$ satisfies

$$P\left(Q_{ij}\left(x_e(\theta_e) l_e\left(\theta_e'|\theta_e\right), \frac{S\left(\theta_e'\right) - y_e\left(\theta_e'\right)}{W}\right), Q_{-ij}(\theta_e), \theta_e\right) Q_{ij}\left(x_e(\theta_e) l_e\left(\theta_e'|\theta_e\right), \frac{S\left(\theta_e'\right) - y_e\left(\theta_e'\right)}{W}\right) = S\left(\theta_e'\right).$$

Two things worth noting here. First, by enforcing tasks of sales income and profit, the planner can directly control the amount of labor inputs. That is to say, $L_w(\theta_e'|\theta_e)$ is independent of θ_e. Second, notice that $\frac{\partial\left[P\left(Q_{ij}, Q_{-ij}(\theta_e), \theta_e\right) Q_{ij}\right]}{\partial Q_{ij}} = P\left(Q_{ij}, Q_{-ij}(\theta_e), \theta_e\right)\left[1 + \varepsilon_{Q_{ij}}^{P,own}(\theta_e)\right]$ and $\varepsilon_{Q_{ij}}^{P,own}(\theta_e) > -1$. As long as Q_{ij} strictly increases in $l_e(\theta_e'|\theta_e)$ with $P\left(Q_{ij}(0,\cdot), Q_{-ij}(\theta_e), \theta_e\right) Q_{ij}(0,\cdot) = 0$, there exists a unique solution $l_e(\theta_e'|\theta_e)$ for any $S(\theta_e') \geq 0$.

Therefore, under our setup, we can reformulate the entrepreneur's problem as:

$$V_e(\theta_e) \equiv \max_{\theta_e'} \; c_e\left(\theta_e'\right) - \phi_e\left(l_e\left(\theta_e'|\theta_e\right)\right). \tag{OC6}$$

Solving the above problem, as in the benchmark model, we have

$$\frac{\partial l_e\left(\theta_e'|\theta_e\right)}{\partial \theta_e} = -\frac{\frac{\partial P\left(Q_{ij}, \theta_e\right)}{\partial \theta_e} Q_{ij} + P\left(Q_{ij}, Q_{-ij}(\theta_e), \theta_e\right)\left[1 + \varepsilon_{Q_{ij}}^{P,own}(\theta_e)\right] \frac{\partial Q_{ij}}{\partial x_e(\theta_e)} x_e'(\theta_e)}{P\left(Q_{ij}, Q_{-ij}(\theta_e), \theta_e\right)\left[1 + \varepsilon_{Q_{ij}}^{P,own}(\theta_e)\right] \frac{\partial Q_{ij}}{\partial l_e\left(\theta_e'|\theta_e\right)}} \tag{OC7}$$

and

$$V_e'(\theta_e) = \frac{\phi_e'(l_e(\theta_e))\, Q_{ij}(\theta_e)}{\frac{\partial Q_{ij}(\theta_e)}{\partial l_e(\theta_e)}} \left[\frac{\frac{d \ln P_{ij}\left(Q_{ij}, Q_{-ij}(\theta_e), \theta_e\right)}{d\theta_e}|_{Q_{ij}=Q_{ij}(\theta_e)}}{1+\varepsilon_{Q_{ij}}^{P,own}(\theta_e)} + \frac{\partial \ln Q_{ij}(\theta_e)}{\partial \ln x_e(\theta_e)} \frac{x_e'(\theta_e)}{x_e(\theta_e)} \right] \tag{OC8}$$

which is equivalent to the benchmark incentive-compatible constraint (29), because in the benchmark model $\mu(\theta) = \frac{1}{1+\varepsilon_{Q_{ij}}^{P,own}(\theta_e)}$, $\frac{Q_{ij}(\theta_e)}{\frac{\partial Q_{ij}(\theta_e)}{\partial l_e(\theta_e)}} = l_e(\theta_e)$ and $\frac{\partial \ln Q_{ij}(\theta_e)}{\partial x_e(\theta_e)} x_e'(\theta_e) = \frac{x_e'(\theta_e)}{x_e(\theta_e)}$.

The worker's problem remains the same as before. Therefore, all incentive-compatible allocations satisfying (26), (29) and the resource constraints are feasible. The planner's problem is similar to the one in the benchmark model, except that the policy constraint $\frac{d \ln \varpi(\theta_e)}{d\theta_e} = 0$ is now relaxed.

OC.5.2 Lagrangian Problem

We prove Theorem 2 following the Lagrangian problem presented in Appendix B.2.1. Note that the expression of markup (19) is now generalized to be $\mu(\theta_e) = \frac{P(\theta_e)\frac{\partial Q_{ij}(\theta_e)}{\partial L_w(\theta_e)}[1-\tau_s(\theta_e)]}{W}$. By the definition of $\varpi(\theta_e)$ (see e.g., (30)), we have $\varpi(\theta_e) = P(\theta_e)\left[1+\varepsilon_{Q_{ij}}^{P,own}(\theta_e)\right]\frac{\partial Q_{ij}(\theta_e)}{\partial L_w(\theta_e)}$. Notice that $1-\tau_s(\theta_e) = \frac{W}{\varpi(\theta_e)}$, as in the benchmark model, we have $\mu(\theta_e) = \frac{1}{1+\varepsilon_{Q_{ij}}^{P,own}(\theta_e)}$.

(i) When the uniform restriction on $\varpi(\theta_e, \theta_e l_e(\theta_e), L_w(\theta_e), Q)$ is loosened and $\varphi(\theta_e) = 0$. In this case, according to the expression of $\frac{\partial \pounds}{\partial L_w(\theta_e)}$ (e.g., (A36)), we have

$$\begin{aligned} P(\theta_e)\frac{\partial Q_{ij}(\theta_e)}{\partial L_w(\theta_e)} &= \frac{\lambda'}{\lambda} - \frac{\kappa'(\theta_e)}{\lambda L_w(\theta_e) N_e f_e(\theta_e)} \frac{\partial \ln Q_{ij}(\theta_e)}{\partial \ln L_w(\theta_e)} \\ &= \frac{\lambda'}{\lambda} - \frac{\kappa'(\theta_e)\xi}{\lambda L_w(\theta_e) N_e f_e(\theta_e)}. \end{aligned} \tag{OC9}$$

Dividing both sides by $\frac{\varepsilon_{L_w}^{\varpi}(\theta_e)}{L_w(\theta_e) N f_e(\theta_e)}$ and integrating across θ_e gives

$$\int_{\theta_e} P(\theta_e) Q_{ij}(\theta_e)\, \xi \frac{N_e f_e(\theta_e)}{\varepsilon_{L_w}^{\varpi}(\theta_e)} d\theta_e = \frac{\lambda'}{\lambda}\int_{\theta_e} \frac{L_w(\theta_e) N_e f_e(\theta_e)}{\varepsilon_{L_w}^{\varpi}(\theta_e)} d\theta_e - \int_{\theta_e} \frac{\kappa'(\theta_e)}{\lambda \varepsilon_{L_w}^{\varpi}(\theta_e)} \xi d\theta_e,$$

Using $\kappa(\underline{\theta}_e) = \kappa(\overline{\theta}_e) = 0$ and integration by parts, we have

$$\begin{aligned} \frac{\lambda'}{\lambda W} &= \frac{\xi \int_{\theta_e} P(\theta_e) Q_{ij}(\theta_e) N_e f_e(\theta_e)\, d\theta_e}{W \int_{\theta_e} L_w(\theta_e) N_e f_e(\theta_e)\, d\theta_e} \\ &= \frac{\xi \int_{\theta_e} \frac{P(\theta_e) Q_{ij}(\theta_e)}{W L_w(\theta_e)} W L_w(\theta_e) N_e f_e(\theta_e)\, d\theta_e}{\int_{\theta_e} W L_w(\theta_e) N_e f_e(\theta_e)\, d\theta_e} \\ &= \frac{\xi \int_{\theta_e} \frac{\mu(\theta_e)}{\xi[1-\tau_s^E(\theta_e)]} W L_w(\theta_e) N f_e(\theta_e)\, d\theta_e}{\int_{\theta_e} W L_w(\theta_e) N_e f_e(\theta_e)\, d\theta_e} \\ &= \mu^*. \end{aligned} \tag{OC10}$$

where the third equation is derived from the generalized definition of markup (i.e., $\mu(\theta_e) = \frac{P(\theta_e)\frac{\partial Q_{ij}(\theta_e)}{\partial L_w(\theta_e)}[1-\tau_s(\theta_e)]}{W}$).

By the definition of tax wedges (see e.g., (31)), we can substitute $P(\theta_e)\frac{\partial Q_{ij}(\theta_e)}{\partial L_w(\theta_e)}$ in (OC9) with $\frac{W\mu(\theta_e)}{1-\tau_s^E(\theta_e)}$:

$$\begin{aligned}
\frac{1}{1-\tau_s^E(\theta_e)} &= \frac{\lambda'}{\lambda W\mu(\theta_e)} - \frac{\kappa'(\theta_e)\xi}{\lambda L_w(\theta_e) N_e f_e(\theta_e) W\mu(\theta_e)} \\
&= \frac{\mu^*}{\mu(\theta_e)} - \frac{\kappa'(\theta_e)\xi}{\lambda L_w(\theta_e) N_e f_e(\theta_e) W\mu(\theta_e)} \\
&= \frac{\mu^*}{\mu(\theta_e)} + \frac{P(\theta_e) Q_{ij}(\theta_e)\left[1-\tau_s^E(\theta_e)\right]\xi\left[1-\tau_e^E(\theta_e)\right]\varepsilon_{Q_{-ij}}^{P,cross}(\theta_e)\begin{bmatrix}\psi_e(\theta_e)\frac{1+\varepsilon_e}{\varepsilon_e}\frac{l_e'(\theta_e)}{l_e(\theta_e)}+\psi_e'(\theta_e) \\ +\psi_e(\theta_e)\frac{d\ln\left[\mu(\theta_e)\varepsilon_{Q_{-ij}}^{P,cross}(\theta_e)\right]}{d\theta_e}\end{bmatrix}}{L_w(\theta_e) W\mu(\theta_e)\lambda N_e f_e(\theta_e)} \\
&= \frac{\mu^*}{\mu(\theta_e)} + \left[1-\tau_e^E(\theta_e)\right]\varepsilon_{Q_{-ij}}^{P,cross}(\theta_e)\begin{bmatrix}\left[1-g_e(\theta_e)\right]- \\ \frac{[1-\bar{g}_e(\theta_e)][1-F_e(\theta_e)]}{f_e(\theta_e)}\left[\frac{1+\varepsilon_e}{\varepsilon_e}\frac{l_e'(\theta_e)}{l_e(\theta_e)}+\frac{d\ln\left[\mu(\theta_e)\varepsilon_{Q_{-ij}}^{P,cross}(\theta_e)\right]}{d\theta_e}\right]\end{bmatrix},
\end{aligned} \tag{OC11}$$

where the third equation is derived by the modified (A45):

$$\begin{aligned}
\kappa'(\theta_e) &= -\frac{d\left[\psi_e(\theta_e)\phi_e'(l_e(\theta_e))l_e(\theta_e)\widehat{\mu}(\theta_e)\varepsilon_{Q_{-ij}}^{P,cross}(\theta_e)\right]}{d\theta_e} \\
&= -\begin{bmatrix}\psi_e'(\theta_e)\phi_e'(l_e(\theta_e))l_e(\theta_e)\widehat{\mu}(\theta_e)\varepsilon_{Q_{-ij}}^{P,cross}(\theta_e)+ \\ \psi_e(\theta_e)\phi_e'(l_e(\theta_e))\frac{1+\varepsilon_e}{\varepsilon_e}l_e'(\theta_e)\widehat{\mu}(\theta_e)\varepsilon_{Q_{-ij}}^{P,cross}(\theta_e)+ \\ \psi_e(\theta_e)\phi_e'(l_e(\theta_e))l_e(\theta_e)\frac{d\ln\left[\widehat{\mu}(\theta_e)\varepsilon_{Q_{-ij}}^{P,cross}(\theta_e)\right]}{d\theta_e}\end{bmatrix} \\
&= -\phi_e'(l_e(\theta_e))l_e(\theta_e)\widehat{\mu}(\theta_e)\varepsilon_{Q_{-ij}}^{P,cross}(\theta_e)\begin{bmatrix}\psi_e(\theta_e)\frac{1+\varepsilon_e}{\varepsilon_e}\frac{l_e'(\theta_e)}{l_e(\theta_e)}+ \\ \psi_e'(\theta_e)+\psi_e(\theta_e)\frac{d\ln\left[\widehat{\mu}(\theta_e)\varepsilon_{Q_{-ij}}^{P,cross}(\theta_e)\right]}{d\theta_e}\end{bmatrix}
\end{aligned}$$

and the tax wedge, $1-\tau_e^E(\theta_e)=\frac{\phi_e'(l_e(\theta_e))}{\frac{P(\theta_e)}{\mu(\theta_e)}\frac{\partial Q_{ij}(\theta_e)}{\partial l_e(\theta_e)}[1-\tau_s^E(\theta_e)]}$ (see e.g., (31)). These two equations implies

$$\kappa'(\theta_e) = -P(\theta_e)Q_{ij}(\theta_e)\left[1-\tau_e^E(\theta_e)\right]\left[1-\tau_s^E(\theta_e)\right]\varepsilon_{Q_{-ij}}^{P,cross}(\theta_e)\begin{bmatrix}\psi_e(\theta_e)\frac{1+\varepsilon_e}{\varepsilon_e}\frac{l_e'(\theta_e)}{l_e(\theta_e)}+ \\ \psi_e'(\theta_e)+\psi_e(\theta_e)\frac{d\ln\left[\mu(\theta_e)\varepsilon_{Q_{-ij}}^{P,cross}(\theta_e)\right]}{d\theta_e}\end{bmatrix}.$$

The last equation of (OC11), i.e.,

$$\frac{\tau_s^E(\theta_e)}{1-\tau_s^E(\theta_e)} = \overbrace{\left[\frac{\mu^*}{\mu(\theta_e)}-1\right]}^{\widetilde{RE}(\theta_e)} + \overbrace{\left[1-\tau_e^E(\theta_e)\right]\varepsilon_{Q_{-ij}}^{P,cross}(\theta_e)\begin{bmatrix}\left[1-g_e(\theta_e)\right]-\frac{[1-\bar{g}_e(\theta_e)][1-F_e(\theta_e)]}{f_e(\theta_e)} \\ \times\left[\frac{1+\varepsilon_e}{\varepsilon_e}\frac{l_e'(\theta_e)}{l_e(\theta_e)}+\frac{d\ln\left[\mu(\theta_e)\varepsilon_{Q_{-ij}}^{P,cross}(\theta_e)\right]}{d\theta_e}\right]\end{bmatrix}}^{\widetilde{IRE}(\theta_e)} \tag{OC12}$$

is derived by (A41), (A42) and $1-\tau_s^E(\theta_e)=\frac{W\mu(\theta_e)}{P(\theta_e)\frac{\partial Q_{ij}(\theta_e)}{\partial L_w(\theta_e)}}=\frac{WL_w(\theta_e)\mu(\theta_e)}{\xi P(\theta_e)Q_{ij}(\theta_e)}$.

(ii) According to the expression of $\frac{\partial \pounds}{\partial l_w(\theta_w)}$ (e.g., (A35)), we have

$$\frac{1}{\frac{\phi_w'(l_w(\theta_w))}{x_w(\theta_w)}} = \frac{\lambda}{\lambda'}\left[1 - \frac{x_w'(\theta_w)}{x_w(\theta_w)}\frac{\psi_w(\theta_w)}{\lambda N_w f_w(\theta_w)}\frac{1+\varepsilon_w}{\varepsilon_w}\right].$$

Substituting $\frac{\phi_w'(l_w(\theta_w))}{x_w(\theta_w)}$ by $\left[1-\tau_w^E(\theta_w)\right]W$ gives:

$$\frac{1}{1-\tau_w^E(\theta_w)} = \frac{1}{\mu^*}\left[1 + [1-\bar{g}_w(\theta_w)]\frac{1-F_w(\theta_w)}{f_w(\theta_w)}\frac{x_w'(\theta_w)}{x_w(\theta_w)}\frac{1+\varepsilon_w}{\varepsilon_w}\right]. \tag{OC13}$$

(iii) According to the expression of $\frac{\partial \pounds}{\partial l_e(\theta_e)}$ (e.g., (A37)), we have:

$$\begin{aligned}\frac{\widetilde{\tau}_e(\theta_e)}{1-\widetilde{\tau}_e(\theta_e)} &= [1-\bar{g}_e(\theta_e)]\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu(\theta_e)\frac{\partial \ln P(Q(\theta_e),\theta_e)}{\partial\theta_e} + \frac{x_e'(\theta_e)}{x_e(\theta_e)}\right] \\ &\quad - \frac{1}{1-\widetilde{\tau}_e(\theta_e)}\frac{WL_w(\theta_e)}{P(\theta_e)Q_{ij}(\theta_e)}\frac{1}{\xi}\frac{\kappa'(\theta_e)\xi}{\lambda N_e f_e(\theta_e) WL_w(\theta_e)}.\end{aligned}$$

Reminber that $1-\widetilde{\tau}_e(\theta_e) = \frac{\phi_e'(l_e(\theta_e))}{P(\theta_e)\frac{\partial Q_{ij}(\theta_e)}{\partial l_e(\theta_e)}} = \frac{\left[1-\tau_e^E(\theta_e)\right]\left[1-\tau_s^E(\theta_e)\right]}{\mu(\theta_e)}$.

Using $\frac{\mu(\theta_e)}{1-\tau_s^E(\theta_e)} = \mu^* - \frac{\kappa'(\theta_e)\xi}{\lambda WL_w(\theta_e)N_e f_e(\theta_e)}$ to substitute $\frac{\kappa'(\theta_e)\xi}{\lambda WL_w(\theta_e)N_e f_e(\theta_e)}$ and $\frac{\xi\left[1-\tau_s^E(\theta_e)\right]}{\mu(\theta_e)}$ to substitute $\frac{WL_w(\theta_e)}{P(\theta_e)Q_{ij}(\theta_e)}$, we have:

$$\begin{aligned}\frac{1}{1-\widetilde{\tau}_e(\theta_e)} &= 1 + [1-\bar{g}_e(\theta_e)]\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu(\theta_e)\frac{\partial \ln P(Q(\theta_e),\theta_e)}{\partial\theta_e} + \frac{x_e'(\theta_e)}{x_e(\theta_e)}\right] \\ &\quad + \frac{1}{1-\widetilde{\tau}_e(\theta_e)}\left[1 - \mu^*\frac{1-\tau_s^E(\theta_e)}{\mu(\theta_e)}\right].\end{aligned}$$

Lastly, substitute $1-\widetilde{\tau}_e(\theta_e)$ by $\frac{\left[1-\tau_e^E(\theta_e)\right]\left[1-\tau_s^E(\theta_e)\right]}{\mu(\theta_e)}$ and rearrange:

$$\frac{\tau_e^E(\theta_e)}{1-\tau_e^E(\theta_e)} = \frac{1-\mu^* + [1-\bar{g}_e(\theta_e)]\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu(\theta_e)\frac{\partial \ln P(Q(\theta_e),\theta_e)}{\partial\theta_e} + \frac{x_e'(\theta_e)}{x_e(\theta_e)}\right]}{\mu^*}. \tag{OC14}$$

OC.5.3 Part One: An Explicit Expression

According to the definitions of tax wedges, we have

$$\frac{\xi\left[1-\tau_s^E(\theta_e)\right]}{\mu(\theta_e)}P(\theta_e)Q(\theta_e) = WL_w(\theta_e) \tag{OC15}$$

and

$$\frac{P(\theta_e)Q_{ij}(\theta_e)\left[1-\tau_s^E(\theta_e)\right]}{\mu(\theta_e)}\left[1-\tau_e^E(\theta_e)\right] = l_e(\theta_e)^{1+\frac{1}{\varepsilon_e}}. \tag{OC16}$$

Substituting $P(\theta_e)$ and $Q(\theta_e)$ in (OC15) by (A1) and $Q_{ij}(\theta_e) = x_e(\theta_e)\, l_e(\theta_e)\, L_w(\theta_e)^{\xi}$, we have:

$$\left(\left[1-\tau_s^E(\theta_e)\right]\frac{\chi(\theta_e)\, x_e(\theta_e)^{1-\frac{1}{\sigma}}}{\mu(\theta_e)}\right)\frac{\xi A^{\frac{\sigma-1}{\sigma}}Q^{\frac{1}{\sigma}}}{N_e^{\frac{1}{\sigma}}W} l_e(\theta_e)^{1-\frac{1}{\sigma}} L_w(\theta_e)^{\xi\frac{\sigma-1}{\sigma}} = L_w(\theta_e),$$

Rearrange the above equation:

$$L_w(\theta_e) = \left(\frac{X_e(\theta_e)\left[1-\tau_s^E(\theta_e)\right]}{\mu(\theta_e)}\right)^{\frac{1}{1-\xi\frac{\sigma-1}{\sigma}}} l_e(\theta_e)^{\frac{1-\frac{1}{\sigma}}{1-\xi\frac{\sigma-1}{\sigma}}}\left(\frac{\xi A^{\frac{\sigma-1}{\sigma}}Q^{\frac{1}{\sigma}}}{WN_e^{\frac{1}{\sigma}}}\right)^{\frac{1}{1-\xi\frac{\sigma-1}{\sigma}}}. \tag{OC17}$$

On the other hand, a combination of the two first-order conditions delivers:

$$\frac{WL_w(\theta_e)}{\xi}\left[1-\tau_e^E(\theta_e)\right] = l_e(\theta_e)^{1+\frac{1}{\varepsilon_e}}.$$

The above two equations imply:

$$L_w(\theta_e) = \left(\frac{X_e(\theta_e)\left[1-\tau_s^E(\theta_e)\right]}{\mu(\theta_e)}\right)^{\frac{1}{1-\xi\frac{\sigma-1}{\sigma}}}\left[\frac{WL_w(\theta_e)}{\xi}\left[1-\tau_e^E(\theta_e)\right]\right]^{\frac{1-\frac{1}{\sigma}}{1-\xi\frac{\sigma-1}{\sigma}}\frac{\varepsilon_e}{1+\varepsilon_e}}\left(\frac{\xi A^{\frac{\sigma-1}{\sigma}}Q^{\frac{1}{\sigma}}}{WN_e^{\frac{1}{\sigma}}}\right)^{\frac{1}{1-\xi\frac{\sigma-1}{\sigma}}}$$

i.e,

$$L_w(\theta_e) = \left[1-\tau_s^E(\theta_e)\right]^{\frac{1+\frac{\sigma-1}{\sigma}\frac{\varepsilon_e}{1+\varepsilon_e}}{1-\frac{\sigma-1}{\sigma}\left(\frac{\varepsilon_e}{1+\varepsilon_e}+\xi\right)}}\left(\frac{X_e(\theta_e)}{\mu(\theta_e)}\right)^{\frac{1}{1-\frac{\sigma-1}{\sigma}\left(\frac{\varepsilon_e}{1+\varepsilon_e}+\xi\right)}}\left[\frac{W}{\xi}\right]^{\frac{\frac{\sigma-1}{\sigma}\frac{\varepsilon_e}{1+\varepsilon_e}}{1-\frac{\sigma-1}{\sigma}\left(\frac{\varepsilon_e}{1+\varepsilon_e}+\xi\right)}}\left(\frac{\xi A^{\frac{\sigma-1}{\sigma}}Q^{\frac{1}{\sigma}}}{WN_e^{\frac{1}{\sigma}}}\right)^{\frac{\frac{1}{1-\xi\frac{\sigma-1}{\sigma}}}{1-\frac{1-\frac{1}{\sigma}}{1-\xi\frac{\sigma-1}{\sigma}}\frac{\varepsilon_e}{1+\varepsilon_e}}}. \tag{OC18}$$

Combining the above two equations, we can solve for $L_w(\theta_e)$ and $l_e(\theta_e)$ and derive:

$$\frac{d\ln L_w(\theta_e)}{d\theta_e} = \frac{\frac{d}{d\theta_e}\left[\ln\frac{X_e(\theta_e)\left[1-\tau_s^E(\theta_e)\right]}{\mu(\theta_e)}\right]}{1-\xi\frac{\sigma-1}{\sigma}} + \frac{\frac{\sigma-1}{\sigma}}{1-\xi\frac{\sigma-1}{\sigma}}\frac{l_e'(\theta_e)}{l_e(\theta_e)} \tag{OC19}$$

and

$$\frac{d\ln l_e(\theta_e)}{d\theta_e} = \frac{\frac{d}{d\theta_e}\left[\ln\frac{X_e(\theta_e)\left[1-\tau_s^E(\theta_e)\right]}{\mu(\theta_e)}\right] + \left[1-\xi\frac{\sigma-1}{\sigma}\right]\frac{d\ln\left[1-\tau_e^E(\theta_e)\right]}{d\theta_e}}{\frac{1+\varepsilon_e}{\varepsilon_e} - \frac{\sigma-1}{\sigma}\left(1+\frac{1+\varepsilon_e}{\varepsilon_e}\xi\right)}. \tag{OC20}$$

Comparing the above two equations to (OA20) and (OA22), one can see that they are the same except that now $X_e(\theta_e)$ is replaced by $X_e(\theta_e)\left[1-\tau_s^E(\theta_e)\right]$. In addition, we have:

$$\begin{aligned}
& \mu(\theta_e)\frac{d\ln P_{ij}\left(Q_{ij},Q_{-ij}(\theta_e),\theta_e\right)}{d\theta_e}|_{Q_{ij}=Q_{ij}(\theta_e)}+\frac{x_e'(\theta_e)}{x_e(\theta_e)} \\
= \;& \mu(\theta_e)\frac{\chi'(\theta_e)}{\chi(\theta_e)}+\frac{x_e'(\theta_e)}{x_e(\theta_e)}+\mu(\theta_e)\varepsilon_{Q_{-ij}}^{P,cross}(\theta_e)\frac{d\ln Q_{ij}(\theta_e)}{d\theta_e} \\
= \;& \mu(\theta_e)\frac{\chi'(\theta_e)}{\chi(\theta_e)}+\frac{x_e'(\theta_e)}{x_e(\theta_e)}+\mu(\theta_e)\varepsilon_{Q_{-ij}}^{P,cross}(\theta_e)\left[\frac{x_e'(\theta_e)}{x_e(\theta_e)}+\frac{l_e'(\theta_e)}{l_e(\theta_e)}+\xi\frac{L_w'(\theta_e)}{L_w(\theta_e)}\right] \\
= \;& \mu(\theta_e)\frac{\chi'(\theta_e)}{\chi(\theta_e)}+\left[1+\mu(\theta_e)\varepsilon_{Q_{-ij}}^{P,cross}(\theta_e)\right]\frac{x_e'(\theta_e)}{x_e(\theta_e)}+\frac{\mu(\theta_e)\varepsilon_{Q_{-ij}}^{P,cross}(\theta_e)}{1-\xi\frac{\sigma-1}{\sigma}}\frac{d\ln l_e(\theta_e)}{d\theta_e} \\
& +\frac{\xi\mu(\theta_e)\varepsilon_{Q_{-ij}}^{P,cross}(\theta_e)}{1-\xi\frac{\sigma-1}{\sigma}}\frac{d\left[\ln\frac{X_e(\theta_e)}{\mu(\theta_e)}+\ln\left[1-\tau_s^E(\theta_e)\right]\right]}{d\theta_e},
\end{aligned}$$

where the third equation is derived by (OC19) and

$$\begin{aligned}
& \mu(\theta_e)\frac{\chi'(\theta_e)}{\chi(\theta_e)}+\left[1+\mu(\theta_e)\varepsilon_{Q_{-ij}}^{P,cross}(\theta_e)\right]\frac{x_e'(\theta_e)}{x_e(\theta_e)} \\
= \;& \mu(\theta_e)\left[\frac{\chi'(\theta_e)}{\chi(\theta_e)}+\frac{\sigma-1}{\sigma}\frac{x_e'(\theta_e)}{x_e(\theta_e)}\right] \\
= \;& \mu(\theta_e)\frac{d\ln\frac{X_e(\theta_e)}{\mu(\theta_e)}}{d\theta_e}+\frac{d\mu(\theta_e)}{d\theta_e}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
& \mu(\theta_e)\frac{d\ln P_{ij}\left(Q_{ij},Q_{-ij}(\theta_e),\theta_e\right)}{d\theta_e}|_{Q_{ij}=Q_{ij}(\theta_e)}+\frac{x_e'(\theta_e)}{x_e(\theta_e)} \\
= \;& \frac{\mu(\theta_e)\varepsilon_{Q_{-ij}}^{P,cross}(\theta_e)}{1-\xi\frac{\sigma-1}{\sigma}}\frac{d\ln l_e(\theta_e)}{d\theta_e}+\frac{\mu(\theta_e)-\xi}{1-\xi\frac{\sigma-1}{\sigma}}\frac{d\ln\left[X_e(\theta_e)/\mu(\theta_e)\right]}{d\theta_e} \\
& +\frac{\xi\mu(\theta_e)\varepsilon_{Q_{-ij}}^{P,cross}(\theta_e)}{1-\xi\frac{\sigma-1}{\sigma}}\frac{d\ln\left[1-\tau_s^E(\theta_e)\right]}{d\theta_e}+\frac{d\mu(\theta_e)}{d\theta_e} \\
= \;& \left[(\mu(\theta_e)-\xi)\frac{1+\varepsilon_e}{\varepsilon_e}-1\right]\frac{\frac{d\ln\left[X_e(\theta_e)/\mu(\theta_e)\right]}{d\theta_e}}{\frac{1+\varepsilon_e}{\varepsilon_e}-\frac{\sigma-1}{\sigma}\left(1+\frac{1+\varepsilon_e}{\varepsilon_e}\xi\right)} \\
& +\mu(\theta_e)\varepsilon_{Q_{-ij}}^{P,cross}(\theta_e)\frac{\left(1+\xi\frac{1+\varepsilon_e}{\varepsilon_e}\right)\frac{d\ln\left[1-\tau_s^E(\theta_e)\right]}{d\theta_e}+\frac{d\ln\left[1-\tau_e^E(\theta_e)\right]}{d\theta_e}}{\frac{1+\varepsilon_e}{\varepsilon_e}-\frac{\sigma-1}{\sigma}\left(1+\frac{1+\varepsilon_e}{\varepsilon_e}\xi\right)}+\frac{d\mu(\theta_e)}{d\theta_e},
\end{aligned}$$

where the second equation is derived by (OC20). In addition, we have

$$
\begin{aligned}
& \frac{1-F_e(\theta_e)}{f_e(\theta_e)} \frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu\left(\theta_e\right) \frac{\partial \ln P\left(Q\left(\theta_e\right), \theta_e\right)}{\partial \theta_e}+\frac{x_e^{\prime}\left(\theta_e\right)}{x_e\left(\theta_e\right)}\right] \\
= & \left[\left(\mu\left(\theta_e\right)-\xi\right) \frac{1+\varepsilon_e}{\varepsilon_e}-1\right] H(\theta_e)+\frac{1-F_e(\theta_e)}{f_e\left(\theta_e\right)} \frac{d \ln \left[\mu\left(\theta_e\right)-\xi\right]}{d \theta_e} \\
& +\mu\left(\theta_e\right) \varepsilon_{Q-ij}^{P, cross}\left(\theta_e\right) \frac{1-F_e(\theta_e)}{f_e(\theta_e)} \frac{1+\varepsilon_e}{\varepsilon_e} \frac{\left(1+\xi \frac{1+\varepsilon_e}{\varepsilon_e}\right) \frac{d \ln \left[1-\tau_s^E\left(\theta_e\right)\right]}{d \theta_e}+\frac{d \ln \left[1-\tau_e^E\left(\theta_e\right)\right]}{d \theta_e}}{\frac{1+\varepsilon_e}{\varepsilon_e}\left(1-\frac{\sigma-1}{\sigma} \xi\right)-\frac{\sigma-1}{\sigma}} .
\end{aligned}
$$

Substituting $\frac{1-F_e(\theta_e)}{f_e(\theta_e)} \frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu\left(\theta_e\right) \frac{\partial \ln P\left(Q(\theta_e), \theta_e\right)}{\partial \theta_e}+\frac{x_e^{\prime}(\theta_e)}{x_e(\theta_e)}\right]$ and $\frac{d \ln l_e(\theta_e)}{d \theta_e}$ in (OC14) and (OC12) by the above equation and (OC20), respectively, and rearrange the formulas, we have

$$
\frac{1}{1-\tau_e^E\left(\theta_e\right)}=\frac{1+\left[1-\bar{g}_e(\theta_e)\right]\left\{\begin{array}{c}
H(\theta_e)\left[\left(\mu\left(\theta_e\right)-\xi\right) \frac{1+\varepsilon_e}{\varepsilon_e}-1\right]+\frac{1-F_e(\theta_e)}{f_e(\theta_e)} \frac{d \ln \left[\mu(\theta_e)-\xi\right]}{d \theta_e}+ \\
\mu\left(\theta_e\right) \varepsilon_{Q-ij}^{P, cross}\left(\theta_e\right) \frac{1-F_e(\theta_e)}{f_e(\theta_e)} \frac{\left(1+\xi \frac{1+\varepsilon_e}{\varepsilon_e}\right) \frac{d \ln \left[1-\tau_s^E(\theta_e)\right]}{d \theta_e}+\frac{d \ln \left[1-\tau_e^E(\theta_e)\right]}{d \theta_e}}{1-\frac{\sigma-1}{\sigma}\left(\frac{\varepsilon_e}{1+\varepsilon_e}+\xi\right)}
\end{array}\right\}}{\mu^*}
$$

and

$$
\begin{aligned}
\frac{1}{1-\tau_s^E\left(\theta_e\right)} & =\frac{\mu^*}{\mu\left(\theta_e\right)}+\left[1-\tau_e^E\left(\theta_e\right)\right] \varepsilon_{Q-ij}^{P, cross}\left(\theta_e\right)\left[\begin{array}{c}
\left[1-g_e(\theta_e)\right]-\frac{\left[1-\bar{g}_e(\theta_e)\right]\left[1-F_e(\theta_e)\right]}{f_e(\theta_e)} \times \\
{\left[\begin{array}{c}
\frac{1+\varepsilon_e}{\varepsilon_e} \frac{\frac{d\left[\ln \frac{X_e(\theta_e)}{\mu(\theta_e)}\right]}{d \theta_e}+\left[\frac{d \ln \left[1-\tau_s^E(\theta_e)\right]}{d \theta_e}+\left(1-\xi \frac{\sigma-1}{\sigma}\right) \frac{d \ln \left[1-\tau_e^E(\theta_e)\right]}{d \theta_e}\right]}{\frac{1+\varepsilon_e}{\varepsilon_e}\left(1-\xi \frac{\sigma-1}{\sigma}\right)-\frac{\sigma-1}{\sigma}} \\
+\frac{d \ln \left[\mu(\theta_e) \varepsilon_{Q-ij}^{P, cross}(\theta_e)\right]}{d \theta_e}
\end{array}\right]}
\end{array}\right] \\
& =\frac{\mu^*}{\mu\left(\theta_e\right)}+\left[1-\tau_e^E\left(\theta_e\right)\right] \varepsilon_{Q-ij}^{P, cross}\left(\theta_e\right)\left[\begin{array}{c}
\left[1-g_e(\theta_e)\right]-\left[1-\bar{g}_e(\theta_e)\right] H(\theta_e) \\
\frac{\left[1-\bar{g}_e(\theta_e)\right]\left[1-F_e(\theta_e)\right]}{f_e(\theta_e)}\left[\begin{array}{c}
\frac{d \ln \left[\mu(\theta_e) \varepsilon_{Q-ij}^{P, cross}(\theta_e)\right]}{d \theta_e}-\frac{d \ln \left[\mu(\theta_e)-\xi\right]}{d \theta_e} \\
+\frac{\frac{d \ln \left[1-\tau_s^E(\theta_e)\right]}{d \theta_e}+\left(1-\xi \frac{\sigma-1}{\sigma}\right) \frac{d \ln \left[1-\tau_e^E(\theta_e)\right]}{d \theta_e}}{\left(1-\xi \frac{\sigma-1}{\sigma}\right)-\frac{\varepsilon_e}{1+\varepsilon_e} \frac{\sigma-1}{\sigma}}
\end{array}\right]
\end{array}\right],
\end{aligned}
$$

where $H(\theta_e)=\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\left[\frac{\frac{1+\varepsilon_e}{\varepsilon_e} \frac{d \ln \left[X_e(\theta_e) / \mu(\theta_e)\right]}{d \theta_e}}{\frac{1+\varepsilon_e}{\varepsilon_e}\left(1-\xi \frac{\sigma-1}{\sigma}\right)-\frac{\sigma-1}{\sigma}}+\frac{d \ln \left[\mu(\theta_e)-\xi\right]}{d \theta_e}\right]$.

Set

$$
\begin{aligned}
H_0(\theta_e) &= \left\{ \begin{array}{c} [1-\bar{g}_e(\theta_e)] \left\{ \frac{1-F_e(\theta_e)}{f_e(\theta_e)} \frac{\frac{\sigma-1}{\sigma}\xi}{\mu(\theta_e)} \frac{d\ln[\mu(\theta_e)-\xi]}{d\theta_e} - \left[\frac{\sigma-1}{\sigma} - \left(1-\frac{\xi}{\mu(\theta_e)}\right)\frac{1+\varepsilon_e}{\varepsilon_e}\right] H(\theta_e) \right\} \\ +\varepsilon_{Q-ij}^{P,cross}(\theta_e)\,[1-g_e(\theta_e)] - \frac{1}{\mu(\theta_e)} \end{array} \right\}, \\
H_1(\theta_e) &= \frac{1+\left\{ \begin{array}{c} [1-\bar{g}_e(\theta_e)]\,H(\theta_e)\left([\mu(\theta_e)-\xi]\frac{1+\varepsilon_e}{\varepsilon_e}-1\right) + [1-\bar{g}_e(\theta_e)]\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{d\ln[\mu(\theta_e)-\xi]}{d\theta_e} \\ -[1-\bar{g}_e(\theta_e)]\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\mu(\theta_e)\,\varepsilon_{Q-ij}^{P,cross}(\theta_e)\frac{1+\frac{1+\varepsilon_e}{\varepsilon_e}\xi}{1-\frac{\sigma-1}{\sigma}\left(\frac{\varepsilon_e}{1+\varepsilon_e}+\xi\right)}\frac{d\ln H_0(\theta_e)}{d\theta_e} \end{array} \right\}}{\mu^*} \\
&= \frac{1+\left\{ \begin{array}{c} [1-\bar{g}_e(\theta_e)]\,H(\theta_e)\left[[\mu(\theta_e)-\xi]\frac{1+\varepsilon_e}{\varepsilon_e}-1\right] \\ +[1-\bar{g}_e(\theta_e)]\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{d\ln[\mu(\theta_e)-\xi]}{d\theta_e} - H_2(\theta_e)\frac{1+\frac{1+\varepsilon_e}{\varepsilon_e}\xi}{\frac{1+\varepsilon_e}{\varepsilon_e}\xi}\mu^*\frac{d\ln H_0(\theta_e)}{d\theta_e} \end{array} \right\}}{\mu^*}, \\
H_2(\theta_e) &= [1-\bar{g}_e(\theta_e)]\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{\mu(\theta_e)}{\mu^*}\varepsilon_{Q-ij}^{P,cross}(\theta_e)\frac{\frac{1+\varepsilon_e}{\varepsilon_e}\xi}{1-\frac{\sigma-1}{\sigma}\left(\frac{\varepsilon_e}{1+\varepsilon_e}+\xi\right)},
\end{aligned}
$$

where $H(\theta_e) = \frac{1-F_e(\theta_e)}{f_e(\theta_e)}\left[\frac{\frac{1+\varepsilon_e}{\varepsilon_e}\frac{d\ln[X_e(\theta_e)/\mu(\theta_e)]}{d\theta_e}}{\frac{1+\varepsilon_e}{\varepsilon_e}\left(1-\xi\frac{\sigma-1}{\sigma}\right)-\frac{\sigma-1}{\sigma}} + \frac{d\ln[\mu(\theta_e)-\xi]}{d\theta_e}\right]$.

We have

$$
\frac{1}{1-\tau_s^E(\theta_e)} = H_0(\theta_e)\left[1-\tau_e^E(\theta_e)\right]
$$

and

$$
\frac{1}{1-\tau_e^E(\theta_e)} = H_1(\theta_e) - H_2(\theta_e)\frac{d\ln\left[1-\tau_e^E(\theta_e)\right]}{d\theta_e}.
$$

Solving the above differential equation, we have

$$
1-\tau_e^E(\theta_e) = \left[1-\tau_e^E(\overline{\theta}_e)\right]e^{-\int_{\theta_e}^{\overline{\theta}_e}\frac{H_1(s)}{H_2(s)}ds} + \int_{\theta_e}^{\overline{\theta}_e} e^{-\int_{\theta_e}^{s}\frac{H_1(u)}{H_2(u)}du}\frac{1}{H_2(s)}ds,
$$

Last, according to (OC14), when $\overline{\theta}_e$ is finite, such that $1-F_e(\overline{\theta}_e)=0$, $\frac{\tau_e^E(\overline{\theta}_e)}{1-\tau_e^E(\overline{\theta}_e)} = \frac{1-\mu^*}{\mu^*}$.

OC.5.4 Part Two: A Result for Comparation

Notice that

$$
\begin{aligned}
\frac{\tau_s^E(\theta_e)}{1-\tau_s^E(\theta_e)} &= \overbrace{\left[\frac{\mu^*}{\mu(\theta_e)}-1\right]}^{\widetilde{RE}(\theta_e)} + \overbrace{\left[1-\tau_e^E(\theta_e)\right]\varepsilon_{Q-ij}^{P,cross}(\theta_e)\left[\begin{array}{c} [1-g_e(\theta_e)] - \frac{[1-\bar{g}_e(\theta_e)][1-F_e(\theta_e)]}{f_e(\theta_e)} \\ \times\left[\frac{1+\varepsilon_e}{\varepsilon_e}\frac{l_e'(\theta_e)}{l_e(\theta_e)} + \frac{d\ln\left[\mu(\theta_e)\varepsilon_{Q-ij}^{P,cross}(\theta_e)\right]}{d\theta_e}\right] \end{array}\right]}^{\widetilde{IRE}(\theta_e)} \\
&= \overbrace{\left[\frac{\mu^*}{\mu(\theta_e)}-1\right]}^{\widetilde{RE}(\theta_e)} + \left[1-\tau_e^E(\theta_e)\right]\varepsilon_{Q-ij}^{P,cross}(\theta_e)\left[\begin{array}{c} [1-g_e(\theta_e)] - \frac{[1-\bar{g}_e(\theta_e)][1-F_e(\theta_e)]}{f_e(\theta_e)}\times \\ \left[\begin{array}{c} \frac{1+\varepsilon_e}{\varepsilon_e}\frac{\frac{d}{d\theta_e}\left[\ln\frac{X_e(\theta_e)\left[1-\tau_s^E(\theta_e)\right]}{\mu(\theta_e)}\right]+\left[1-\xi\frac{\sigma-1}{\sigma}\right]\frac{d\ln\left[1-\tau_e^E(\theta_e)\right]}{d\theta_e}}{\frac{1+\varepsilon_e}{\varepsilon_e}-\frac{\sigma-1}{\sigma}\left(1+\frac{1+\varepsilon_e}{\varepsilon_e}\xi\right)} \\ +\frac{d\ln\left[\mu(\theta_e)\varepsilon_{Q-ij}^{P,cross}(\theta_e)\right]}{d\theta_e} \end{array}\right] \end{array}\right].
\end{aligned}
$$

When $\tau_s^{E\prime}(\theta_e) = \tau_e^{E\prime}(\theta_e) = 0$,

$$\frac{\tau_s^E(\theta_e)}{1-\tau_s^E(\theta_e)} = \left[\frac{\mu^*}{\mu(\theta_e)} - 1\right] + \left[1-\tau_e^E(\theta_e)\right] \varepsilon_{Q-ij}^{P,cross}(\theta_e) \left[\left[1-g_e(\theta_e)\right] - \left[1-\bar{g}_e(\theta_e)\right] H(\theta_e)\right].$$

Notice that

$$\frac{1}{1-\tau_e^E(\theta_e)} = \frac{1}{\mu^*}\left[1+\left[1-\bar{g}_e(\theta_e)\right]\left\{\begin{array}{c} \left[(\mu(\theta_e)-\xi)\frac{1+\varepsilon_e}{\varepsilon_e} - 1\right] H(\theta_e) + \frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{d\ln[\mu(\theta_e)-\xi]}{d\theta_e} + \\ \mu(\theta_e)\,\varepsilon_{Q-ij}^{P,cross}(\theta_e)\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{\left(1+\xi\frac{1+\varepsilon_e}{\varepsilon_e}\right)\frac{d\ln\left[1-\tau_s^E(\theta_e)\right]}{d\theta_e} + \frac{d\ln\left[1-\tau_e^E(\theta_e)\right]}{d\theta_e}}{1-\frac{\sigma-1}{\sigma}\left(\frac{\varepsilon_e}{1+\varepsilon_e}+\xi\right)} \end{array}\right\}\right].$$

We have the following results when $\frac{d\ln\left[1-\tau_e^E(\theta_e)\right]}{d\theta_e} = \frac{d\ln\left[1-\tau_s^E(\theta_e)\right]}{d\theta_e} = 0$:

$$\frac{1}{1-\tau_e^E(\theta_e)} = \frac{1+\left[1-\bar{g}_e(\theta_e)\right]\left\{H(\theta_e)\left[(\mu(\theta_e)-\xi)\frac{1+\varepsilon_e}{\varepsilon_e} - 1\right] + \frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{d\ln[\mu(\theta_e)-\xi]}{d\theta_e}\right\}}{\mu^*}.$$

If in addition $\mu'(\theta_e) = 0$:

$$\frac{1}{1-\tau_e^E(\theta_e)} = \frac{1+\left[1-\bar{g}_e(\theta_e)\right] H(\theta_e)\left[\left[\mu(\theta_e)-\xi\right]\frac{1+\varepsilon_e}{\varepsilon_e} - 1\right]}{\mu^*}.$$

OC.6 Proof of Proposition 6

(i) Notice that (OA26). Denote by y_e^i, $F_{y_e}^i(y_e^i)$, and $f_{y_e}^i(y_e^i)$ the initial profit, CDF, and PDF of initial profit, respectively. $\frac{1-F_{y_e}^i(y_e^i(\widehat{\theta}_e))}{y_e^i(\widehat{\theta}_e) f_{y_e}^i\left(y_e^i(\widehat{\theta}_e)\right)}$ is the initial hazard ratio of profit. When the profit and sales taxes are linear in the initial, the initial hazard ratio of profit $\frac{1-F_{y_e}^i(y_e^i(\widehat{\theta}_e))}{y_e^i(\widehat{\theta}_e) f_{y_e}^i\left(y_e^i(\widehat{\theta}_e)\right)} = H(\theta_e)$. Thus $H(\theta_e)$ is constant for $\theta_e \geq \widehat{\theta}_e$, if the initial hazard ratio of profit is constant for $\theta_e > \widehat{\theta}_e$.

(ii) Notice that if for any $\theta_e > \widehat{\theta}_e$, $g_e(\theta_e) = \widehat{g}_e$ is constant, $\overline{g}_e(\theta_e)$ is also constant and equal to $\widehat{g}_e$ for any $\theta_e \geq \widehat{\theta}_e$. With the above two findings, part two of Proposition 6 follow directly from Theorem 1.

(iii) To prove part three of Proposition 6, first note that $\frac{1}{1-\widehat{\tau}_e}$ and thus $\widehat{\tau}_e$ increases with μ. Notice that for any $\theta_e \in \Theta_e$, $\mu(\theta_e)$ non-decreases with the decrease of I, μ as a weighted average of $\mu(\theta_e)$ must non-decreases with the decrease of I. Thus a condition guarantees that, given μ, $\frac{1}{1-\widehat{\tau}_e}$ increases with $\widehat{\mu}$ is a sufficient condition for $\widehat{\tau}_e$ to be increasing with the decrease of I. We now show that (OC21) is such a condition.

To see this, we first treat $\widehat{g}_e$ and μ as given and take the derivative of $\frac{1}{1-\widehat{\tau}_e}$ with respect to $\widehat{\mu}$:

$$\frac{d\left(\frac{1}{1-\widehat{\tau}_e}\right)}{d\widehat{\mu}} = \frac{-\frac{1+(1-\widehat{g}_e)\widehat{H}\left[\frac{1+\varepsilon_e}{\varepsilon_e}(\widehat{\mu}-\xi)-1\right]}{\widehat{\mu}^2}+\frac{(1-\widehat{g}_e)\widehat{H}\frac{1+\varepsilon_e}{\varepsilon_e}}{\widehat{\mu}}+\frac{\frac{\sigma}{\sigma-1}\frac{1}{(\widehat{\mu})^2}}{\frac{\sigma}{\sigma-1}-\xi}\left(1-\widehat{g}_e\right)\left(1-\widehat{H}\right)}{1-\frac{\xi}{\frac{\sigma}{\sigma-1}-\xi}\frac{\mu-\widehat{\mu}}{\widehat{\mu}}} - \frac{\frac{\xi}{\frac{\sigma}{\sigma-1}-\xi}\frac{\mu}{\widehat{\mu}^2}\left[\begin{array}{c}\frac{1}{\widehat{\mu}}\left[1+\left(1-\widehat{g}_e\right)\widehat{H}\left[\frac{1+\varepsilon_e}{\varepsilon_e}\left(\widehat{\mu}-\xi\right)-1\right]\right]\\ +\frac{1-\frac{\sigma}{\sigma-1}\frac{1}{\widehat{\mu}}}{\frac{\sigma}{\sigma-1}-\xi}\left(1-\widehat{g}_e\right)\left(1-\widehat{H}\right)\end{array}\right]}{\left[1-\frac{\xi}{\frac{\sigma}{\sigma-1}-\xi}\frac{\mu-\widehat{\mu}}{\widehat{\mu}}\right]^2}$$

Rearranging the right side of the above equation, we find that $\frac{d}{d\widehat{\mu}}\left(\frac{1}{1-\widehat{\tau}_e}\right) > 0$ if and only if

$$\left(1-\widehat{g}_e\right)\widehat{H}\left[\begin{array}{c}\left(\frac{\sigma}{\sigma-1}-\xi\right)\left[\left(\widehat{\mu}-\xi\mu\frac{\sigma-1}{\sigma}\right)\frac{1+\varepsilon_e}{\varepsilon_e}-\frac{1}{\varepsilon^{y_e,top}_{1-\widehat{\tau}_e}}\right]\\ +\left(\frac{\sigma}{\sigma-1}-\xi\mu\frac{\sigma-1}{\sigma}\right)\end{array}\right] > \frac{\sigma}{\sigma-1}-\xi.$$

Note that the term in the bracket is positive, because for any $\theta_e \in \Theta_e$, $\mu\left(\theta_e\right) \leq \frac{\sigma}{\sigma-1}$. Thus, we have

$$1-\widehat{g}_e > \frac{\frac{\sigma}{\sigma-1}-\xi}{\widehat{H}\left[\begin{array}{c}\left(\frac{\sigma}{\sigma-1}-\xi\right)\left[\left(\widehat{\mu}-\xi\mu\frac{\sigma-1}{\sigma}\right)\frac{1+\varepsilon_e}{\varepsilon_e}-\frac{1}{\varepsilon^{y_e,top}_{1-\widehat{\tau}_e}}\right]\\ +\left(\frac{\sigma}{\sigma-1}-\xi\mu\frac{\sigma-1}{\sigma}\right)\end{array}\right]}.$$

Dividing both the numerator and denominator of the right side of the above inequality by $\frac{\sigma}{\sigma-1}-\xi$ and substituting $\frac{1}{\varepsilon^{y_e,top}_{1-\widehat{\tau}_e}}$ by $\frac{1+\varepsilon_e}{\varepsilon_e}\left(\widehat{\mu}-\xi\right)-1$, we have

$$\widehat{g}_e < 1-\frac{1}{\left[\xi\left(1-\mu\frac{\sigma-1}{\sigma}\right)\frac{1+\varepsilon_e}{\varepsilon_e}+1+\frac{\frac{\sigma}{\sigma-1}-\xi\mu\frac{\sigma-1}{\sigma}}{\frac{\sigma}{\sigma-1}-\xi}\right]\widehat{H}}. \tag{OC21}$$

(OC21). Notice that $\mu \leq \frac{\sigma}{\sigma-1}$, the term in the bracket of the right side of (A64) is not less than 2. Thus, condition (A64) satisfies if $\widehat{g}_e < 1-\frac{1}{2\widehat{H}}$.

Last, since for any $I > 1$, (A65) is a sufficient condition for $\widehat{\tau}_e$ to be increasing with the marginal decrease of I (just suppose that I is continuous), and $\widehat{\tau}_e$ is twice continuously differentiable with respect to I, according to the mean value theorem, (A65) is also a sufficient condition for $\widehat{\tau}_e$ to be increasing with the decrease of I from $n+1$ to n, where $n \in \mathbb{N}_+$.■

OC.7 Proof of Proposition 5

(i) Rewriting the general tax formula for the entrepreneurs in equation (36) for the case with uniform markups, we have part one of Proposition 5. By the definitions of $\varepsilon^{y_e}_{1-\tau_e}\left(\theta_e\right)$ and $\varepsilon^{P,cross}_{Q_{-ij}}\left(\theta_e\right)$ (see section

A.2 for details), we have:

$$\frac{1}{1-\tau_e(\theta_e)} = \frac{1+[1-\bar{g}_e(\theta_e)]\,H(\theta_e)\left[\frac{1+\varepsilon_e}{\varepsilon_e}[\mu-\xi]-1\right]}{\mu} + \frac{\frac{\sigma}{\sigma-1}}{\frac{\sigma}{\sigma-1}-\xi}\left[\frac{\sigma-1}{\sigma}-\frac{1}{\mu}\right]\{[1-g_e(\theta_e)]-[1-\bar{g}_e(\theta_e)]\,H(\theta_e)\}.$$

(ii) Supposing $g_o(\cdot)$ (such that $\bar{g}_o(\cdot)$) is exogenous, we have:

$$\begin{aligned}
\frac{d\left[\frac{1}{1-\tau_e(\theta_e)}\right]}{d\mu} &= \frac{1}{\mu^2}\left[\begin{array}{c}\mu[1-\bar{g}_e(\theta_e)]\,H(\theta_e)\frac{1+\varepsilon_e}{\varepsilon_e}-[1-\bar{g}_e(\theta_e)]\,H(\theta_e)\left[(\mu-\xi)\frac{1+\varepsilon_e}{\varepsilon_e}-1\right]\\ +\frac{\frac{\sigma}{\sigma-1}}{\frac{\sigma}{\sigma-1}-\xi}\{[1-g_e(\theta_e)]-[1-\bar{g}_e(\theta_e)]\,H(\theta_e)\}-1\end{array}\right]\\
&= \frac{1}{\mu^2}\left[\begin{array}{c}[1-\bar{g}_e(\theta_e)]\,H(\theta_e)\left[\xi\frac{1+\varepsilon_e}{\varepsilon_e}+1\right]\\ +\frac{\frac{\sigma}{\sigma-1}}{\frac{\sigma}{\sigma-1}-\xi}\{[1-g_e(\theta_e)]-[1-\bar{g}_e(\theta_e)]\,H(\theta_e)\}-1\end{array}\right]\\
&= \frac{1}{\mu^2}\left[\begin{array}{c}[1-\bar{g}_e(\theta_e)]\,H(\theta_e)\xi\left[\frac{1+\varepsilon_e}{\varepsilon_e}-\frac{1}{\frac{\sigma}{\sigma-1}-\xi}\right]\\ +[1-g_e(\theta_e)]\frac{\frac{\sigma}{\sigma-1}}{\frac{\sigma}{\sigma-1}-\xi}-1\end{array}\right]
\end{aligned}$$

and

$$\begin{aligned}
\frac{d\left[\frac{1-\tau_w(\theta_w)}{1-\tau_e(\theta_e)}\right]}{d\mu} &= \frac{1-\tau_w(\theta_w)}{\mu}\frac{d\left[\frac{\mu}{1-\tau_e(\theta_e)}\right]}{d\mu}\\
&= \frac{1-\tau_w(\theta_w)}{\mu}\left[\begin{array}{c}[1-\bar{g}_e(\theta_e)]\,H(\theta_e)\frac{1+\varepsilon_e}{\varepsilon_e}+\\ \frac{1}{\frac{\sigma}{\sigma-1}-\xi}\{[1-g_e(\theta_e)]-[1-\bar{g}_e(\theta_e)]\,H(\theta_e)\}\end{array}\right].
\end{aligned}$$

According to the above equations, $\tau_e(\theta_e)$ increases in μ if and only if:

$$[1-\bar{g}_e(\theta_e)]\,H(\theta_e)\xi\left[\frac{1+\varepsilon_e}{\varepsilon_e}-\frac{1}{\frac{\sigma}{\sigma-1}-\xi}\right]+[1-g_e(\theta_e)]\frac{\frac{\sigma}{\sigma-1}}{\frac{\sigma}{\sigma-1}-\xi}-1>0,$$

i.e.,

$$\begin{aligned}
g_e(\theta_e) &< 1-\frac{1-[1-\bar{g}_e(\theta_e)]\,H(\theta_e)\xi\left[\frac{1+\varepsilon_e}{\varepsilon_e}-\frac{1}{\frac{\sigma}{\sigma-1}-\xi}\right]}{\frac{\frac{\sigma}{\sigma-1}}{\frac{\sigma}{\sigma-1}-\xi}}\\
&= \frac{\xi(\sigma-1)}{\sigma}\left\{1+[1-\bar{g}_e(\theta_e)]\,H(\theta_e)\left[\frac{1+\varepsilon_e}{\varepsilon_e}\left(\frac{\sigma}{\sigma-1}-\xi\right)-1\right]\right\}.
\end{aligned} \tag{OC22}$$

$\frac{1-\tau_w(\theta_w)}{1-\tau_e(\theta_e)}$ increases in μ iff

$$[1-\bar{g}_e(\theta_e)]\,H(\theta_e)\frac{1+\varepsilon_e}{\varepsilon_e}+\frac{1}{\frac{\sigma}{\sigma-1}-\xi}\{[1-g_e(\theta_e)]-[1-\bar{g}_e(\theta_e)]\,H(\theta_e)\}>0,$$

i.e.,

$$g_e(\theta_e)<1+[1-\bar{g}_e(\theta_e)]\,H(\theta_e)\left[\frac{1+\varepsilon_e}{\varepsilon_e}\left(\frac{\sigma}{\sigma-1}-\xi\right)-1\right], \tag{OC23}$$

or equivalently

$$g_e(\theta_e) - 1 + [1 - \bar{g}_e(\theta_e)]\, H(\theta_e) < [1 - \bar{g}_e(\theta_e)]\, H(\theta_e) \frac{1+\varepsilon_e}{\varepsilon_e} \left(\frac{\sigma}{\sigma - 1} - \xi \right),$$

which must be true when $[1 - g_e(\theta_e)] - [1 - \bar{g}_e(\theta_e)]\, H(\theta_e) > 0$ and $H(\theta_e) > 0$.

(iii) When $g_e(\theta_e) = \bar{g}_e(\theta_e)$ and $H(\theta_e) > 0$, inequality (OC22) is equivalent to

$$g_e(\theta_e) < \frac{\frac{\xi(\sigma-1)}{\sigma} \left\{ 1 + H(\theta_e) \left[\frac{1+\varepsilon_e}{\varepsilon_e} \left(\frac{\sigma}{\sigma-1} - \xi \right) - 1 \right] \right\}}{1 + \frac{\xi(\sigma-1)}{\sigma} H(\theta_e) \left[\frac{1+\varepsilon_e}{\varepsilon_e} \left(\frac{\sigma}{\sigma-1} - \xi \right) - 1 \right]},$$

where $\frac{1+H(\theta_e)\left[\frac{1+\varepsilon_e}{\varepsilon_e}\left(\frac{\sigma}{\sigma-1}-\xi\right)-1\right]}{\frac{\sigma}{\xi(\sigma-1)}+H(\theta_e)\left[\frac{1+\varepsilon_e}{\varepsilon_e}\left(\frac{\sigma}{\sigma-1}-\xi\right)-1\right]}$ increases in $H(\theta_e) \left[\frac{1+\varepsilon_e}{\varepsilon_e} \left(\frac{\sigma}{\sigma-1} - \xi \right) - 1 \right]$, because $\frac{\sigma}{\xi(\sigma-1)} > 1$. Besides, under condition (24), we have $\frac{1+\varepsilon_e}{\varepsilon_e} \left(\frac{\sigma}{\sigma-1} - \xi \right) - 1 > 0$ and that $H(\theta_e) \left[\frac{1+\varepsilon_e}{\varepsilon_e} \left(\frac{\sigma}{\sigma-1} - \xi \right) - 1 \right]$ increases in $H(\theta_e)$. Thus, when $H(\theta_e) > 0$, the above inequality holds if

$$g_e(\theta_e) \leq \frac{\xi\,(\sigma - 1)}{\sigma}.$$

Analogously, when $g_e(\theta_e) = \bar{g}_e(\theta_e) < 1$ and $H(\theta_e) > 0$, inequality (OC23) must holds.■

OD Discussion and Robustness

OD.1 Kimball Technology

Set the Lagrangian function as

$$\begin{aligned}
& \pounds\left(L_w, l_w, l_e, V_w, V_e, Q; \lambda, \lambda', \psi_w, \psi_e, \kappa, \varphi\right) \\
= \quad & \sum_{o\in\{w,e\}} N_o \int_{\theta_o} G\left(V_o(\theta_o)\right) \widetilde{f}_o\left(\theta_o\right) d\theta_o + \lambda \left[Q - \sum_{o\in\{w,e\}} N_o \int_{\theta_o} \left[V_o\left(\theta_o\right) + \phi_o\left(l_o\left(\theta_o\right)\right)\right] f_o\left(\theta_o\right) d\theta_o - R\right] \\
& + \lambda' \left[N_w \int_{\theta_w} x_w\left(\theta_w\right) l_w\left(\theta_w\right) f_w\left(\theta_w\right) d\theta_w - N_e \int_{\theta_e} L_w\left(\theta_e\right) f_e\left(\theta_e\right) d\theta_e\right] \\
& + \lambda'' \left[1 - \int_{\theta_e} \chi\left(\theta_e\right) Y \left(\frac{Q\left(\theta_e\right)}{Q/N_e}\right) dF_{\theta_e}\left(\theta_e\right)\right] \\
& + \int_{\theta_w} \psi_w\left(\theta_w\right) \left[l_w\left(\theta_w\right) \phi_w'\left(l_w\left(\theta_w\right)\right) \frac{x_w'\left(\theta_e\right)}{x_w\left(\theta_e\right)} - V_w'(\theta_w)\right] d\theta_w \\
& + \int_{\theta_e} \psi_e\left(\theta_e\right) \left[\phi_e'\left(l_e\left(\theta_e\right)\right) l_e\left(\theta_e\right) \left[\mu(\theta_e) \frac{\chi'\left(\theta_e\right)}{\chi\left(\theta_e\right)} + \frac{x_e'\left(\theta_e\right)}{x_e\left(\theta_e\right)}\right] - V_e'(\theta_e)\right] d\theta_e,
\end{aligned}$$

where $\mu(\theta_e)$ is a function of $\frac{Q(\theta_e)}{Q}$.

Taking partial integrals yields the following

$$-\int_{\theta_e} \psi_o(\theta_e) V_o'(\theta_e) d\theta_e = V_o(\underline{\theta}_o)\psi_o(\underline{\theta}_o) - V_o(\overline{\theta}_o)\psi_o(\overline{\theta}_o) + \int_{\theta_o} \psi_o'(\theta_o) V_o(\theta_o) d\theta_o.$$

The derivatives with respect to the endpoint conditions yield boundary conditions:

$$\psi_o(\underline{\theta}_o) = \psi_o(\overline{\theta}_o) = 0, o \in \{w, e\}.$$

Thus,

$$\int_{\theta_o} \psi_o'\left(\theta_o\right) d\theta_o = 0,$$

Substituting the above conditions into the Lagrangian function, yields the following first-order conditions:

$$\frac{\partial \pounds}{\partial Q} = \lambda + N_e \lambda'' \int_{\theta_e} \chi\left(\theta_e\right) Y' \left(\frac{Q\left(\theta_e\right)}{Q/N_e}\right) \frac{Q\left(\theta_e\right)}{Q^2} dF_e\left(\theta_e\right) = 0, o \in \{w, e\}, \tag{OD2}$$

$$\frac{\partial \pounds}{\partial V_o(\theta_o)} = G'(V_o(\theta_o)) N_o \widetilde{f}_o\left(\theta_o\right) + \psi_o'(\theta_o) - \lambda N_o f_o\left(\theta_o\right) = 0, \tag{OD3}$$

$$\frac{\partial \pounds}{\partial l_w\left(\theta_w\right)} = -\lambda N_w \phi_w'\left(l_w\left(\theta_w\right)\right) f_w\left(\theta_w\right) + \lambda' N_w x_w\left(\theta_w\right) f_w\left(\theta_w\right) + \psi_w\left(\theta_w\right) \frac{\phi_w'\left(l_w\left(\theta_w\right)\right)}{x_w\left(\theta_w\right)} \frac{1+\varepsilon_w}{\varepsilon_w} = 0, \tag{OD4}$$

$$\frac{\partial \pounds}{\partial L_w(\theta_e)} = \left[\begin{array}{c} -\lambda'' N_e \chi\left(\theta_e\right) Y' \left(\frac{Q(\theta_e)}{Q/N_e}\right) \frac{1}{Q} \frac{\partial Q(\theta_e)}{\partial L_w(\theta_e)} f_e\left(\theta_e\right) - \lambda' N_e f_e\left(\theta_e\right) \\ + \psi_e\left(\theta_e\right) \phi_e'\left(l_e\left(\theta_e\right)\right) l_e\left(\theta_e\right) \frac{\partial \mu(\theta_e)}{\partial Q(\theta_e)} \frac{\partial Q(\theta_e)}{\partial L_w(\theta_e)} \frac{\chi'(\theta_e)}{\chi(\theta_e)} \end{array}\right] = 0, \tag{OD5}$$

and

$$\frac{\partial \pounds}{\partial l_e(\theta_e)} = \psi_e(\theta_e)\,\phi_e'(l_e(\theta_e))\,\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu(\theta_e)\frac{\chi'(\theta_e)}{\chi(\theta_e)}+\frac{x_e'(\theta_e)}{x_e(\theta_e)}\right]+ \tag{OD6}$$
$$\psi_e(\theta_e)\,\phi_e'(l_e(\theta_e))\,l_e(\theta_e)\,\frac{\partial\mu(\theta_e)}{\partial Q(\theta_e)}\frac{\partial Q(\theta_e)}{\partial l_e(\theta_e)}\frac{\chi'(\theta_e)}{\chi(\theta_e)}+$$
$$\left[-\lambda'' N_e\chi(\theta_e)\,Y'\left(\frac{Q(\theta_e)}{Q}\right)\frac{1}{Q}\frac{\partial Q(\theta_e)}{\partial l_e(\theta_e)}f_e(\theta_e)-\lambda\phi_e'(l_e(\theta_e))\,N_e\right]f_e(\theta_e)=0,\forall\theta_o\in\Theta_o.$$

Substitute λ'' in (OD5) by (OD2) and use (A69):$\tau_s(\theta_e) = 1-\frac{W}{\varpi(\theta_e)} = 1-\frac{W}{\frac{P(\theta_e)}{\mu(\theta_e)}\frac{\partial Q(\theta_e)}{\partial L_w(\theta_e)}}$; $P(\theta_e)\frac{\partial Q(\theta_e)}{\partial L_w(\theta_e)} = \frac{\mu(\theta_e)W}{1-\tau_s(\theta_e)}$; $\frac{P(\theta_e)Q(\theta_e)}{WL_w(\theta_e)}\xi = \frac{\mu(\theta_e)}{1-\tau_s(\theta_e)}$

$$P(\theta_e)\frac{\partial Q(\theta_e)}{\partial L_w(\theta_e)} = \frac{\lambda'}{\lambda}+\frac{[1-\bar{g}_e(\theta_e)]\,[1-F_e(\theta_e)]}{f_e(\theta_e)}\phi_e'(l_e(\theta_e))\,l_e(\theta_e)\,\frac{\partial\mu(\theta_e)}{\partial Q(\theta_e)}\frac{\partial Q(\theta_e)}{\partial L_w(\theta_e)}\frac{\chi'(\theta_e)}{\chi(\theta_e)},$$

where we substitute $\frac{\psi_e(\theta_e)}{\lambda N_e f_e(\theta_e)}$ by (A42). Substituting $P(\theta_e)\frac{\partial Q_{ij}(\theta_e)}{\partial L_w(\theta_e)}$ by $\frac{\mu(\theta_e)W}{1-\tau_s(\theta_e)}$ gives:

$$\frac{1}{1-\tau_s(\theta_e)} = \frac{\frac{\lambda'}{\lambda}}{\mu(\theta_e)\,W}+\frac{[1-\bar{g}_e(\theta_e)]\,[1-F_e(\theta_e)]}{f_e(\theta_e)}\frac{\phi_e'(l_e(\theta_e))\,l_e(\theta_e)}{WL_w(\theta_e)}\frac{\partial\ln\mu(\theta_e)}{\partial\ln Q(\theta_e)}\xi\frac{\chi'(\theta_e)}{\chi(\theta_e)}, \tag{OD7}$$

Equivalently,

$$\frac{1}{1-\tau_s(\theta_e)} = \frac{\frac{\lambda'}{\lambda}}{\mu(\theta_e)\,W}+\frac{[1-\bar{g}_e(\theta_e)]\,[1-F_e(\theta_e)]}{f_e(\theta_e)}\frac{\phi_e'(l_e(\theta_e))}{P(\theta_e)\frac{\partial Q(\theta_e)}{\partial l_e(\theta_e)}}\frac{Q(\theta_e)\,P(\theta_e)}{WL_w(\theta_e)}\frac{\partial\ln\mu(\theta_e)}{\partial\ln Q(\theta_e)}\xi\frac{\chi'(\theta_e)}{\chi(\theta_e)}.$$

Substitute $\frac{Q(\theta_e)P(\theta_e)}{WL_w(\theta_e)}$ by $\frac{P(\theta_e)Q(\theta_e)}{WL_w(\theta_e)}\xi = \frac{\mu(\theta_e)}{1-\tau_s(\theta_e)}$ and substitute $\frac{\phi_e'(l_e(\theta_e))}{P(\theta_e)\frac{\partial Q_{ij}(\theta_e)}{\partial l_e(\theta_e)}}$ by $\frac{1-\tau_e(\theta_e)}{\mu(\theta_e)}$

$$\frac{1}{1-\tau_s(\theta_e)} = \frac{\frac{\lambda'}{\lambda}}{\mu(\theta_e)\,W}+\frac{[1-\bar{g}_e(\theta_e)]\,[1-F_e(\theta_e)]}{f_e(\theta_e)}\frac{1-\tau_e(\theta_e)}{1-\tau_s(\theta_e)}\frac{\partial\ln\mu(\theta_e)}{\partial\ln Q(\theta_e)}\frac{\chi'(\theta_e)}{\chi(\theta_e)}.$$

Rearrange the above equation to derive

$$1-\tau_s(\theta_e) = \frac{1-\frac{[1-\bar{g}_e(\theta_e)][1-F_e(\theta_e)]}{f_e(\theta_e)}\,[1-\tau_e(\theta_e)]\,\frac{\partial\ln\mu(\theta_e)}{\partial\ln Q(\theta_e)}\frac{\chi'(\theta_e)}{\chi(\theta_e)}}{\frac{\lambda'}{\lambda\mu(\theta_e)W}}. \tag{OD8}$$

Therefore,

$$\frac{1-\tau_s(\theta_e')}{1-\tau_s(\theta_e)} = \frac{1-\frac{[1-\bar{g}_e(\theta_e')][1-F_e(\theta_e')]}{f_e(\theta_e')}\,[1-\tau_e(\theta_e')]\,\frac{\partial\ln\mu(\theta_e')}{\partial\ln Q(\theta_e')}\frac{\chi'(\theta_e')}{\chi(\theta_e')}}{1-\frac{[1-\bar{g}_e(\theta_e)][1-F_e(\theta_e)]}{f_e(\theta_e)}\,[1-\tau_e(\theta_e)]\,\frac{\partial\ln\mu(\theta_e)}{\partial\ln Q(\theta_e)}\frac{\chi'(\theta_e)}{\chi(\theta_e)}}.$$

Substitute $\lambda'' N_e$ in (OD6) by (OD2) and use (A69):

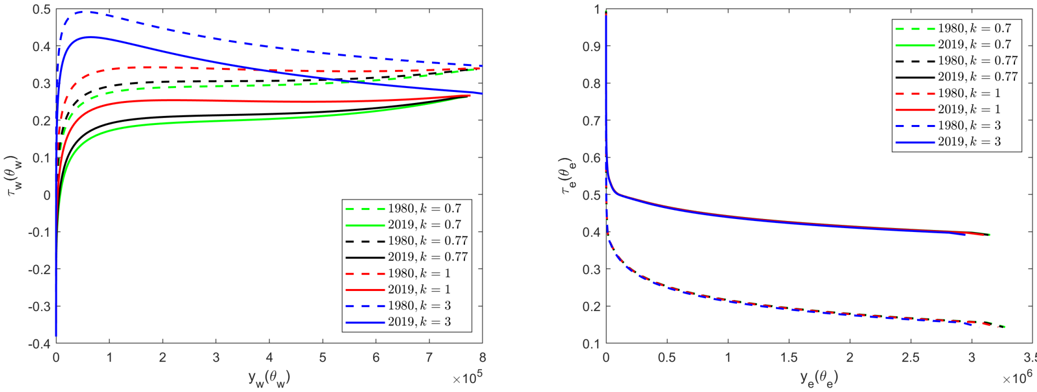
$$\frac{1-\frac{\phi_e'(l_e(\theta_e))}{P(\theta_e)\frac{\partial Q(\theta_e)}{\partial l_e(\theta_e)}}}{\frac{\phi_e'(l_e(\theta_e))}{P(\theta_e)\frac{\partial Q(\theta_e)}{\partial l_e(\theta_e)}}} = [1-\bar{g}_e(\theta_e)]\,\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\left[\begin{array}{c}\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu(\theta_e)\frac{\chi'(\theta_e)}{\chi(\theta_e)}+\frac{x_e'(\theta_e)}{x_e(\theta_e)}\right]\\ +\mu(\theta_e)\frac{\partial\ln\mu(\theta_e)}{\partial\ln Q(\theta_e)}\frac{\chi'(\theta_e)}{\chi(\theta_e)}\end{array}\right] \tag{OD9}$$

where we substitute $\frac{\psi_e(\theta_e)}{\lambda N_e f_e(\theta_e)}$ by (A42). Using $\frac{1-\tau_e(\theta_e)}{\mu(\theta_e)} = \frac{\phi_e'(l_e(\theta_e))}{P(\theta_e)\frac{\partial Q_{ij}(\theta_e)}{\partial l_e(\theta_e)}}$ to substitute $\frac{\phi_e'(l_e(\theta_e))}{P(\theta_e)\frac{\partial Q(\theta_e)}{\partial l_e(\theta_e)}}$ in (OD9), delivers (OD10).

$$\frac{1-\frac{1-\tau_e(\theta_e)}{\mu(\theta_e)}}{\frac{1-\tau_e(\theta_e)}{\mu(\theta_e)}} = [1-\bar{g}_e(\theta_e)]\,\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\left[\begin{array}{c} \frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu(\theta_e)\frac{\chi'(\theta_e)}{\chi(\theta_e)}+\frac{x_e'(\theta_e)}{x_e(\theta_e)}\right] \\ +\mu(\theta_e)\frac{\partial\ln\mu(\theta_e)}{\partial\ln Q(\theta_e)}\frac{\chi'(\theta_e)}{\chi(\theta_e)} \end{array}\right]. \tag{OD10}$$

OE Supplements to Quantitative Analysis

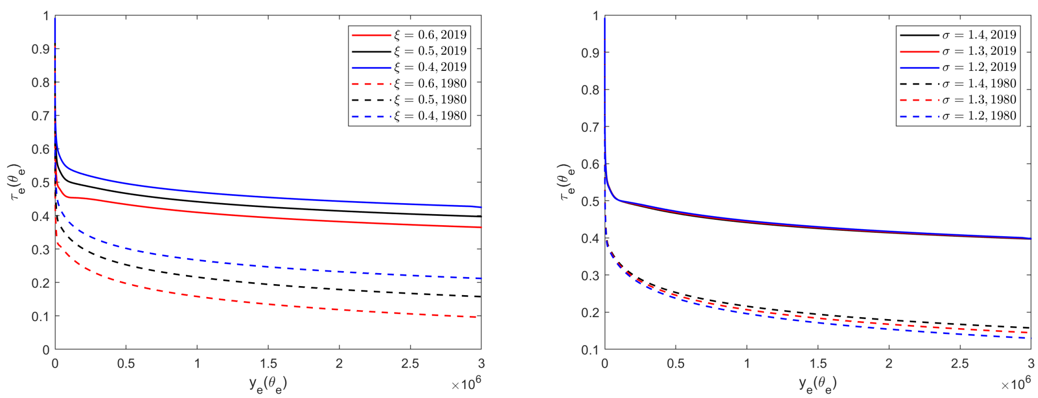
Figure OE2 gathers results regarding to the changes with k.



(a) Labor Tax Wedges (b) Profit Tax Wedges

Figure OE2: Tax Wedges Change with k

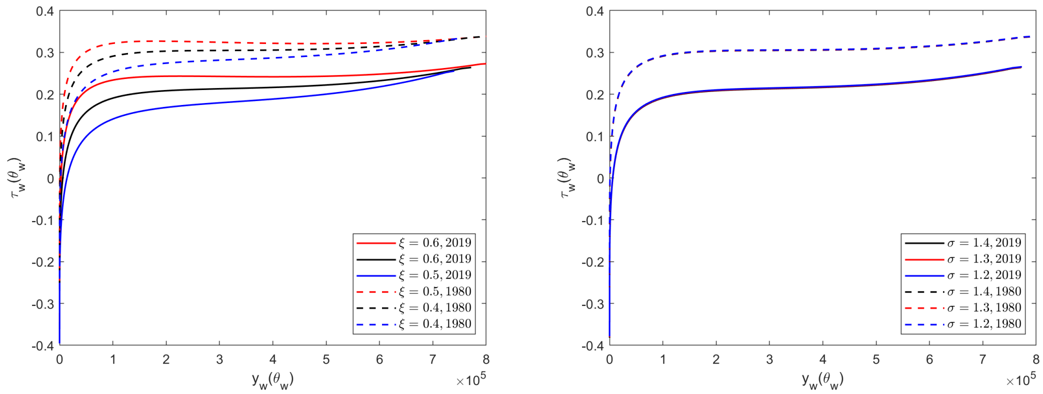
Figure OE3 gathers changes in profit tax rates with the changes in ξ and σ.



(a) Profit Tax Wedges (b) Profit Tax Wedges

Figure OE3: Profit Tax Wedges Change with ξ and σ

Figure OE4 gathers changes in labor income tax rates with the changes in ξ and σ.



(a) Labor Income Tax Wedges

(b) Labor Income Tax Wedges

Figure OE4: Labor Income Tax Wedges Change with ξ and σ