OPTIMAL TAXATION AND MARKET POWER*

Jan Eeckhout, UPF Barcelona†

Chunyang Fu, UCASS‡

Wenjian Li, Zhejiang University§

Xi Weng, Peking University¶

November 27, 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 derives optimal tax rates that depend on markups and identifies four distinct components 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. Our theory provides concrete proposals for how policymakers can optimally use income taxation to redistribute income to the poor, while incentivizing production in the presence of market power.

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, 72225001, and 72192844), 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 firms (profits) and less to workers (wages). Additionally, market power creates inefficiencies in the allocation of resources as prices become excessively high. Moreover, production is misallocated across heterogeneous firms, which leads to deadweight loss and a reduction in welfare. This raises the question of whether, and how, optimal tax design can mitigate the effect of market power and enhance efficiency and social welfare. In light of the rise in market power since 1980, answering this question by incorporating market power in optimal tax analysis has significant policy implications.

In this paper, we answer this question both from normative and positive standpoints. Normatively, we analytically derive optimal tax formulas for an economy with market power. These formulas break down the various mechanisms through which the optimal tax regime enhances efficiency and welfare. Positively, we use these formulas to make concrete tax policy recommendations for the joint design of labor and profit income taxes. By matching the model to moments from the US economy between 1980 and 2019 – a period marked by increased market power – our quantitative model prescribes a decline in labor income tax rate by 10 percentage points and an increase in the profit rate by 3 percentage points on average, and by 25 percentage points at the top.

To explore optimal taxation in the context of market power, we develop a canonical Mirrlees (1971) taxation framework augmented with (i) oligopolistic markets that feature endogenous pricing and markups, (ii) heterogenous firm productivities, and (iii) both wage-earning workers and profit-earning entrepreneurs. Central to our setup is that entrepreneurs in our model are agents with private information and market power. Worker and entrepreneurial behavior, including labor supply and pricing, respond to the tax system, which in turn determines the primary distribution of income and the efficiency of production. A key innovation of our model is endogenous market power in oligopolistic markets. Variable markups are crucial because they generate inequality between firms, both in markups and profits, which aligns with the data.

Our *main contribution* is to demonstrate that strategic pricing and market structure play a critical role in optimal taxation, both theoretically and quantitatively. The model captures several empirically relevant features that link inequality and inefficiency to market power. The observed rise in economy-wide market power is driven by increased inequality of markups, which directly influences income inequality.[1] The model allows us to formulate specific taxation proposals for policymakers to address market power.[2] The novel elements in our framework are crucial for policy design. The traditional Mirrleesian tax provides the correct incentives by balancing efficient effort supply against inequality. In addition, our optimal tax formulas simultaneously address externalities, inefficiency, and inequality of profits stemming from endogenous markups.

[1]As market power increases, labor income decreases while there is an increase in the level and inequality of income of entrepreneurs. The rise in market power also leads to a decrease in output and social welfare. These outcomes are consistent with the documented decline in the labor share, which coincides with the rise in market power. See Karabarbounis and Neiman (2014), De Loecker et al. (2020), and Autor et al. (2020).

[2]We design optimal policy relying solely on labor income, profits, and commodity taxes. The most straightforward way to mitigate the distortionary effects of market power is to eliminate its root causes through antitrust policies. However, optimal antitrust policies may not always be feasible. Antitrust policy faces significant challenges because market power has multiple origins: technological factors, such as entry barriers, returns to scale, and firm heterogeneity; and market structure factors, such as mergers and acquisitions (M&A) (see Sutton (1991, 2001) and De Loecker et al. (2019)).

Our primary theoretical finding is the derivation of an optimal tax formula comprising four components, which illustrate how optimal taxes deviate from classic Mirrleesian taxes due to endogenous market power.[3] (i) the traditional Mirrleesian component; (ii) the Pigouvian component; (iii) the reallocation effect (RE); and (iv) the indirect redistribution effect (IRE).

(i) The traditional Mirrleesian component balances the redistributive benefits of higher tax rates against the associated efficiency losses. Both labor income and profit taxes exhibit Mirrleesian characteristics, as both worker and entrepreneur exert endogenous effort. The other three components address externalities stemming from market power.

(ii) The Pigouvian component corrects the externality from market power, which results in insufficient output due to high prices. Thus, the tax formulas incorporate a Pigouvian correction, encouraging policymakers to subsidize both labor and entrepreneurial effort. Notably, the Pigouvian components differ between labor and profits taxes. The labor income tax's Pigouvian component depends on the *average* markup weighted by factor inputs, while the profit tax's Pigouvian component is determined by firm-level markups. Intuitively, the average markup measures the efficiency of labor supply, whereas firm-level markups reflect the insufficient labor inputs and entrepreneurial effort of each firm.

(iii) The RE component applies only to profit taxes. Total output increases when labor inputs shift from low-markup to high-markup firms. Thus, the RE is determined by the gap between firm-level markup and average markups. In competitive markets, this reallocation of factors has no first-order effect on total output.

(iv) The IRE allows policymakers to exploit price competition between firms to adjust the primary distribution of profits through market incentives. Lowering marginal tax rates increases output, which subsequently reduces competitor prices. Interestingly, this IRE emerges only under strategic interaction where firms have incomplete pricing power. It vanishes in monopolistic markets, where a single firm dominates, and pricing externalities disappear. This distinction highlights the role of market structure in policy design.

These four components illustrate how market power shapes optimal tax policy. Omitting any of these elements can lead to substantially different conclusions, distinguishing our results from prior research. The net effect on labor income taxes combines the Pigouvian correction of markup externalities with the Mirrleesian tradeoff between production and redistribution. For entrepreneurial incomes, the skill gap widens as markups rise, amplifying the Mirrleesian component and increasing tax rates, especially for top incomes. Similarly, the net effect on profit taxes depends on the interplay between the IRE[4] and the RE, which incentivizes high-skill entrepreneurial production. We formalize these theoretical insights in a series of theorems and propositions, deriving explicit tax wedges in terms of economic primitives. We also present statistics-based formulas in the appendix.

One key contribution of this paper is linking optimal tax formulas to empirical data. Using the normative analysis, we provide precise guidance for policymakers on addressing the rise in market power from 1980 to 2019. Our *quantitative analysis* estimates model parameters and quantifies optimal tax rules, decomposing

[3] In the benchmark, we focus on a tax system consisting of non-linear taxes on the income of workers and entrepreneurs and a linear tax on the sales of consumer goods. However, we can restrict our attention exclusively to the taxes on entrepreneurs and workers, disregarding the sales tax. It is well known (e.g., Chari and Kehoe (1999) and Golosov et al. (2003)) that multiple tax policies can achieve the same second-best outcome. In our setting, a linear sales tax can be substituted with a uniform tax on labor income and profits. Therefore, we assume that sales taxes are zero.

[4] The government can reduce the income gap between entrepreneurs by lowering the prices of products produced by high-skill entrepreneurs. To achieve this, it decreases the profit tax rate for high-skill entrepreneurs.

them into the four identified components. These results underscore the central role of market structure and endogenous markups in shaping optimal tax policy.

Our model prescribes a 10 percentage point decline in labor income tax rates from 1980 to 2019. Conversely, profit tax rates rise by 3 percentage points on average and by 25 percentage points at the top. Policymakers trade off equity and efficiency, making the profit tax more progressive overall while remaining regressive for large firms. Regressive profit taxes promote efficient production reallocation toward more productive firms.

The quantitative exercise allows us to link the positive findings to the various normative forces in the model, thereby quantifying the different components of the planner's tradeoff between redistribution and efficiency. In particular, we find that the decline in the labor income tax rate is driven by the rise in average markups, which reduces the Pigouvian component and, consequently, labor income taxes. This Pigouvian effect also influences the profit tax, but the other three components play a more dominant role. The ambiguous effect of market power on the profit tax rate in the normative analysis is resolved in our estimated quantitative model.

We find that the overall impact of the rise in average markups and the corresponding increase in the dispersion of markups between 1980 and 2019 results in a *higher profit tax rate*. The Pigouvian and reallocation components exert a negative effect on the top tax rates,[5] whereas changes in the indirect redistribution effect and the Mirrleesian component increase the top profit tax rates. Among these, the Mirrleesian effect on the profit tax rate is the strongest, outweighing the influence of the other components and leading to an overall increase in the profit tax rate.

We finalize our analysis with a robustness exercise, examining three alternative specifications of our baseline model: introducing non-linear sales taxes, allowing the planner to condition taxes on markups or outputs, and incorporating capital investment. This analysis demonstrates that our results are robust across these variations in the model setup.[6]

Relation to the Literature. Starting with Mirrlees (1971), an extensive and influential literature on optimal taxation has analyzed the factors that determine the properties of income tax schedules. Within this literature, our paper contributes to three strands: (i) market power and optimal policies; (ii) endogenous pricing and optimal taxation; and (iii) externalities and optimal taxation. Additionally, our paper is related to the literature on optimal taxation and entrepreneurship, as well as optimal taxation and technology.

(i) In recent years, a growing policy literature has analyzed the relationship between markups and inequality (e.g., Stiglitz (2012); Atkinson (2015); Baker and Salop (2015); Khan and Vaheesan (2017)). Our paper differs from existing research in several respects. First, most existing papers consider representative agents. These studies 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 lump-sum taxes are unenforceable and study how governments can raise rev-

[5] Higher markups require correcting the externality and reallocating resources toward more productive firms.

[6] An important 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 first-best because the markup is endogenous to the unobservable decisions of entrepreneurs, which respond to the planner's tax schedule. In other words, there is an incentive constraint that the planner must account for when determining the optimal tax rate, even when conditioning on markups.

enue efficiently. In contrast, we examine optimal taxation in the spirit of Mirrleesian taxation, highlighting the trade-off between efficiency and equality.

Second, in studies that consider redistribution in the presence of market power, the equilibrium often assumes monopolistic competition (see, e.g., Gürer (2021) and Boar and Midrigan (2019)). Our framework features oligopolistic competition modeled as in Jaimovich and Floetotto (2008), a variation on Atkeson and Burstein (2008). This setting allows us to analyze how market structure influences the optimal design of the tax system, a topic largely absent in the literature.

Current studies of oligopolistic markets generally do not consider strategic pricing and profit tax design (see, e.g., Kushnir and Zubrickas (2019) and Jaravel and Olivi (2019)). In these studies, agents treat prices and profits as given. In contrast, we assume entrepreneurs strategically set prices based on their reported types. Consequently, the role of market power in our principal-agent problem is notably distinct from that in the existing literature.

The market structure and agents' pricing power are not merely theoretical curiosities; they have important implications for concrete policy recommendations. Unlike previous studies (see, e.g., Kaplow (2019) and Boar and Midrigan (2019)), our formula suggests that even without changes in social welfare weights, rising market power may increase the optimal profit tax rate. One important reason is that it alters the Mirrleesian rule.

(ii) The secondary literature on optimal taxation and endogenous prices (wages) typically assumes competitive markets (see, e.g., Stiglitz (1982); Naito (1999); Naito (2004); Saez (2004); Rothschild and Scheuer (2013); Scheuer (2014); Sachs et al. (2020); Cui et al. (2021)). This literature emphasizes the general equilibrium effects of taxes on factor prices, leading to indirect redistribution between agents providing different factors. We demonstrate that the indirect redistribution effect (IRE) critically depends on market structure. Lowering the profit tax encourages entrepreneurial effort and output, reducing competitors' product prices and leading to indirect redistribution. This IRE vanishes when there is no competition, as in monopolistic markets.

(iii) The third strand of research examines 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)). A notable finding in this literature is the "additivity property" (see Kopczuk (2003)),[7] which states that the Pigouvian component is separable from other elements of the optimal tax, suggesting that rising markups decrease the optimal tax rate. In our model, the additivity property does not hold. Markups enter other components of the optimal tax in addition to the Pigouvian component, and do not necessarily decrease the optimal tax.

It is also worth noting that other factors besides market power introduce externalities. For instance, Rothschild and Scheuer (2013) and Scheuer (2014) study optimal taxation in Roy models, which include both RE and IRE in their tax formulas. Additionally, Rothschild and Scheuer (2014) and Rothschild and Scheuer (2016) incorporate a Pigouvian effect, with externalities arising from the gap between the market and social value of occupations.

Beyond the three strands discussed, our paper also relates to the literature on optimal taxation and entrepreneurship (see, e.g., Scheuer (2014); Ales and Sleet (2016); Ales et al. (2017); Scheuer and Werning

[7] The additivity property is a special case of the "principle of targeting" proposed by Dixit (1985).

(2017)). Among these, Boar and Midrigan (2019) introduces market power but studies an alternative incentive problem where profit taxes do not affect entrepreneurial incentives. Their setup overlooks the influence of market power on the Mirrleesian component. Moreover, their policies, production technologies, and market structures differ from ours.[8]

Lastly, our paper 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 (i.e., the curvature of firm-level production with respect to labor inputs). Our results extend their findings to monopolistically competitive economies but suggest their result no longer holds in oligopolistic economies. The span of control influences the parametric tax rule by magnifying the indirect redistribution effect.

The paper is organized as follows. Section 2 presents the model setup, the equilibrium definition, and the planner's problem. Our main results are in Section 3. Section 4 discusses the robustness of the main results. Section 5 contains the numerical analysis. We conclude in Section 6.

2 The Model

We first lay out the model setup in Section 2.1. Then, in Section 2.2 we derive the agents' best responses, the market clearing conditions, and define equilibrium, along with the properties of the Laissez-faire economy. In Section 2.3 we present the solution to the planner's problem.

2.1 Setup

Environment. The economy is static. Agents belong to one of two occupations, $o \in e, w$: entrepreneur or worker. Occupational types are predetermined. The measure of workers is N_w; the measure of entrepreneurs, N_e, is normalized to one. A representative firm produces final goods in a competitive market and makes zero profits. Production of final goods requires a composite input of firm-level intermediate goods. Each intermediate good is produced by an entrepreneur (idea) and the effort of workers.

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 $f_o(\theta_o)$. Let $x_o(\theta_o)$ represent the efficiency labor provided by θ_o per unit of time. It is convenient to order ability on the unit interval and assume a 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, this assumption about the distribution of ability entails no loss of generality.

Preferences. Workers and entrepreneurs derive utility from consumption and effort. We assume a quasi-linear utility function:[9]: $u_o(c_o, l_o) = c_o - \phi_o(l_o)$, where l_o refers to working hours and c_o to consumption. $\phi_o(\cdot)$ is twice continuously differentiable and strictly convex. To simplify, we assume a utility function with constant elasticity of labor supply. Let $\varepsilon_o \equiv \frac{\phi_o'(l_o)}{l_o \phi_o''(l_o)}$ denote the elasticity of labor supply. Denote by $V_o(\theta_o)$ the indirect utility of an agent with type θ_o.

[8] They study quantity regulation rather than profit taxes. Market power in our model arises from the number of firms in oligopolistic competition, unlike their model, which uses preferences (via the Kimball aggregator) under monopolistic competition. These differences yield novel policy implications.

[9] The quasi-linear utility function eliminates the income effect and the complementarity between consumption and labor. This assumption simplifies the analysis and is not crucial to the paper's main economic implications. See Atkinson and Stiglitz (1976), Mirrlees (1976), and Christiansen (1984) for discussions on how omitted elements affect optimal taxation.

Markets. Each firm is indexed by a triple (i, j, θ_e), where $\theta_e \in \Theta_e$ is the entrepreneur's ability, $j \in [0, J(\theta_e)]$ represents the market order, and $i = 1, ..., I$ indicates the firm's order in the most granular market. $I \in \mathbb{N}+$ is the number of firms in a granular market, and $J(\theta_e) \in \mathbb{R}+$ is the measure of markets for θ_e entrepreneurs. A differentiated input is fully identified by (i, j, θ_e). Intermediate goods produced by I entrepreneurs of type θ_e in the same market are used to produce market-level intermediate goods (θ_e, j). The final good aggregates intermediate goods across θ_e and j. The labor and final goods markets are perfectly competitive, while the intermediate goods market exhibits oligopolistic competition.

Market Structure and Oligopolistic Competition. A key feature of our setup is the strategic interaction between firms in the intermediate goods market, which compete oligopolistically. This strategic interaction has fundamental implications for optimal taxation. We derive general conclusions from a generic specification of market structure and technology.

For certain analytical results, we specify the production technology and market structure explicitly. We model the input market as in Jaimovich and Floetotto (2008), a variation of the nested-CES structure in Atkeson and Burstein (2008), with product differentiation in production rather than preferences.

Technology. An agent of ability θ_o who works l_o hours supplies $x_o(\theta_o)l_o$ units of effective labor in the intermediate goods sector.[10] 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). Here, W is the competitive wage rate paid for one efficiency unit of labor.[11] Firm-level output $Q_{ij}(\theta_e)$ is a function of entrepreneurial effort $l_e(\theta_e)$ and labor inputs $L_w(\theta_e)$. The profit of firm (i, j, θ_e) is given by:

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

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

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 function highlights the novelty of our setup: 1. Entrepreneurs select output strategically to influence the prices of their own products; 2. Entrepreneurs competing in the same market affect each other's product prices.

The profit tax impacts firm-level profit both through its effect on the firm's behavior and on the behavior of its competitors. This setup distinguishes our paper from the classic studies by Mirrlees (1971) and Stiglitz (1982). It also sets our work apart from the current literature on optimal taxation under monopolistic competition, where firms have monopoly pricing power in granular markets, yet there is no strategic interaction (see, e.g., Gürer (2021) and Boar and Midrigan (2019)). This strategic interaction proves crucial for determining the optimal tax rate, both theoretically and quantitatively.

[10] The efficiency units assumption simplifies the solution but is not innocuous. It rules out sorting, as firms are indifferent to worker types if they provide the same efficiency units. See Sattinger (1975a), Sattinger (1993), and Eeckhout and Kircher (2018) on how efficiency units imply no sorting. Solving the optimal taxation problem with market power and sorting remains an open challenge.

[11] Labor supplied by different worker abilities is assumed to be perfectly substitutable. For imperfect substitutability, see Sachs et al. (2020) and Cui et al. (2021).

The firm-level production technology for intermediate goods follows Lucas (1978), with a heterogeneous entrepreneur hiring an endogenous number of workers to maximize profits. Since the productivities of entrepreneurs and workers are expressed in efficiency units, the technology uses efficiency units as inputs instead of physical labor. The output of a θ_e entrepreneur is:[12]

$$Q_{ij}(\theta_e) = x_e(\theta_e) l_{e,ij}(\theta_e) \cdot L_{w,ij}(\theta_e)^{\xi}, 0 < \xi \leq 1. \tag{2}$$

There is no capital in our model. Thus, as in Lucas (1978) or Prescott and Visscher (1980), the entrepreneur is the residual claimant, i.e., the entrepreneur "owns" the technology θ_e and hires labor to maximize profits.

For some results, we assume a specific *nested-CES technology*. The technology aggregates a finite number I of close substitutes (e.g., Coke and Pepsi) in the same market with elasticity $\eta(\theta_e)$ to $Qj(\theta_e)$, then across a continuum $J(\theta_e)$ of markets with elasticity σ to $Q(\theta_e)$, and finally across a continuum of less substitutable input goods $Q(\theta_e), \theta_e \in \Theta_e$ (e.g., soft drinks and cars) with elasticity σ to the final good Q:[13]

$$Q_j(\theta_e) = \left[I^{-\frac{1}{\eta(\theta_e)}} \sum_{i=1}^{I} Q_{ij}(\theta_e)^{\frac{\eta(\theta_e)-1}{\eta(\theta_e)}} \right]^{\frac{\eta(\theta_e)}{\eta(\theta_e)-1}}, \tag{3}$$

$$Q(\theta_e) = \left[J(\theta_e)^{-\frac{1}{\sigma}} \int_j Q_j(\theta_e)^{\frac{\sigma-1}{\sigma}} dj \right]^{\frac{\sigma}{\sigma-1}}, \tag{4}$$

$$Q = A \left[\int_{\theta_e} \zeta(\theta_e) Q(\theta_e)^{\frac{\sigma-1}{\sigma}} d\theta_e \right]^{\frac{\sigma}{\sigma-1}}. \tag{5}$$

The elasticity of substitution σ across markets (e.g., between soft drinks and cars) is smaller than within markets (e.g., between Coke and Pepsi): $\sigma < \eta(\theta_e)$. To rule out abnormal markups, we assume $\sigma > 1$ throughout. $\zeta(\theta_e)$ is a distribution parameter. As illustrated by Ales et al. (2015), variations in $\zeta(\theta_e)$ capture technological or preference-based demand variations for different skills. To abstract from the love-of-variety effect related to I, we normalize the firm-level and market-level outputs by $I^{-\frac{1}{\eta(\theta_e)}}$ and $J(\theta_e)^{-\frac{1}{\sigma}}$.[14] We then introduce $\chi(\theta_e) = \zeta(\theta_e) f_e(\theta_e)^{-\frac{1}{\sigma}}$ as a modified distribution parameter. Finally, define:

$$X_e(\theta_e) = A^{\frac{\sigma-1}{\sigma}} x_e(\theta_e)^{\frac{\sigma-1}{\sigma}} \chi(\theta_e) \tag{6}$$

as the composite ability of θ_e entrepreneur.[15] The components $\chi(\theta_e)$ and $x_e(\theta_e)$ are determinants of firm productivity. Although equilibrium allocations are determined by composite productivity, $\chi(\theta_e)$ and $x_e(\theta_e)$ are not perfect substitutes, as equilibrium prices depend on their specific values. $\chi(\theta_e)$ enters the demand function and interacts with the markup, while $x_e(\theta_e)$ does not. Importantly, composite ability determines optimal tax formulas (see Proposition 2). For general results, see equation (38) and Theorem 1.

Information and Policy. In the tradition of Mirrleesian taxation, we assume that types θ_o and effort l_o are unobservable, while incomes y_w and y_e are observable. As additional actions are introduced, further clarification of information is desirable. In particular, we assume that factor inputs L_w are also observable. Thus, the unobservability of θ_e is equivalent to saying that firm-level output Q_{ij} and price P_{ij} cannot both

[12] 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. Moreover, because all goods are imperfect substitutes, there is no indeterminacy in the firm size.

[13] The elasticities of substitution between intermediates across markets are uniform. In more general settings, elasticities may be asymmetric, affecting how endogenous wages respond to tax policy (see Rothschild and Scheuer (2013) and Sachs et al. (2020)).

[14] The measure of markets $J(\theta_e) = \frac{Nef(\theta_e)}{I}$ also represents the measure of varieties provided by θ_e entrepreneurs.

[15] Online Appendix OA.2.3 shows that equilibrium allocations depend on composite ability, while prices depend on its structure.

be observed. Otherwise, the government could infer the type of entrepreneur from the relationship $P_{ij} = P_{ij}(Q_{ij}, Q_{-ij}{}_{-i\neq i}, \theta_e)$.

There are several justifications for this information constraint. Effective output Q_{ij}, like effort l_o, is difficult to measure objectively. Although working hours can be observed, effort l_o is hard to gauge because the intensity of work cannot be measured objectively. Similarly, while product quantities may be counted, quality—and thus effective output—remains difficult to assess.[16]

The government can impose arbitrarily non-linear profit and labor income taxes, $T_e : \mathbb{R}+ \mapsto \mathbb{R}$ and $T_w : \mathbb{R}+ \mapsto \mathbb{R}$, following the Mirrlees tradition. Together with a sales tax $t_s \in \mathbb{R}$, these direct taxes form the tax policy system $\mathcal{T} \equiv \{T_e, T_w, t_s\}$, which we consider in our benchmark model. Since labor and profit income taxes are flexible, 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 imposing a linear sales tax, as this is commonly used in practice.[17] Therefore, the government aims for a third-best allocation rather than a second-best, under both information and policy constraints. In Section 4, we extend the model to allow for non-linear sales taxes and compare the third-best with the second-best allocation.

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

$$\sum_{o\in\{w,e\}} N_o \int_{\theta^o} G\left(V_o(\theta_o)\right) \widetilde{f}_o\left(\theta_o\right) d\theta_o, \tag{7}$$

subject to the budget constraint

$$N_e \int_{\theta_e} T_e\left(y_e\left(\theta_e\right)\right) f_e\left(\theta_e\right) d\theta_e + N_w \int_{\theta_w} T_w\left(y_w\left(\theta_w\right)\right) f_w\left(\theta_w\right) d\theta_w = R \tag{8}$$

and agents' responses to the taxes. Here, $R \in \mathbb{R}_+$ represents exogenous tax revenue. The social welfare function $G : \mathbb{R}_+ \longmapsto \mathbb{R}_+$ is twice differentiable and concave. The PDF $\widetilde{f}_\theta(\cdot)$ is a Pareto weights schedule, assumed to be continuous.[18]

Next, in Section 2.2, we set up and define equilibrium, deriving key concepts such as markups, elasticities, and the labor share, as well as the main properties of the laissez-faire equilibrium. Equilibrium is defined for a given tax schedule. In Section 2.3, we present the planner's problem and explain how the planner chooses the optimal tax schedule given the equilibrium in the economy.

2.2 Equilibrium

Below, we formally define equilibrium after laying out all agents' best responses and market-clearing.

Final Goods Market. We begin with the final goods market, where the price of the final good is normalized to one. The final goods producer chooses the inputs of intermediate goods to maximize its profit. The demand for the intermediate input solves:

---

[16] In reality, many other factors complicate the observation of prices and yields. For example, there may be collusion between companies or between companies and buyers to hide prices or efficiencies to reduce tax burdens.

[17] Alternatively, one could consider a linear tax on labor inputs (e.g., a salary tax). Both the linear sales tax and the linear salary tax act as tax wedges between the marginal cost and income of labor inputs $L_{w,ij}$. Since optimal taxation focuses on tax wedges rather than specific tax policies (see, e.g., Chari and Kehoe (1999); Golosov et al. (2003); Salanié (2003), pages 64-66), introducing both indirect taxes is unnecessary. For instance, in equation (12), if an additional tax t_l is levied on labor inputs, the ratio of marginal income to marginal cost of $L_{w,ij}$ becomes $\frac{1+t_l}{1-t_s}$, meaning t_l's role as a tax wedge can be replaced by t_s.

[18] The use of general Pareto weights in the optimal tax literature dates back to Diamond and Mirrlees (1971a) and Diamond and Mirrlees (1971b), and is applied in the context of our model in Scheuer (2014).

$$\Pi = \max_{Q_{ij}^D(\theta_e)} Q - \int_{\theta_e} \int_j \left[\sum_i Q_{ij}^D(\theta_e) P_{ij}(\theta_e) \right] dj d\theta_e, \tag{9}$$

where $Q_{ij}^D(\theta_e)$ is the quantity demanded from firm (i, j, θ_e).

Entrepreneurs. In our benchmark model, we consider a Cournot Competitive Tax Equilibrium in intermediate goods market j with I firms. There is a continuum of intermediate goods markets j and θ_e, so there strategic interaction only within market j, and all firms treat output decisions in other markets as given.

The problem for the entrepreneur in firm (i, j, θ_e) is:

$$V_{e,ij}(\theta_e) \equiv \max_{l_{e,ij}, L_{w,ij}} c_e - \phi_e(l_e) \tag{10}$$

$$\text{s.t. } c_{e,ij} = y_{e,ij} - T_e(y_{e,ij}) \tag{11}$$

$$y_{e,ij} = (1 - t_s) P_{ij} \left(Q_{ij}, \{Q_{-ij}(\theta_e)\}_{-i \neq i}, \theta_e \right) Q_{ij} - W L_{w,ij}, \tag{12}$$

where $Q_{ij} = Q_{ij}(x_e(\theta_e) l_e, L_w)$ is the quantity of intermediate goods supplied. 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 entrepreneur (i, j, θ_e), respectively.

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

$$V_w(\theta_w) \equiv \max_{l_w} c_w - \phi_w(l_w) \tag{13}$$

$$\text{s.t. } c_w = W x_w(\theta_w) l_w - T_w(W x_w(\theta_w) l_w). \tag{14}$$

We 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. The commodity and labor markets clear when the quantity demanded in the output sector $Q_{ij}^D(\theta_e)$ from equation (9) equals the quantity supplied $Q_{ij}^S(\theta_e)$ from equation (10):

$$Q_{ij}^D(\theta_e) = Q_{ij}^S(\theta_e) \tag{15}$$

$$\text{and} \quad Q = \int_{\theta_w} c_w(\theta_w) f_w(\theta_w) d\theta_w + \int_{\theta_e} \int_j \left[\sum_i c_{e,ij}(\theta_e) \right] dj d\theta_e + R, \tag{16}$$

$$\text{and} \quad \int_{\theta_w} x_w(\theta_w) l_w(\theta_w) f_w(\theta_w) d\theta_w = \int_{\theta_e} \int_j \left[\sum_i L_{w,ij}(\theta_e) \right] dj d\theta_e, \tag{17}$$

where R is exogenous government revenue.

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

$$P_{ij}(\theta_e) = \frac{\partial Q}{\partial Q_{ij}(\theta_e)}, \tag{18}$$

$$W = (1 - t_s) \frac{\partial \left[P_{ij}(\theta_e) Q_{ij}(\theta_e) \right]}{\partial L_{w,ij}(\theta_e)}, \tag{19}$$

$$\phi_w'(l_w(\theta_w)) = W x_w(\theta_w) \left[1 - T_w'(W x_w(\theta_w) l_w(\theta_w)) \right], \tag{20}$$

$$\phi_e'(l_{e,ij}(\theta_e)) = (1 - t_s) \frac{\partial \left[P_{ij}(\theta_e) Q_{ij}(\theta_e) \right]}{\partial Q_{ij}(\theta_e)} \frac{\partial Q_{ij}(\theta_e)}{\partial l_{e,ij}(\theta_e)} \left[1 - T_e'(y_{e,ij}(\theta_e)) \right]. \tag{21}$$

When the first-order conditions are both necessary and sufficient for solving the individuals' and the final goods producer's problems, the equilibrium allocations are determined by (15) through (21) and the individuals' budget constraints.

Equilibrium. Throughout this paper, we 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 schedules $L_w : \Theta_w \longmapsto \mathbb{R}+$, which are independent of (i, j). Prices $\mathcal{P} = \{P, W\}$ in equilibrium consist of the wage rate W and the price schedule $P : \Theta_e \longmapsto \mathbb{R}+$, both independent of (i, j). Formally, we define 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 (9); maximizes the entrepreneurs' and workers' utilities (10) and (13) subject to the budget constraints (11) and (14); the price system satisfies (19) and (18); and labor and commodity markets are cleared, i.e., (15) to (17) are satisfied.*

Note that, due to Walras's law, we do not need to explicitly impose the government's budget constraint in our definition of SCCTE. Given the agents' budget constraints and market-clearing conditions, the government's budget constraint will automatically be satisfied.

We now introduce some common restrictions on the equilibrium that we consider throughout the paper. First, we assume that the mechanisms (tax policies) are twice continuously differentiable and that the first-order conditions are both necessary and sufficient for the agents' optimization problems. This is a common assumption in the optimal tax literature, effectively ensuring that the optimal tax schedules are not excessively regressive (see, e.g., Jacquet et al. (2013)). We will demonstrate the sufficiency of first-order conditions under our leading production technology (see equations (2) to (5)). 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 with respect to wages.[19] For simplicity, we assume that $y_w(\theta_w)$ is strictly increasing in θ_w, which in turn implies $x'_w(\cdot) > 0$. Assumption 1 also excludes cases with mass points.

In what follows, where there is no ambiguity, we will drop the subscript ij. For example, in a symmetric equilibrium, the markup is identical for all entrepreneurs of the same type. Thus, we often denote the markup $\mu_{ij}(\theta_e)$ as $\mu(\theta_e)$ and the labor demand $L_{w,ij}(\theta_e)$ as $L_w(\theta_e)$. The inverse demand function, accounting for strategic interactions between firm i and its competitors $-i$, simplifies to $P_{ij}(Q_{ij}, Q_{ij}(\theta_e), \theta_e)$, assuming competitors produce the same quantity of products.

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

$$\varepsilon^{P,own}_{Q_{ij}}(\theta_e) \equiv \frac{\partial \ln P_{ij}\left(Q_{ij}, Q_{ij}(\theta_e), \theta_e\right)}{\partial \ln Q_{ij}}\Big|_{Q_{ij}=Q_{ij}(\theta_e)} \text{ and } \varepsilon^{P,cross}_{Q_{-ij}}(\theta_e) \equiv \frac{\partial \ln P_{ij}\left(Q_{ij}, Q_{ij}(\theta_e), \theta_e\right)}{\partial \ln Q_{ij}(\theta_e)}\Big|_{Q_{ij}=Q_{ij}(\theta_e)}. \quad (22)$$

[19]See, e.g., Salanié (2003), p. 87.

The cross-inverse demand elasticity measures the elasticity of an entrepreneur's inverse demand function with respect to their competitor's output. The presence of cross-inverse demand elasticity highlights the key difference between Cournot competition and monopoly. In markets with strategic interaction, a firm's demand depends not only on its own output but also on the output decisions of its competitors.

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

$$\mu_{ij}(\theta_e) \equiv \frac{P_{ij}(\theta_e)}{MC_{ij}(\theta_e)} = \frac{P_{ij}(\theta_e)}{\frac{W}{\frac{\partial Q_{ij}(\theta_e)}{\partial L_{w,ij}(\theta_e)}(1-t_s)}}. \tag{23}$$

The Lerner Rule establishes the relationship between inverse-demand elasticity and markups:

$$\mu_{ij}(\theta_e) = \frac{1}{1+\varepsilon_{Q_{ij}}^{P,own}(\theta_e)}. \tag{24}$$

The higher the demand elasticity, the lower the markup. Lastly, we define an economy-wide aggregate markup, which proves to be an important factor in optimal taxation:

$$\mu \equiv \int_{\theta_e} \mu(\theta_e)\,\omega(\theta_e)\,d\theta_e, \text{ where } \omega(\theta_e) \equiv \frac{L_w(\theta_e) f_e(\theta_e)}{\int_{\theta_e} L_w(\theta_e) f_e(\theta_e)\,d\theta_e}. \tag{25}$$

The Laissez-faire Economy. We now examine the effect of market power on the equilibrium allocation in the laissez-faire economy, i.e., when government revenue R is zero and no taxes are levied. The laissez-faire economy serves as a benchmark, highlighting how market power influences key equilibrium outcomes: labor share, output, wages, and profits. This analysis helps us understand the model's mechanics before introducing optimal taxation. It also validates the model setup underlying the tax design and provides intuition for parameter selection in the quantitative analysis. We find that, under common parameter values, the model reproduces crucial features associated with the rise of market power, as empirically documented.

In our model, a firm's labor share is defined as the ratio of its total wage bill to its revenue.[20] Denote the labor share of firm as:

$$\nu_{ij}(\theta_e) \equiv \frac{WL_{w,ij}(\theta_e)}{P_{ij}(\theta_e)\,Q_{ij}(\theta_e)\,(1-t_s)}. \tag{26}$$

At first glance, this expression suggests a positive relationship between the sales tax rate t_s (an increase in t_s increases the labor share). However, taxes also affect other endogenous variables such as $L_{w,ij}$, P_{ij}, and Q_{ij}. Using the firm's first-order condition, we can rewrite the labor share as:

$$\nu_{ij}(\theta_e) = \frac{\xi}{\mu_{ij}(\theta_e)}, \tag{27}$$

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

$$\nu \equiv \frac{W\int x_w(\theta_w)\,l_w(\theta_w)\,f_w(\theta_w)\,d\theta_w}{Q}. \tag{28}$$

Although the firm-level labor share is independent of the tax rate, the aggregate labor share depends on it.

Before introducing the planner's problem and solution, we summarize the properties of the laissez-faire economy in Proposition 1:

Proposition 1 *In the nested-CES economy:*

(i) The firm labor share $\nu_{ij}(\theta_e)$ decreases in the markup $\mu_{ij}(\theta_e)$;

[20] In the absence of capital, the residual income is attributed to the entrepreneur, i.e., the profit share.

(ii) In the Laissez-faire economy, the aggregate labor share ν decreases in market power (increase in I), when

$$\frac{1}{1+\varepsilon_e}+\frac{1}{\sigma-1}>\xi. \tag{29}$$

(iii) Consider the Laissez-faire economy with constant $\eta(\theta_e)$. Under conditions (29) and (30)

$$\frac{1+\varepsilon_w}{\varepsilon_w}-\xi\left(1+\varepsilon_e\right)>0, \tag{30}$$

we have the following results with the decrease of I (rise of market power).

(a) The output Q, wage rate W, and aggregate labor share ν decrease;

(b) 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 entrepreneur utility increases 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.3. ■

The parameter restriction (29) is intuitive. It ensures that as I increases (markups decrease), the entrepreneurial effort of smaller firms rises relative to that of larger firms, preventing excessive market concentration and a consequent increase in average markups. Condition (30) guarantees that the demand elasticity of labor is smaller than the supply elasticity of labor. This implies that labor input decreases as markups rise, or equivalently, that the equilibrium wage increases with TFP.[21] These two parameter restrictions are relatively weak and are satisfied within the range of values commonly used in the quantitative literature.

Part (i) of Proposition 1 confirms a well-known theoretical property: firms with higher individual markups tend to have a lower labor share. Higher markups indicate that the firm sells and produces fewer units, reducing its demand for labor inputs and, consequently, its labor share. The negative markup-labor share relationship is documented empirically by De Loecker et al. (2020) and Autor et al. (2020).

Part (ii) extends this insight, showing that under common conditions, the aggregate labor share declines as competition decreases (i.e., as I falls).

Part (iii) of Proposition 1 establishes that an increase in market power negatively affects the aggregate economy: the wage rate, aggregate output, sales, and labor share all decline. Additionally, Part (iii) identifies the conditions under which firm-level profits increase as competition decreases. The conditions for increased profits and utility presented in Proposition 1 align with typical values found in the literature.[22]

2.3 The Planner's Problem

The planner's problem can be approached in 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, subject to information, resource, and policy constraints.[23] A

[21] See Online Appendix equation (OA15) for details.

[22] 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 condition for increasing utility is satisfied when $\mu<2$. If ξ increases to 0.6, these conditions loosen, with the first and second thresholds rising to $\mu<1.5$ and $\mu<2.125$, respectively. Overall, under common parameters, the rise in market power has negative effects on workers, lowering their income and utility while benefiting entrepreneurs by increasing both their profits and utility.

[23] In our benchmark, we consider the planner's third-best solution, where the sales tax is restricted to be uniform. This constraint ensures that the tax wedges between the marginal cost and benefit of labor inputs are uniform across firms.

worker who reports θ_w' receives labor income $y_w(\theta_w')$, resulting in after-tax income $c_w(\theta_w')$. Similarly, an entrepreneur who reports θ_e' receives profit $y_e(\theta_e')$ and after-tax profit $c_e(\theta_e')$.

Following standard practice in the literature, we apply the first-order approach to simplify the incentive constraints and solve the planner's problem. See Online Appendices OB.1 and OB.2 for the validity of this approach.[24]

Worker. The incentive-compatibility conditions for workers follow the standard framework in the Mirrleesian tax literature:

$$V_w'(\theta_w) = l_w(\theta_w)\phi_w'(l_w(\theta_w))\frac{x_w'(\theta_w)}{x_w(\theta_w)}, \forall \theta_w \in \Theta_w, \tag{31}$$

where $V_w(\theta_w) = c_w(\theta_w) - \phi_w\left(\frac{y_w(\theta_w)}{x_w(\theta_w)W}\right)$. When the Spence-Mirrlees condition is satisfied, the first-order incentive condition is not only necessary but also sufficient.[25] Further clarification is required for the entrepreneur's incentive-compatible condition.

Entrepreneur. In the subsequent analysis, we provide conditions under which a SCCTE exists where all firms report truthfully. This requires that, given other entrepreneurs report their true types, any individual entrepreneur also finds it optimal to report truthfully. Formally, an entrepreneur chooses a reporting type θ_e' to maximize:

$$V_e(\theta_e) = \max_{\theta' \in \Theta_e} V_e(\theta_e'|\theta_e), \tag{32}$$

where $V_e(\theta_e'|\theta_e) = c_e(\theta_e') - \phi_e(l_e(\theta_e'|\theta_e))$ is the utility of the θ_e entrepreneur who reports θ_e' and $l_e(\theta_e'|\theta_e)$ is the entrepreneurial labor supply needed to finish the task:[26]

$$l_e(\theta_e'|\theta_e) = \min_{L_w, l_e} l_e \tag{33}$$

$$\text{s.t. } (1-t_s) P_{ij}(Q_{ij}, Q_{ij}(\theta_e), \theta_e) Q_{ij} - WL_w = y_e(\theta_e').$$

In Online Appendix OB.2, we show when the first-order conditions of the above problem are both necessary and sufficient. 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 the following:*

$$V_e'(\theta_e) = \phi_e'(l_e(\theta_e)) l_e(\theta_e)\left[\mu(\theta_e)\frac{d\ln P_{ij}(Q_{ij}, Q_{ij}(\theta_e), \theta_e)}{d\theta_e}|_{Q_{ij}=Q(\theta_e)} + \frac{x_e'(\theta_e)}{x_e(\theta_e)}\right]. \tag{34}$$

Proof. See Appendix B.1. ■

Lemma 1 is useful because it simplifies the incentive-compatible constraint to condition (34). Moreover, it highlights what is new in our model. Three points are worth noting here:

First, compared to the incentive condition in the canonical Mirrlees (1971) taxation framework (see the worker's incentive condition (31)), an additional term appears in the entrepreneur's incentive condition. This term captures the influence of market power and endogenous pricing. Intuitively, the gross utility of an entrepreneur depends not only on effort $l_e(\theta_e)$ and labor capacity $x_e(\theta_e)$, but also on the price of the product $P_{ij}(\theta_e)$ and market power $\mu(\theta_e)$.

[24] Additionally, in Online Appendix OB.1, we show how the constrained optimal allocation can be implemented using the tax system $\mathcal{T}$. Furthermore, we demonstrate that t_s is redundant in implementing the constrained optimal allocation, allowing us to focus on a tax system with $t_s = 0$.

[25] See Salanié (2003) p.88-90 for details.

[26] In what follows, we denote by $L_w(\theta_e'|\theta_e)$ the solution to (33), and let $Q_{ij}(\theta_e'|\theta_e) = Q_{ij}(x_e(\theta_e) l_e(\theta_e'|\theta_e), L_w(\theta_e'|\theta_e))$ and $P_{ij}(\theta_e'|\theta_e) = P_{ij}\left(Q_{ij}(\theta_e'|\theta_e), Q_{ij}(\theta_e), \theta_e\right)$.

Second, the traditional indirect redistribution route is closed. To see this, note that the influence of price on gross utility depends on $\frac{d\ln P_{ij}\left(Q_{ij},Q_{ij}(\theta_e),\theta_e\right)}{d\theta_e}|Q_{ij}=Q_{ij}\left(\theta_e\right)$ instead of $\frac{d\ln P(\theta_e)}{d\theta_e}$. Since the entrepreneur sets prices by choosing the firm-level output Q_{ij}, the tax has no first-order effect on prices through changes in the firm's own output Q_{ij}. Therefore, the traditional indirect redistribution route is closed.

Third, the tax exhibits an indirect redistribution effect through its influence on the competitor's output. To see this, note that

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

This equation highlights the difference between Cournot competition and monopoly. In a submarket with strategic interaction among competitors, prices depend on competitors' outputs (see the cross-inverse demand elasticity). This introduces a novel indirect redistribution route.

Relaxed Planner's Problem. In the relaxed planner's problem, the incentive constraints are replaced by first-order incentive conditions. The planner then chooses the allocation $\{c_w(\theta_w),l_w\left(\theta_w\right),c_e(\theta_e),L_w(\theta_e),l_e\left(\theta_e\right)\}$ to maximize the social welfare (7), subject to the incentive conditions (31) and (34), where the inverse demand function satisfies (18); the market clear conditions (15) to (17); and condition (19). Condition (19) serves as a policy constraint in the planner's problem. Essentially, it requires the marginal revenue of labor inputs:

$$\varpi\left(\theta_e\right)\equiv\frac{\partial\left[P_{ij}\left(\theta_e\right)Q_{ij}\left(\theta_e\right)\right]}{\partial L_{w,ij}\left(\theta_e\right)}\tag{35}$$

to be equal across firms. Thus, the uniform-sales-tax policy constraint can be rewritten as $\frac{d\ln\varpi(\theta_e)}{d\theta_e}=0$. Finally, it is often more convenient to take $V_o\left(\theta_o\right)$ as the planner's decision variable instead of $c_o(\theta_o)$. After solving the planner's problem, the price system and tax system can be constructed using the FOCs.

Next, we introduce some concepts that are useful to derive our main results: tax wedges, the hazard rate, the skill gap, and social welfare weights.

Tax Wedges Marginal distortions introduced by taxes in agents' decisions can be described using tax wedges. Entrepreneurs face three choices (consumption, effort, and hiring workers), while workers face two (consumption and working hours). This results in three tax wedges:

$$\tau_s\left(\theta_e\right)=1-\frac{W}{\frac{\partial\left[P_{ij}(\theta_e)Q_{ij}(\theta_e)\right]}{\partial L_{w,ij}(\theta_e)}},\tau_w\left(\theta_w\right)=1-\frac{\phi'_w\left(l_w\left(\theta_w\right)\right)}{Wx_w\left(\theta_w\right)},\tau_e\left(\theta_e\right)=1-\frac{\phi'_e\left(l_e\left(\theta_e\right)\right)}{\frac{\partial\left[P_{ij}(\theta_e)Q_{ij}(\theta_e)\right]}{\partial l_{e,ij}(\theta_e)}}.$$

$\tau_s\left(\theta_e\right)$ represents the tax wedge between the marginal cost and marginal income of labor inputs $L_w\left(\theta_e\right)$. Similarly, $\tau_w(\theta_w)$ is the tax wedge between the marginal disutility and income of the labor supply l_w, while $\tau_e(\theta_e)$ represents the tax wedge between the marginal disutility and income of the entrepreneur's labor supply l_e. These wedges are determined by the taxes.[27]

In our benchmark model, due to the policy constraint on the sales tax, $\tau_s\left(\theta_e\right)$ is uniform. Without loss of generality, we set τ_s, or t_s, to zero.[28] Consequently, $\tau_w\left(\cdot\right)$ and $\tau_e\left(\cdot\right)$ effectively capture the tax rates on $l_w\left(\cdot\right)$ and $l_e\left(\cdot\right)$, respectively.

[27] From the FOCs (18) - (21), we obtain $\tau_s\left(\cdot\right)=t_s$, $\tau_w\left(\theta_w\right)=T'_w\left(y_w\left(\theta_w\right)\right)$, and $\tau_e\left(\theta_e\right)=1-\left(1-t_s\right)\left[1-T'_e\left(y_e\left(\theta_e\right)\right)\right]$.

[28] As is well-established in the optimal tax literature, multiple tax systems can typically implement the second-best allocation (see, e.g., Chari and Kehoe (1999); Golosov et al. (2003)). In our model, as long as $\tau_s\left(\theta_e\right)$ is restricted to being uniform and income taxes are unconstrained, there is no need to enforce sales taxes alongside direct taxes. Furthermore, as long as $\tau_s\left(\cdot\right)$ is uniform, a sales tax can be replaced by a consumption tax. This equivalence suggests that a uniform sales tax should be treated not as a tax borne by the firm but rather as a uniform tax on labor factors and entrepreneurial efforts.

Income Elasticity. Define the non-linear elasticity of labor income as $\varepsilon^{y_w}_{1-\tau_w}(\theta_e) = \frac{1}{1+1/\varepsilon_w}\frac{d\ln y_w(\theta)}{d\ln x_w(\theta_w)}$, a familiar from the optimal income tax literature (see, e.g., Sachs et al. (2020)). For profit elasticity, consider a marginal increase $d\tau$) in the marginal tax rate faced by an entrepreneur of type θ_e. The non-linear profit elasticity is defined as $\varepsilon^{y_e}_{1-\tau_e}(\theta_e) \equiv -\frac{dy_e(\theta_e)}{y_e(\theta_e)} / \frac{d\tau}{1-T'_e(y_e(\theta_e))}$. The entrepreneur takes actions of others as given and responds rationally to the tax reform. We then have (see Appendix A.2.2 for details):

$$\varepsilon^{y_e}_{1-\tau_e}(\theta_e) = \frac{1}{\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu(\theta_e) - \xi\right] - 1 + \frac{y_e(\theta_e)T''_e(y_e(\theta_e))}{1-T'_e(y_e(\theta_e))}}.$$

Using the superscript o to denote variables in the Laissez-faire economy, let $y^o_e : \Theta_e \mapsto \mathbb{R}_+$ denote the profit in the Laissez-faire economy. The profit elasticity is then given by:

$$\varepsilon^{y^o_e}_{1-\tau_e}(\theta_e) = \frac{1}{\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu(\theta_e) - \xi\right] - 1}, \tag{36}$$

which is observable, as both $\varepsilon^{y_e}_{1-\tau_e}(\theta_e)$ and the progressivity of the profit tax $\frac{y_e(\theta_e)T''_e(y_e(\theta_e))}{1-T'_e(y_e(\theta_e))}$ are observable.

Hazard Ratio. Let $F^o_{y_e} : \mathbb{R}_+ \mapsto \mathbb{R}_+$ denote the CDF of profit in the Laissez-faire economy and let $f^o_{y_e}(y^o_e) = F^{o\prime}_{y_e}(y^o_e)$ be the PDF. The hazard ratio of profit in the Laissez-faire economy is:

$$H(\theta_e) \equiv \frac{1 - F^o_{y_e}(y^o_e(\theta_e))}{y^o_e(\theta_e)\, f^o_{y_e}(y^o_e(\theta_e))} = \frac{d\ln y^o_e(\theta_e)}{d\theta_e}\frac{1-F_e(\theta_e)}{f_e(\theta_e)} \tag{37}$$

The hazard ratio $H(\theta_e)$ is observable. For linear profit taxes, $H(\theta_e) = \frac{1-F_{y_e}(y_e(\theta_e))}{y_e(\theta_e)f_{y_e}(y_e(\theta_e))}$. In the general case, $H(\theta_e)$ can be derived using the real hazard ratio $\frac{1-F_{y_e}(y_e(\theta_e))}{y_e(\theta_e)f_{y_e}(y_e(\theta_e))}$ and the progressivity of the profit tax (equations (OA23)–(OA25)).

Skill Gap. A central concept in the Mirrleesian optimal taxation framework is the skill gap, $\frac{x'_w(\theta_w)}{x_w(\theta_w)}$, which determines how a worker's wage rate varies with their type and drives the income gap. The analogous concept for entrepreneurs is $\frac{\gamma'_e(\theta_e)}{\gamma_e(\theta_e)}$:

$$\frac{\gamma'_e(\theta_e)}{\gamma_e(\theta_e)} \equiv \left[\kappa\frac{d\ln X_e(\theta_e)}{d\theta_e} - \kappa\frac{d\ln\mu(\theta_e)}{d\theta_e} + \frac{d\ln\left[\mu(\theta_e)-\xi\right]}{d\theta_e}\right]\left[\mu(\theta_e) - \xi - \frac{\varepsilon_e}{1+\varepsilon_e}\right], \tag{38}$$

where $\kappa = \frac{\frac{\sigma}{\sigma-1}\frac{1+\varepsilon_e}{\varepsilon_e}}{\frac{1+\varepsilon_e}{\varepsilon_e}\left(\frac{\sigma}{\sigma-1}-\xi\right)-1}$ and $\mu(\theta_e) - \xi - \frac{\varepsilon_e}{1+\varepsilon_e} = \frac{1}{\varepsilon^{y^o_e}_{1-\tau_e}(\theta_e)}\frac{\varepsilon_e}{1+\varepsilon_e}$ under the nested-CES technology.

In competitive markets, where $\mu(\theta_e) = 1$, $\frac{\gamma'_e(\theta_e)}{\gamma_e(\theta_e)}$ equals $\frac{d\ln X_e(\theta_e)}{d\theta_e}$ multiplied by a constant, which indicates $\gamma_e(\theta_e)$ represents the effective skill of the entrepreneur. However, in markets with imperfect competition, the skill gap also depends on the markup $\mu(\theta_e)$.[29] This indicates that rising market power influences optimal taxation by reducing the profit elasticity on the RHS of (38) and increasing the skill gap.

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

$$g_o(\theta_o) \equiv \frac{G'(V_o(\theta_o))\widetilde{f}_o(\theta_o)}{\lambda f_o(\theta_o)} \quad \text{and} \quad \bar{g}_o(\theta_o) \equiv \frac{\int_{\theta_o}^{\bar{\theta}_o} g(x)\widetilde{f}_o(x)\,dx}{1-F_o(\theta_o)}, \tag{39}$$

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. These weights depend on the Pareto weights, the social welfare function, and the marginal utility, reflecting the planner's preference for equity.

[29] See Online Appendix OA.2.2 for details.

3 Main Results

We now analyze the properties of the economy under optimal taxation as determined by the planner in solving the benchmark model. We begin by presenting the most general result on the tax formula in Section 3.1. In Section 3.2, we explore the relationship between market structure and optimal taxation, which is the central theme of our model and findings. We derive a series of results that apply to special cases: (i) a competitive economy, (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 progressively reveals the distinct components of the optimal tax wedges. We summarize our findings in Section 3.3.

3.1 General Tax Formulas

The framework developed in this paper applies to general, non-parametric technological specifications. In Online Appendix OC.3, we present statistic-based optimal tax formulas that are independent of specific technological assumptions (see Theorem OC1). For analytical tractability, we now provide parameter-based optimal tax formulas in the nested-CES economy. These formulas illustrate how the optimal tax deviates from the canonical Mirrleesian tax in the presence of 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 nested-CES economy:*

$$\frac{1}{1-\tau_w(\theta_w)} = \frac{1}{\mu}\left[1+\left[1-\bar{g}_w(\theta_w)\right]\frac{1-F_w(\theta_w)}{f_w(\theta_w)}\frac{x_w'(\theta_w)}{x_w(\theta_w)}\frac{1+\varepsilon_w}{\varepsilon_w}\right], \tag{40}$$

$$\frac{1}{1-\tau_e(\theta_e)} = \frac{\frac{1}{\mu(\theta_e)}\left[1+\left[1-\bar{g}_e(\theta_e)\right]\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{\gamma_e'(\theta_e)}{\gamma_e(\theta_e)}\frac{1+\varepsilon_e}{\varepsilon_e}\right]+\frac{\frac{\sigma}{\sigma-1}}{\frac{\sigma}{\sigma-1}-\xi}IRE(\theta_e)}{1-\frac{\xi}{\frac{\sigma}{\sigma-1}-\xi}RE(\theta_e)}. \tag{41}$$

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

$$RE(\theta_e) \equiv \frac{\mu}{\mu(\theta_e)} - 1, \tag{42}$$

$$IRE(\theta_e) \equiv \varepsilon_{Q_{-ij}}^{P,cross}(\theta_e)\left\{\left[1-g_e(\theta_e)\right]-\left[1-\bar{g}_e(\theta_e)\right]H(\theta_e)\right\}. \tag{43}$$

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

$$\omega(\theta_e) = \frac{\left[\left[1-\tau_e(\theta_e)\right]\left(\frac{X_e(\theta_e)}{\mu(\theta_e)}\right)^{\frac{1+\varepsilon_e}{\varepsilon_e}\frac{\sigma}{\sigma-1}}\right]^{\frac{1}{\frac{1+\varepsilon_e}{\varepsilon_e}\left(\frac{\sigma}{\sigma-1}-\xi\right)-1}} f_e(\theta_e)}{\int_{\theta_e}\left[\left[1-\tau_e(\theta_e)\right]\left(\frac{X_e(\theta_e)}{\mu(\theta_e)}\right)^{\frac{1+\varepsilon_e}{\varepsilon_e}\frac{\sigma}{\sigma-1}}\right]^{\frac{1}{\frac{1+\varepsilon_e}{\varepsilon_e}\left(\frac{\sigma}{\sigma-1}-\xi\right)-1}} f_e(\theta_e)\,d\theta_e}. \tag{44}$$

Proof. See Appendix B.3. ■

Two things are worth noting here. First, equations (40), (41), along with the weights for firm-level markups (equation 44), describe 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 to capture the influence of the market structure I

and the elasticity of substitution $\eta(\theta_e)$. In other words, I and $\eta(\theta_e)$ only affect the optimal tax formulas and equilibrium conditions through $\mu(\theta_e)$. The general tax formula can be decomposed into four components, each of which drives increases in the tax rates:

(i) Mirrleesian Component. In a competitive goods market, the tax rules reduces to the traditional Mirrleesian tax formulas. That is, $\tau_o(\theta_o) = \tau_o^M(\theta_o)$, and

$$\begin{aligned} \frac{\tau_w^M(\theta_w)}{1-\tau_w^M(\theta_w)} &= [1-\bar{g}_w(\theta_w)] \frac{1-F_w(\theta_w)}{f_w(\theta_w)} \frac{x_w'(\theta_w)}{x_w(\theta_w)} \frac{1+\varepsilon_w}{\varepsilon_w}, \\ \frac{\tau_e^M(\theta_e)}{1-\tau_e^M(\theta_e)} &= [1-\bar{g}_e(\theta_e)] \frac{1-F_e(\theta_e)}{f_e(\theta_e)} \frac{\gamma_e'(\theta_e)}{\gamma_e(\theta_e)} \frac{1+\varepsilon_e}{\varepsilon_e}. \end{aligned} \tag{45}$$

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

A key driver is the skill gap $\frac{\gamma_e'(\theta_e)}{\gamma_e(\theta_e)}$. Market power influences the optimal profit tax through multiple channels, with the most significant being its impact on this skill gap. Our quantitative analysis below demonstrates that observed increases in markups raise the top profit tax rate by increasing the skill gap.

(ii) Pigouvian Component. Market power directly affects optimal taxation through the reciprocal of the markup, $\frac{1}{\mu}$ and $\frac{1}{\mu(\theta_e)}$. We refer to these effects as the *Pigouvian component* for labor and profit taxes, respectively. An important insight is that the optimal tax rate depends on the employment-weighted average markup, which assigns a higher weight to larger firms with higher markups. This weighted average typically exceeds the unweighted markup average, magnifying the tax impact on larger firms.[30]

(iii) Reallocation Effect. The Reallocation Effect term, $RE(\theta_e)$, reflects the efficiency of labor allocation. This effect raises the optimal tax rate when $\mu(\theta_e) < \mu$ and lowers it otherwise. The rationale is that labor should be reallocated to high-markup firms where labor demand is inefficiently low.

(iv) Indirect Redistribution Effect. Indirect Redistribution Effect, $IRE(\theta_e)$, captures profit tax-induced redistribution. It consists of two parts: a local redistribution effect $\varepsilon_{Q_{-ij}}^{P,cross}(\theta_e)[1-g_e(\theta_e)]$, and a cumulative redistribution effect $\varepsilon_{Q_{-ij}}^{P,cross}(\theta_e)[1-\bar{g}_e(\theta_e)]\, H(\theta_e)$.

Depending on the distribution of skills and welfare weights, the indirect redistribution effect may either increase or decrease the optimal tax rate. For example, a reduction in $\tau_e(\theta_e)$ increases output in submarket θ_e, lowering prices. If $g_e(\theta_e) < 1$, this price decrease promotes redistribution. However, maintaining entrepreneurial effort requires further reductions in $\tau_e(\theta_e)$, triggering cumulative effects on higher-skilled entrepreneurs' tax liabilities. Whether $IRE(\theta_e)$ ultimately increases or decreases $\tau_e(\theta_e)$ depends on the net impact of these opposing forces.

Comparison with the Existing Literature. Unlike prior studies (e.g., Kaplow (2019), Boar and Midrigan (2019)), our framework shows that rising market power can increase optimal profit tax rates, even without changes in welfare weights. Additionally, it suggests a decrease in labor income tax rates as market power grows. These differences arise from two factors: First, we consider both workers and entrepreneurs, with

[30] Additionally, changes in employment-weighted markups tend to be larger than those in unweighted averages, especially if markup changes are driven by large firms. This amplifies the reallocation effect.

only the latter exerting market power. Since both the labor supply of workers and entrepreneurs are inefficiently low, the Pigouvian component requires a decrease in both the labor and profit tax rates. Second, market power influences profit tax rates through several channels (RE, IRE, and Mirrleesian Component), while labor income tax rates are primarily affected by the Pigouvian Component.

Next, we explore optimal taxation under various market structures to further clarify these distinctions.

3.2 Market Structure and Optimal Taxation

We explore how changes in market structure affect optimal tax rates and equilibrium outcomes. Table OE3 summarizes the influence of market structure on each component of the optimal tax formula.

Table 1: Impact of Market Structure on Optimal Taxation

	Mirrleesian component	Pigouvian component	IRE	RE
Competitive Economy	✓		✓	
Monopolistic Competition	✓	✓		✓⋆
Oligopolistic Competition	✓	✓	✓	✓⋆

⋆heterogenous markups

The canonical Mirrleesian tax literature studies optimal taxation in a competitive economy with heterogenous labor (see, e.g., Mirrlees (1971), Piketty (1997), Diamond (1998), Saez (2001)). In such settings, the optimal tax formula includes only the Mirrleesian component when factors are perfectly substitutable, whereas the Indirect Redistribution Effect (IRE) arises when factors are imperfectly substitutable (see, e.g., Stiglitz (1982)).

Under monopolistic competition (see, e.g., Gürer (2021) and Boar and Midrigan (2019)), the Pigouvian component emerges. However, the IRE disappears because entrepreneurs' pricing behavior eliminates the first-order effect of taxes on prices. As entrepreneurs choose prices to maximize their utility, the envelope theorem implies that changes in the tax system have no first-order effect on utility.

This changes under oligopolistic competition. Entrepreneurs no longer hold monopoly power, and their pricing behavior cannot fully offset the first-order effects of taxes on prices. The Reallocation Effect (RE) emerges when firm-level markups are heterogeneous, as the marginal productivity of labor differs across firms. Reallocating labor across firms can enhance total factor productivity, making taxation a tool for improving resource allocation.

It is noteworthy that factors beyond market structure can also lead to Pigouvian and reallocation effects. For instance, Rothschild and Scheuer (2013) identify a reallocation effect[31] and an indirect redistribution effect in their tax formula, while Rothschild and Scheuer (2014) and Rothschild and Scheuer (2016) introduce the Pigouvian effect as well. In their competitive economy, the reallocation and Pigouvian effects arise from the externalities of agents' occupational choices.

Our paper breaks new ground by identifying strategic pricing and markup inequality as novel sources for the IRE and RE effects in taxation. It demonstrates that whether the optimal profit tax rate rises with market power critically depends on the underlying market structure. We now explore special cases that our model encompasses.

[31]See the Sectoral Shift Effect in their paper.

(i) Competitive Economy. To enhance transparency, we now analyze optimal taxation under different market structures, considering each scenario individually. When $I = 1$ and $\sigma \to \infty$, the economy reverts to the canonical Mirrlees (1971) taxation framework. The general tax formula simplifies to:

$$\tau_w(\theta_w) = \tau_w^M(\theta_w) \quad \text{and} \quad \tau_e(\theta_e) = \tau_e^M(\theta_e). \tag{46}$$

Here, the optimal tax rate reflects a tradeoff between redistributive benefits and labor distortions.

(ii) Monopolistic Competition. Consider an economy where each market has a single monopolistic producer, i.e., $I = 1$. In this setting, markups are uniform and equal to $\mu = \frac{\sigma}{\sigma-1}$. The solution to the planner's problem under this structure yields the following optimal tax rules:

Proposition 2 *When $I = 1$ the optimal labor income tax formula (40) holds. The optimal profit tax can be simplified as:* [32]

$$\frac{1}{1-\tau_e(\theta_e)} = \frac{1 + [1-\bar{g}_e(\theta_e)]\frac{1+\varepsilon_e}{\varepsilon_e}\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\left[\mu\frac{\chi'(\theta_e)}{\chi(\theta_e)} + \frac{x_e'(\theta_e)}{x_e(\theta_e)}\right]}{\mu}. \tag{47}$$

Proof. Omitted.[33] ■

Three observations are worth noting. First, the Pigouvian component now enters the optimal tax formula. When workers and entrepreneurs are homogeneous, the optimal tax wedges simplify:

$$\frac{1}{1-\tau_w} = \frac{1}{1-\tau_e} = \frac{1}{\mu}. \tag{48}$$

These tax wedges serve as Pigouvian taxes, counteracting the externalities of market power by subsidizing entrepreneurs and workers to boost output. Since firm-level markups are uniform, the Pigouvian components in labor and profit taxes are identical. Unlike the markup in the labor income tax, the markup in the profit tax also enters the Mirrleesian component. The optimal profit tax is determined by the composite ability $\frac{X_e'(\theta_e)}{X_e(\theta_e)} = \mu\frac{\chi'(\theta_e)}{\chi(\theta_e)} + \frac{x_e'(\theta_e)}{x_e(\theta_e)}$, which depends on μ. This difference between the labor income and profit taxes leads to the second noteworthy insight.

Second, with rising markups, the after-tax retention rate of labor income generally increases relative to profit income. Specifically,

$$\frac{1-\tau_w(\theta_w)}{1-\tau_e(\theta_e)} = \frac{1 + [1-\bar{g}_e(\theta_e)]\frac{1+\varepsilon_e}{\varepsilon_e}\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\left[\mu\frac{\chi'(\theta_e)}{\chi(\theta_e)} + \frac{x_e'(\theta_e)}{x_e(\theta_e)}\right]}{1 + [1-\bar{g}_w(\theta_w)]\frac{1+\varepsilon_w}{\varepsilon_w}\frac{1-F_w(\theta_w)}{f_w(\theta_w)}\frac{x_w'(\theta_w)}{x_w(\theta_w)}}. \tag{49}$$

This ratio increases in μ if $\frac{\chi'(\theta_e)}{\chi(\theta_e)} > 0$, which typically holds.

Third, in a monopolistically competitive economy with exogenous social welfare weights, the optimal profit tax rate tends to decrease with increasing markups if entrepreneurial skill $x_e(\theta_e)$ rises with ability. This is not necessarily the case under oligopolistic competition.

(iii) Oligopolistic Competition with Uniform Markups. Next, we examine oligopolistic competition $I > 1$ with uniform markups ($\eta(\theta_e)$ constant). Here, inter-firm strategic interactions exist, but markup inequality is absent.

[32]For $I = 1$, we have $\mu\frac{\chi'(\theta_e)}{\chi(\theta_e)} + \frac{x_e'(\theta_e)}{x_e(\theta_e)} = \mu\frac{X'(\theta_e)}{X(\theta_e)}$, because $\mu = \frac{\sigma}{\sigma-1}$. Therefore, we have another form of equation (47): $\frac{1}{1-\tau_e(\theta_e)} = \frac{1+[1-\bar{g}_e(\theta_e)]\frac{1+\varepsilon_e}{\varepsilon_e}\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\mu\frac{X'(\theta_e)}{X(\theta_e)}}{\mu}$, $\forall\theta_e \in \Theta_e$. This finding again suggests that the optimal tax rate is irrelevant to the specific composition of $X(\theta_e)$.

[33]For $I = 1$, we have $\mu(\theta_e) = \mu$ and $\varepsilon_{Q_{-ij}}^{P,cross}(\theta_e) = 0$. Under these conditions, equation (41)) reduces to (47). When workers and entrepreneurs are homogeneous, $\bar{g}_o = g_o = 1$, and the optimal tax formulas (40) and (41) simplify to (48).

Proposition 3 *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:

$$\frac{1}{1-\tau_e(\theta_e)} = \frac{1 + [1 - \bar{g}_e(\theta_e)] \frac{H(\theta_e)}{\varepsilon^{y_e^o}_{1-\tau_e}}}{\mu} + \frac{\frac{\sigma}{\sigma-1}}{\frac{\sigma}{\sigma-1} - \xi} IRE(\theta_e), \tag{50}$$

where $\varepsilon^{y_e^o}_{1-\tau_e} = \frac{1}{\frac{1+\varepsilon_e}{\varepsilon_e}(\mu-\xi)-1}$.[34]

(ii) Given $H(\theta_e)$, for any $\theta_e \in \Theta_e$, $\frac{1-\tau_w(\theta_w)}{1-\tau_e(\theta_e)}$ increases in μ iff $g_e(\theta_e) < 1 + [1 - \bar{g}_e(\theta_e)] \frac{H(\theta_e)}{\varepsilon^{y_e^o}_{1-\tau_e}}$; $\tau_e(\theta_e)$ increases in μ iff $g_e(\theta_e) < \frac{\xi(\sigma-1)}{\sigma}\left[1 + [1 - \bar{g}_e(\theta_e)] \frac{H(\theta_e)}{\varepsilon^{y_e^o}_{1-\tau_e}}\right]$.

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

Proof. See Online Appendix OC.7. ■

Part (i) of Proposition 3 provides an explicit formula for the optimal profit tax when social welfare weights are exogenous. Compared to the tax formula under monopolistic competition (47), an additional term appears: $\frac{\frac{\sigma}{\sigma-1}}{\frac{\sigma}{\sigma-1}-\xi} IRE(\theta_e)$.

As illustrated earlier, $IRE(\theta_e)$ represents the indirect redistribution effect of changing $Q_{ij}(\theta_e)$. The term $\frac{\frac{\sigma}{\sigma-1}}{\frac{\sigma}{\sigma-1}-\xi} > 1$ measures the percentage change in $Q_{ij}(\theta_e)$ resulting from a one-percent increase in $l_e(\theta_e)$. Specifically, a one-percent increase in $l_e(\theta_e)$ directly induces a one-percent increase in $Q_{ij}(\theta_e)$. Additionally, it indirectly crowds in $-\frac{\varepsilon^{\varpi}_{l_e}(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)}$ (or $-\xi \frac{\varepsilon^{\varpi}_{l_e}(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)}$) percent of $L_w(\theta_e)$ ($Q_{ij}(\theta_e)$), lowering the marginal productivity of $L_w(\theta_e)$ to the market level.

Here, $\varepsilon^{\varpi}_{l_e}(\theta) \equiv \frac{\partial \ln \varpi(\theta_e)}{\partial \ln l_e(\theta_e)}$ and $\varepsilon^{\varpi}_{L_w}(\theta_e) \equiv \frac{\partial \ln \varpi(\theta_e)}{\partial \ln L_w(\theta_e)}$ denote the own elasticities of productivity $\varpi(\theta_e)$ with respect to $l_e(\theta_e)$ and $L_w(\theta_e)$, respectively (see Appendix A.2.2).[35] In sum, a one-percent increase in $l_e(\theta_e)$ triggers a $1 - \xi \frac{\varepsilon^{\varpi}_{l_e}(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)} = \frac{\frac{\sigma}{\sigma-1}}{\frac{\sigma}{\sigma-1}-\xi} > 1$ percent increase in $Q_{ij}(\theta_e)$. Consequently, the term $\frac{\frac{\sigma}{\sigma-1}}{\frac{\sigma}{\sigma-1}-\xi} IRE(\theta_e)$ captures the total indirect redistribution effect caused by the increase in $l_e(\theta_e)$.

The literature has highlighted that the indirect redistribution effect (IRE) generally leads to a lower top tax rate (see, e.g., Stiglitz (1982) and Rothschild and Scheuer (2013)). To understand this, consider the scenario where $g_e(\theta_e)$ and $\bar{g}_e(\theta_e)$ approach zero as $\theta_e \to \overline{\theta}_e$. In our framework of oligopolistic competition with $\eta(\theta) = \eta$, the cross-inverse demand elasticity, $\varepsilon^{P,cross}_{Q_{-ij}}(\theta_e) \leq 0$, remains constant. As $\theta_e \to \overline{\theta}_e$, the indirect redistribution effect $IRE(\theta_e)$ converges to $\varepsilon^{P,cross}_{Q_{-ij}}(\theta_e)[1 - H(\theta_e)]$. This term is negative if $H(\theta_e) < 1$. Empirical evidence from the United States suggests that $H(\theta_e) < 1$ when θ_e is sufficiently large. Thus, IRE typically necessitates a lower profit tax rate for high-skilled entrepreneurs in the U.S.[36]

Parts (ii) and (iii) of Proposition 3 further analyze when the after-tax income ratio of labor to profit income, $\frac{1-\tau_w(\theta_w)}{1-\tau_e(\theta_e)}$, and the profit tax rate $\tau_e(\theta_e)$ increase with the markup μ. The findings indicate that given

[34] $\frac{H(\theta_e)}{\varepsilon^{y_e^o}_{1-\tau_e}} = \frac{1-F_e(\theta_e)}{f_e(\theta_e)} \frac{\gamma_e'(\theta_e)}{\gamma_e(\theta_e)} \frac{1+\varepsilon_e}{\varepsilon_e}$. See equation (OA25) for the relationship between $H(\theta_e)$ and $\frac{\gamma_e'(\theta_e)}{\gamma_e(\theta_e)}$.

[35] In the nested-CES economy, $\varepsilon^{\varpi}_{L_w}(\theta_e) = \xi \frac{\sigma-1}{\sigma} - 1$ and $\varepsilon^{\varpi}_{l_e}(\theta_e) = \frac{\sigma-1}{\sigma}$.

[36] In 2007, the hazard ratios for top labor, capital, and total incomes in the U.S. were approximately 0.62, 0.76, and 0.71, respectively (see, e.g., Saez and Stantcheva (2018)).

$H(\theta_e)$, both $\frac{1-\tau_w(\theta_w)}{1-\tau_e(\theta_e)}$ and $\tau_e(\theta_e)$ increase with μ when $g_e(\theta_e)$ is sufficiently small. Since $g_e(\theta_e)$ approaches zero as θ_e becomes large, there exists a range of θ_e for which these tax rates increase with μ.

These findings suggest that without accounting for market power inequality and changes in the hazard ratio of profits, the optimal profit tax rates for top firms will generally rise as markups increase. In the next section, we delve into the impact of the reallocation effect.

(iv) Oligopolistic Competition with Heterogeneous Markups. Finally, we consider the most general case, incorporating both oligopolistic competition and heterogeneous markups. from Theorem 1. For the workers, the tax formula (40) remains unchanged compared to the case with uniform markups. As for the entrepreneurs, the planner now uses the tax to implement an efficiency-enhancing reallocation of factors, which is captured by the denominator on the right side of (41), i.e., $1 - RE\left(\theta_e\right) \frac{\xi}{\frac{\sigma}{\sigma-1}-\xi} = 1 - \frac{\mu-\mu(\theta_e)}{\mu(\theta_e)} \frac{\xi}{\frac{\sigma}{\sigma-1}-\xi}$.

Remember 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 equalize the marginal productivities of labor inputs between firms 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 $-\frac{\varepsilon^{\varpi}_{l_e}(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)}$ the percentage increase in $L_w(\theta_e)$.

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

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

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.

Our optimal tax formula suggests that, with an increase in the firm-level markup, the reallocation effect requires a lower tax rate for the firm.[37] The above finding 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 (2013)). Note that the rise in markups does not decrease the optimal profit tax rates of all firms through the RE. Only when the growth rate of a firm's markup is faster than the average markup does changes in the reallocation effect require a reduction in the firm's marginal profit tax.

3.3 Summary of Results

There are four components determining the optimal profit tax rate, in addition to the social welfare weights. The Mirrleesian component reflects the trade-off between direct redistribution and the revenue effect of profit tax. The Pigouvian component restores productive efficiency by offsetting the externality from the 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. Lastly, the indirect redistribution effect captures the tax's redistribution effect through prices.

These four components suggest that changing market power has an amalgam of different, often opposing forces on the optimal profit tax rate. As an illustration, consider the increase of $\mu\left(\theta_e\right)$ under nested-CES:

1. Immediately, the Pigouvian component $\frac{1}{\mu(\theta_e)}$ decreases to reduce the tax rate.
2. The reallocation effect $RE\left(\theta_e\right) = \frac{\mu}{\mu(\theta_e)} - 1$ decreases if the firm-level markup $\mu\left(\theta_e\right)$ increases relative to the average markup.

[37] Since there is no use in setting a marginal tax rate larger than one, the right side of (41) is positive. Assuming $\tau_e\left(\theta_e\right) < 1$, the numerator of the right side of (41) is positive if the denominator $1 - \frac{\xi}{\frac{\sigma}{\sigma-1}-\xi} RE\left(\theta_e\right) > 0$, which is true because $\mu < \frac{\sigma}{\sigma-1}$.

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. The main ingredient of the Mirrleesian component – the skill gap $\frac{\gamma_e'(\theta_e)}{\gamma_e(\theta_e)}$ – generally increases because $\varepsilon^{y^o_e}_{1-\tau_e}(\theta_e) = \frac{1}{\mu(\theta_e)-\xi-\frac{\varepsilon_e}{1+\varepsilon_e}}$ decreases (see equation (38)).

In conclusion, our theoretical analysis identifies different forces behind the effect of market power on the profit tax rate, especially for the top firms. Our findings suggest that the labor income tax rate should decrease while the top profit tax rate should increase with the rise of market power. To evaluate the net effect as well as the contribution of each force, below in Section 5 we perform a quantitative exercise based on US micro-data to provide a concrete policy prescription on how the the tax authorities should react to the change in market power since the 1980s. Before this, we discuss the robustness of our results.

4 Discussion and Robustness: Policy-relevant Specifications

We consider alternative specifications that are relevant to put our results in perspective concerning concrete tax policy prescriptions.[38] In Appendix C, we discuss further extensions and robustness.

(i) Second-Best: Non-linear Sales Taxes. As we have emphasized before, 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 the policy constraint on the optimal profit tax. Set $\tau^E_w(\theta_w)$, $\tau^E_e(\theta_e)$ and $\tau^E_s(\theta_e)$ as the marginal labor income tax rate, profit tax rate, and *non-linear* sales income tax rate, respectively: $\tau^E_w(\theta_w) = \tau_w(\theta_w)$, $\tau^E_s(\theta_e) = \tau_s(\theta_e)$ and $1-\tau^E_e(\theta_e) = \frac{1-\tau_e(\theta_e)}{1-\tau^E_s(\theta_e)}$. 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:*

$$\frac{1}{1-\tau^E_w(\theta_w)} = \frac{1+[1-\bar{g}_w(\theta_w)]\frac{1-F_w(\theta_w)}{f_w(\theta_w)}\frac{x_w'(\theta_w)}{x_w(\theta_w)}\frac{1+\varepsilon_w}{\varepsilon_w}}{\mu^*}, \tag{51}$$

$$\frac{1}{1-\tau^E_e(\theta_e)} = \frac{1+[1-\bar{g}_e(\theta_e)]H(\theta_e)\left[[\mu(\theta_e)-\xi]\frac{1+\varepsilon_e}{\varepsilon_e}-1\right]}{\mu^*} + \frac{[1-\bar{g}_e(\theta_e)]\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\left[\frac{d\ln[\mu(\theta_e)-\xi]}{d\theta_e} + \frac{\mu(\theta_e)\varepsilon^{P,cross}_{Q_{-ij}}(\theta_e)\left(1+\xi\frac{1+\varepsilon_e}{\varepsilon_e}\right)\left[\frac{d\ln\left[1-\tau^E_s(\theta_e)\right]}{d\theta_e}+\frac{d\ln\left[1-\tau^E_e(\theta_e)\right]}{d\theta_e}\right]}{1-\frac{\sigma-1}{\sigma}\left(\frac{\varepsilon_e}{1+\varepsilon_e}+\xi\right)}\right]}{\mu^*}, \tag{52}$$

$$\frac{\tau^E_s(\theta_e)}{1-\tau^E_s(\theta_e)} = \left[\frac{\mu^*}{\mu(\theta_e)}-1\right] + \left[1-\tau^E_e(\theta_e)\right]\varepsilon^{P,cross}_{Q_{-ij}}(\theta_e)\{[1-g_e(\theta_e)]-[1-\bar{g}_e(\theta_e)]H(\theta_e)\} \tag{53}$$

$$+\left[1-\tau^E_e(\theta_e)\right]\varepsilon^{P,cross}_{Q_{-ij}}(\theta_e)\frac{[1-\bar{g}_e(\theta_e)][1-F_e(\theta_e)]}{f_e(\theta_e)}\left[\begin{array}{c}\frac{d\ln\left[\mu(\theta_e)\varepsilon^{P,cross}_{Q_{-ij}}(\theta_e)\right]}{d\theta_e} - \frac{d\ln[\mu(\theta_e)-\xi]}{d\theta_e} \\ +\frac{\frac{d\ln\left[1-\tau^E_s(\theta_e)\right]}{d\theta_e}+\left(1-\xi\frac{\sigma-1}{\sigma}\right)\frac{d\ln\left[1-\tau^E_e(\theta_e)\right]}{d\theta_e}}{1-\frac{\sigma-1}{\sigma}\left(\frac{\varepsilon_e}{1+\varepsilon_e}+\xi\right)}\end{array}\right] \tag{54}$$

[38] To facilitate the analysis, in this section, we assume that the relevant monotonicity hypothesis of the incentive problem is 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:*

$$\frac{1}{1-\tau_e^E(\theta_e)} = \frac{1 + \left[1 - \bar{g}_e(\theta_e)\right] H(\theta_e) \left[\frac{1+\varepsilon_e}{\varepsilon_e}\left(\mu(\theta_e) - \xi\right) - 1\right]}{\mu^*}, \tag{55}$$

and

$$\frac{\tau_s^E(\theta_e)}{1-\tau_s^E(\theta_e)} = \underbrace{\left[\frac{\mu^*}{\mu(\theta_e)} - 1\right]}_{RE^*(\theta_e)} + \left[1 - \tau_e^E(\theta_e)\right] \underbrace{\varepsilon_{Q_{-ij}}^{P,cross}(\theta_e)\left\{\left[1 - g_e(\theta_e)\right] - \left[1 - \bar{g}_e(\theta_e)\right] H(\theta_e)\right\}}_{IRE^*(\theta_e)}. \tag{56}$$

Proof. Ommited. ■

Corollary 1 directly follows from Theorem 2. $RE^*(\theta_e)$ and $IRE^*(\theta_e)$ are analogous in Theorem 1. 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 components. Moreover, the Pigouvian component 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. The tax wedge on entrepreneurial effort $1 - \tau_e(\theta_e) = \left[1 - \tau_e^E(\theta_e)\right]\left[1 - \tau_s^E(\theta_e)\right]$ still incorporates all of the four elements.

Equation (55) 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, 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 small firms with low markups and negative for large firms with high markups. Second, the optimal profit tax depends on the set of enforceable policies. Third, findings in our benchmark model regarding the optimal profit tax wedge pertain to the total tax rate on entrepreneurial effort enforced by the profit and sales taxes instead of the nominal profit tax alone. That said, irrespective of whether the non-linear sales tax is enforced, tax design critically depends on the four elements highlighted in the benchmark model.

(ii) Conditioning Taxes on Markups. In our setup so far, the planner cannot 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 nested-CES technology, the optimal solution where taxes condition on markups, as well as profits, is equivalent to the solution with non-linear sales tax schedules (see the Online Appendix OC.5.1 for further details). 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:

$$\frac{P_{ij}}{W}\frac{\partial Q_{ij}}{\partial L_w} = \mu\left(\theta_e'\right) \quad \text{and} \quad P_{ij}Q_{ij} - WL_w = y_e\left(\theta_e'\right), \tag{57}$$

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.

We reformulate the entrepreneur's problem and incentive constraints in Online Appendix OC.5.1 (see equations (OC6) and (OC7)). We find that $L_w\left(\theta_e'|\theta_e\right)$ is independent of θ_e as it is in our benchmark model.[39] This finding suggests that the incentive condition remains the same. In particular, the first-best optimum is not achievable even if the markup is observable.[40]

(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 (2019).[41] 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:

$$V_e\left(\theta_e\right) \equiv \max_{\theta_e'} V_e\left(\theta_e'|\theta_e\right) \tag{58}$$

$$\text{where} \quad V_e\left(\theta_e'|\theta_e\right) = \max_{L_w, l_e} P\left(Q_{ij}\left(\theta_e'\right), \theta_e\right) Q_{ij}\left(\theta_e'\right) - WL_w - T_e\left(\theta_e'\right) - \phi_e\left(l_e\right) \tag{59}$$

$$\text{s.t. } Q_{ij}\left(\theta_e'\right) = Q_{ij}\left(x_e\left(\theta_e\right) l_e, L_w\right). \tag{60}$$

[39] 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 (57) to derive $L_w\left(\theta_e'|\theta_e\right) = \frac{y_e(\theta_e')}{W[\mu(\theta_e')/\xi - 1]}$, which is independent of θ_e.

[40] 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.

[41] Boar and Midrigan (2019) cannot consider profit taxes because entrepreneurs provide no effort, and profit taxes therefore have no effect on behavior.

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

$$V_e'(\theta_e) = \frac{\partial V_e(\theta_e'|\theta_e)}{\partial \theta_e}|_{\theta_e'=\theta_e} = \frac{\partial P\left(Q_{ij}(\theta_e), \theta_e\right)}{\partial \theta_e} Q_{ij}(\theta_e) + \phi_e'(l_e(\theta_e)) l_e(\theta_e) \frac{x_e'(\theta_e)}{x_e(\theta_e)}. \quad (61)$$

The above incentive condition equals the original one if and only if $P(\theta_e) Q_{ij}(\theta_e) = \phi_e'(l_e(\theta_e)) l_e(\theta_e) \mu(\theta_e)$. Since, by the FOCs of the entrepreneur, $\phi_e'(l_e(\theta_e)) l_e(\theta_e) = \frac{WL_w(\theta_e)}{\xi}$ and $P(\theta_e) Q_{ij}(\theta_e) = \frac{WL_w(\theta_e)\mu(\theta_e)}{\xi}$, equation $P(\theta_e) Q_{ij}(\theta_e) = \phi_e'(l_e(\theta_e)) l_e(\theta_e) \mu(\theta_e)$ naturally holds. 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. This finding suggests that our main result is independent of the policy instrument and that quantity regulation can generally be replaced by a profit tax.

5 Quantitative Analysis

To provide concrete guidance for policymakers and to approximate the theoretical results to reality, we perform a quantitative exercise using microdata from the United States economy, estimating the model to match key moments regarding markups and income distribution. In Section 5.1 we lay out the parameterization and underlying assumptions. In Section 5.2 we describe the data and calibrate the distributions of $x_w(\theta_w)$, $X_e(\theta_e)$, and $\mu(\theta_e)$ using data on income and markups. In Section 5.3, we derive our main quantitative results: solving for the optimal tax rates and tax revenue, quantifying the four components from the tax formula, analyzing the role of market structure on optimal taxes, performing counterfactual exercises and robustness checks, and discussing the policy implications. We conduct the analysis for two different years, 1980 and 2019, to evaluate how optimal taxation changes during the period in which market power has risen.

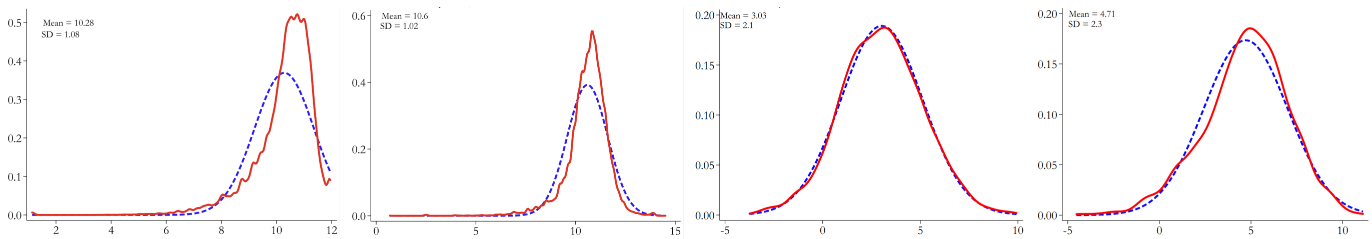
5.1 Parameterization

For our benchmark 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 nested-CES production technology specified in equations (2) to (5). To capture social preferences for redistribution, we consider a concave social welfare function $G(V) = \frac{V^{1-k}-1}{1-k}$, 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 5.3, we investigate the robustness of our main results to the choice of $\{k, \sigma, \xi\}$.

We assume θ_o equals the quantile of $y_o(\theta_o)$, meaning $f_o(\theta_o) = 1$ and $\Theta_o = [0, 1]$. Since the functions $x_w(\theta_w)$ and $X_e(\theta_e)$ govern abilities, there is no loss in assuming that the distributions of skills are uniform. As is well-known from the taxation literature, a uniform linear sales tax on output is equivalent to a uniform linear tax on all factors. Therefore, we can focus on income taxes and set sales tax to zero. Following Saez and Stantcheva (2018), we set linear and uniform income tax rates to match the U.S. average tax rate on total income, which means both the labor income and profit income tax rates are 25.6% in 1980 and 25% in 2019. Tax revenue is returned to the agents through lump sum transfer payments.

5.2 Data and Calibration

We calibrate our model to match the distributions of labor income, profits, and markups in the U.S. economy.



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

Figure 1: Labor Income and Profit Distributions: data and lognormal fit

Labor Income. We obtain the income distribution of salaried workers in 2019 and 1980 from the Current Population Survey (CPS) and use a 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. We also report the mean and standard deviation of the log wages.[42]

Profits. We use profit data from De Loecker et al. (2020) based on a sample of publicly traded firms and apply a lognormal fit to estimate the model. Figures 1c and 1d plot the distribution of log profits (red) and their lognormal fit (blue) for 1980 and 2019, reporting the mean and standard deviation of log profits.[43]

Markups. We obtain firm-level markup data using the method in De Loecker et al. (2020). A firm's markup is defined as the ratio of the output price to marginal cost and is estimated using the cost approach.[44] One insight from our theoretical analysis is that the firm-level markup is a sufficient statistic capturing the influence of market power on equilibrium allocation and optimal taxation. There is no need to calibrate $\eta(\theta_e)$ and I separately, as long as we have $\mu(\theta_e)$.

The cost-weighted average markup μ enters the optimal taxation formulas.[45] To match the model's markup distribution to the data, we rank firms by their firm-level markups μ_{ij}. Denote by $\vartheta_{ij} = \frac{\sum_{\mu_{i'j'} \leq \mu_{ij}} L_{w,i'j'}}{\sum_{i'j'} L_{w,i'j'}}$ the cumulative cost weight of firms with markups not exceeding μ_{ij}. We match the firm-level markup of firms at each ϑ_{ij} from the model to the data using a polynomial fit. Figure 2 reports the data and the model fit for the cost-weighted markup distribution in 1980 and 2019. The cost-weighted average markup increases from 1.26 in 1980 to 1.38 in 2019. The plot shows that the rise in the average markup is driven predominantly by an increase in markups in the top percentiles of the distribution.

Skill Gap. Finally, we calibrate $x_w(\cdot)$ and $X_e(\cdot)$ using the agents' first-order conditions. Our theoretical analysis shows 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 $\chi(\theta_e)$ separately. Instead, we calibrate $X_e(\theta_e)$, or alternatively, $\gamma_e(\theta_e)$.[46] Equation (38) provides the expression for the skill gap $\frac{\gamma_e'(\theta_e)}{\gamma e(\theta_e)}$,

[42] 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 greater than 0.

[43] Profits are in millions of dollars, adjusted to 2019 prices, and truncated at 0.

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

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

[46] 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.

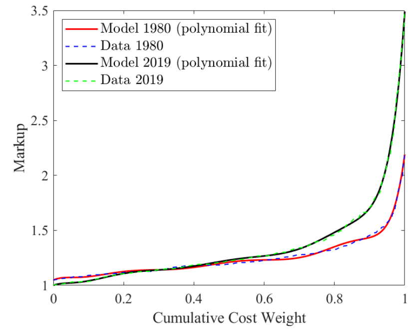
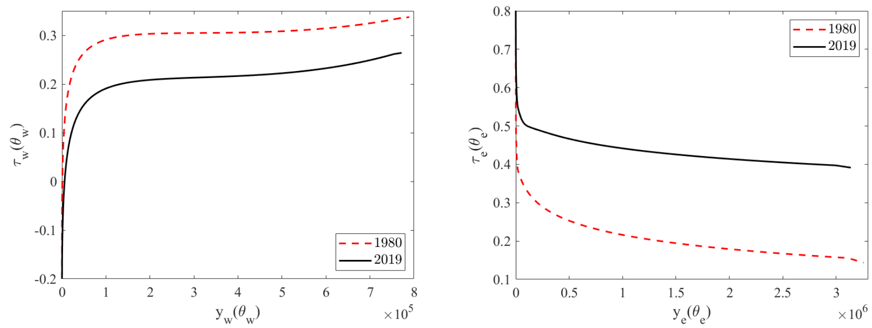


Figure 2: Markup Distribution: data and calibrated model



(a) Income-based Income Tax Wedges (b) Profit-based Profit Tax Wedges

Figure 3: Optimal Tax Wedges in 1980 and 2019

which we plot and discuss in Figure 5 below. Finally, we set $Ne = 1$ and derive Nw using the labor market clearing condition, which is equivalent to (because $\frac{y_e(\theta_e)}{W}\frac{\xi}{\mu(\theta_e)-\xi} = L_w(\theta_e)$):

$$N_w \int_{\theta_w} y_w(\theta_w) f_{\theta_w}(\theta_w)\, d\theta_w = N_e \int_{\theta_e} \frac{y_e(\theta_e)}{W} \frac{\xi}{\mu(\theta_e)-\xi} f_{\theta_e}(\theta_e)\, d\theta_e. \tag{62}$$

5.3 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 from 1980 to 2019, while the optimal effective tax rate on profit increases during the same period.

Figure 3 plots the optimal 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 the labor income tax.

The optimal average labor income tax rate decreased from 21.4% in 1980 to 11.5% in 2019, while the optimal average profit tax rate increases from 58.6% in 1980 to 61.3% in 2019.[47] Even though the top profit tax rate increases by 25 percentage points at the top, the average tax rate only increases by 2.7 percentage points. This is because only the marginal tax rate of firms at the very top increases significantly (see Figure OE2 in Appendix OE). 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.[48]

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

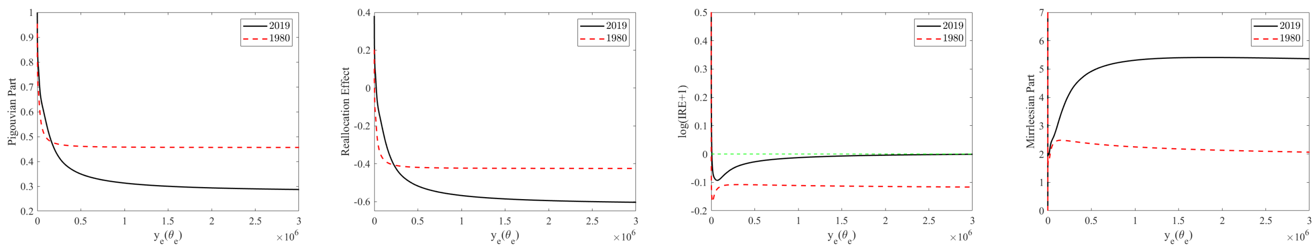
[47] The after-tax labor income of θ_o refers to $y_o - T_o(y_o) - t_o$, where $t_o = -T_o(y_o(\underline{\theta}_o))$ is a lump-sum transfer (tax if negative). Then $T_o(y_o) = \int_{y_o(\underline{\theta}_o)}^{y_o} T_o'(y)\,dy - t_o$. $\overline{Y}_o = \int_{\underline{y}_o}^{\overline{y}_o} [y - T_o(y) - t_o] f_{y_o}(y)\,dy$ is the average after-tax 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 $\overline{y}_o = y_o(\overline{\theta}_o)$. The average tax rate is $\overline{Y}_o/\overline{Y}_o$.

[48] The original after-tax share of labor income to the total income 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.

efficiency. However, it also increases market concentration and the average markup.[49]

In the robustness section of the quantitative exercise below, we explore a wide range of parameters to analyze how our results vary under different configurations. We find that our main conclusions remain robust, including the substantial gap between the average profit tax and the profit tax wedge schedule.

Decomposition of Optimal Taxation. The primary driver behind the decrease in the labor income tax rate is the rise in average markups, which lowers the Pigouvian component of the tax rule. The changes in the profit tax rate, however, are more complex, as they result from the interplay of multiple factors. In Figure 4, we decompose the profit tax rate into its constituent elements: the Pigouvian component, the reallocation effect, the indirect redistribution effect, and the Mirrleesian component.



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

Figure 4: Four Components of the Profit Tax Wedge

Figure 3 shows that the combined effect of the four components results in an overall increase in the profit tax rate. In Figure 4, we observe a decline in the top profit tax rate due to the Pigouvian and reallocation components, while the indirect redistribution effect and the Mirrleesian component drive an increase in the top profit tax rate. Overall, the positive impact of the latter components dominates. Notably, the Mirrleesian component is the primary contributor to the increase in the top profit tax rate.

It is important to highlight that the Mirrleesian component is driven by the skill gap, which is influenced by both the ability distribution and the markup (see equation (38)). The markup affects the Mirrleesian component via profit elasticity. Figure 5 illustrates that the change in the skill gap is almost entirely due to changes in profit elasticity.

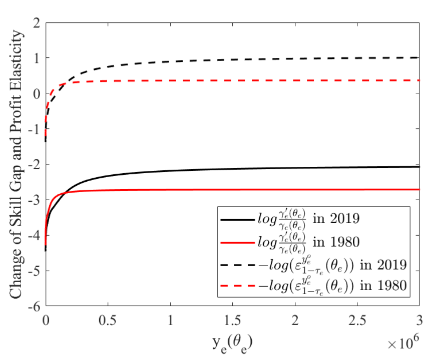


Figure 5: Skill Gap and Profit Elasticity

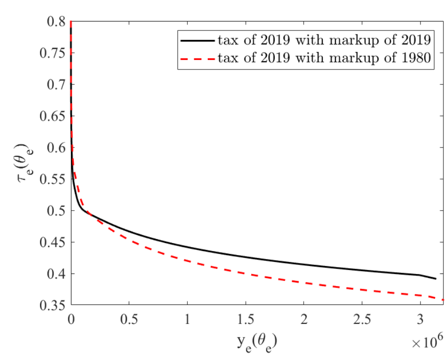


Figure 6: Counterfactual Profit Tax Wedges

The solid lines in Figure 5 represent the logarithms of the skill gap in 1980 and 2019, while the dashed lines show the logarithms of the inverse of profit elasticity for the same years. Recall that profit elasticity is

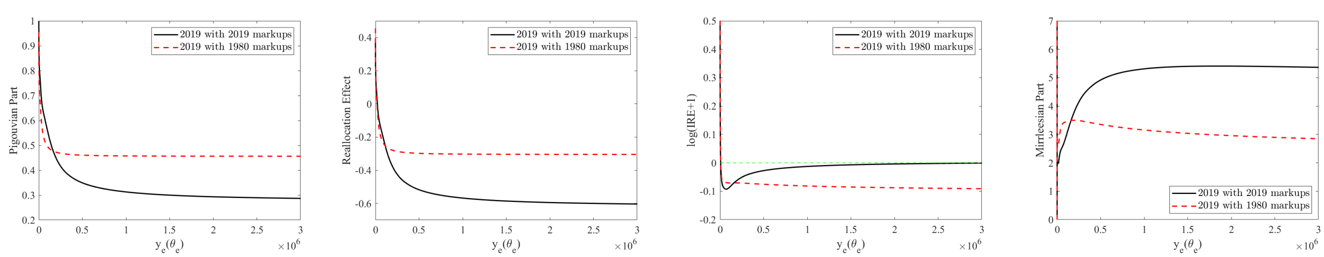
[49]The cost-weighted average markup in 2019 increases slightly from 1.375 in the initial economy to 1.382 in the optimum. For 1980, the increase is from 1.257 to 1.261.

a key component of the skill gap (see equations (38) and (36)). The rise in the skill gap from 1980 to 2019 is primarily driven by the decrease in profit elasticities, as indicated by the similar patterns of both curves. Since changes in profit elasticity are driven by changes in markups (as seen in equation (36)), this finding suggests that rising markups generally increase the Mirrleesian component by reducing profit elasticity.

It's important to note that markup levels affect all four components of optimal taxation.[50] Whether rising market power increases or decreases the optimal tax rate depends on the interplay of these four components. For the top profit tax rate, the Mirrleesian component has the largest influence. As long as the government prioritizes equality, recent increases in market power justify a higher top-profit tax rate.

Market Structure and Optimal Taxation. How do assumptions about market structure affect tax design? Our findings show that the labor tax rate is significantly lower than in models with a competitive economy. For instance, in a purely competitive economy, the top labor income tax rate is around 60% (see Saez (2001)), compared to about 25% in our model, which incorporates market power. This difference arises primarily from the Pigouvian component in the labor income tax formula (40) and the statistics-based formula (OC4). When calibrated to the labor income distribution, the Mirrleesian component remains unchanged. However, the profit tax rate is influenced by three additional components beyond the Mirrleesian component. Through counterfactual analysis, we demonstrate that rising market power increases the top profit tax rate.

Counterfactual Analysis. To explore the impact of changes in markups on optimal profit tax wedges, we perform a counterfactual analysis of $\mu(\cdot)$. Specifically, we examine how the optimal tax rate in 2019 would change if the markup distribution remained at its 1980 level. Figure 6 compares the profit tax wedges for 2019 under counterfactual 1980-level markups to the actual tax wedges for 2019. The results show lower taxes for high-profit incomes and higher taxes for low-profit incomes under the counterfactual scenario. This analysis isolates the effect of markups on the optimal tax rate, independent of changes in productivity. As illustrated in Figure 6, fixing markups at their lower 1980 level reduces the tax wedge for high-profit entrepreneurs. We then decompose the components of the optimal tax rates to demonstrate how the top profit tax rate changes in this counterfactual exercise, where only markups vary while the productivity distribution remains at its 2019 level.



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

Figure 7: Counterfactual: The Four Components 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 component 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 IRE increases the tax rates of low and middle types and decreases the tax

[50] The indirect redistribution effect also depends on market structure due to pricing power.

rates of high types. While it significantly decreases the top profit tax with markups of 1980, it has minor effects on the top tax rate in 2019 because of the cross-inverse demand elasticity becomes smaller for the top. Finally, we see that the increase in the Mirrleesian component due to the increase of the markup is primarily responsible for the increase in the top profit tax.

Robustness. We evaluate the robustness of our main results by considering different parameter values: $k \in \{0.7, 0.77, 1, 3\}, \xi \in \{0.4, 0.5, 0.6\}$, and $\sigma \in \{1.4, 1.3, 1.2\}$. These choices reflect common values used in the literature.[51] To ensure the concavity of the entrepreneur's incentive problem, we restrict ξ to values that prevent the span of control from becoming excessively large. Furthermore, σ cannot exceed 1.4 because the theoretical maximum value of firm-level markup decreases with σ. Since the highest observed markup in our dataset is 3.5, $\sigma > 1.4$ would not be feasible.

Our findings remain consistent across this broad parameter range: labor income tax rates decrease, and profit tax rates increase. The optimal profit tax rate is largely insensitive to k. For all parameter configurations, the optimal average profit tax rate is approximately 58% in 1980 and 61% in 2019. However, the optimal average labor income tax rate is sensitive to k. For $k = 0.7, 1$, and 3, the average labor income tax rates are 19.7%, 26.1%, and 44% in 1980; and 9.6%, 16.7%, and 37% in 2019. These results are visualized in Figure OE3 of Online Appendix OE.

We also test the robustness of our result by varying the values of ξ and σ. While changes in ξ and σ do not affect the optimal tax wedge $\tau_w(\theta)$, they do influence the tax schedule $T_w(y_w)$ through the associated changes in $y_w(\cdot)$. Additionally, ξ and σ impact $T_e(\cdot)$ both directly and indirectly. Figures OE4 and OE5 in Online Appendix OE confirm that our conclusions hold because the impact of ξ and σ on the optimal tax policies of 1980 and 2019 align in the same direction.

Policy Implications. Our quantitative results yield three key *policy implications*. First, given the changes in market structure observed in recent years, labor income tax rates should be reduced, while profit tax rates—particularly for top firms—should be increased. Second, for large, high-productivity firms, the profit tax rate should be designed to be appropriately regressive to enhance production efficiency via the reallocation effect. Third, while the optimal profit tax rate should be regressive for large firms, this regressivity decreases as market power rises.

In contrast to existing studies (see, e.g., Kaplow (2019) and Boar and Midrigan (2019)), our findings highlight that (i) even without accounting for changes in social welfare weights, rising market power necessitates an increase in the optimal profit tax rate, and (ii) the optimal tax system not only enhances allocative efficiency but also mitigates inequality.

6 Conclusion

The most effective way to address market power is to eliminate its root cause through competition policy. In its absence, we ask what role income taxation can play in addressing the inefficiency and inequality caused by 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 influence

[51] 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 an early version of their paper, Sachs et al. (2016) consider $k = 1$ and 3.

optimal production through the incentive constraint, endogenous labor supply, and the general equilibrium wage effect.

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

Our theory integrates the Mirrleesian approach to optimal income taxation with markets featuring oligopoly pricing, resulting in a tractable framework. We derive tax wedges for labor and entrepreneurial income taxes, which can be decomposed into four components: (i) a Mirrleesian component; (ii) a Pigouvian correction for the externality from market power; (iii) an indirect redistribution effect; and (iv) a reallocation effect toward more productive firms.

We conduct a detailed quantitative analysis, calibrating our model economy to match key moments of the U.S. economy in 1980 and 2019, a period marked by rising market power. Our estimates allow us to decompose the optimal tax rules into the four components identified 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, while optimal profit taxes are higher on average. Additionally, we find that the optimal profit tax is regressive to enhance allocative efficiency, though it is less regressive in 2019 than in 1980, reflecting the trade-off between efficiency and equality.

Optimal income taxation in the presence of market power falls short of the first-best outcome but can still reduce inequality and incentivize production by lowering taxes on labor, thereby increasing the after-tax labor share. Meanwhile, profit tax rates rise to fund transfers to poorer households. At the same time, policymakers can use taxation to mitigate misallocation between firms with different productivity levels, which results in a relatively regressive profit tax compared to the labor income tax.

APPENDIX

A Environment

A.1 The Cournot Competitive Tax Equilibrium

When the first-order conditions are both necessary and sufficient for the individuals' and final good producer's problems, the equilibrium allocations are governed by conditions (15) to (21) alongside the individuals' budget constraints. Given the technology considered in this paper and assuming $\phi_o(l_o) = \frac{l_o^{1+\frac{1}{\varepsilon_o}}}{1+\frac{1}{\varepsilon_o}}$, the following conditions hold 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 (23) (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 (24). Substituting $1 + \frac{\partial \ln P\left(Q_{ij}(\theta_e), \theta_e\right)}{\partial \ln Q_{ij}(\theta_e)}$ by (24), 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

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

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 nested-CES 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:

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

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 Nested-CES Economy

Wage Elasticity. Recall that in the nested-CES 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. By solving the final good producer's problem, we immediately derive the price equation (A1) and the inverse demand function (A6). Utilizing the definitions of price elasticities, we obtain the following results for 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

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

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 a 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. Thus, aggregate variables, including the outputs of final goods and the price of labor factors, remain unchanged. As in Scheuer and Werning (2017), the elasticity derived here is a micro elasticity.

The optimal choices 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 (19) and (24), and equation (A14) is derived by (21) and (24).

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) yields:

$$\frac{1+\varepsilon_e}{\varepsilon_e}\frac{dl_e}{l_e}=\frac{dL_w}{L_w}\left[1-\frac{y_eT_e''(y_e)}{1-T_e'(y_e)}\right]-\frac{d\tau}{1-T_e'(y_e)}.\tag{A17}$$

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. Consider again a small increase (i.e., $d\tau$) in the marginal tax rate faced by the θ_e-type entrepreneur. While this tax reform has no first-order effects on aggregate values or the actions of other types, it now affects aggregate variables, including the outputs of final goods and the price of labor factors, through changes in $Q_{-ij}(\theta_e)$. Define $\widetilde{\varepsilon}_{1-\tau_e}^{y_e}(\theta_e) \equiv -\frac{dy_e(\theta_e)}{y_e(\theta_e)} / \frac{d\tau}{1-T_e'(y_e(\theta_e))}$ as the elasticity of profit in this context. We refer to $\widetilde{\varepsilon}_{1-\tau_e}^{y_e}(\theta_e)$ 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.

The conditions in equations (A15) and (A16) still hold. However, when taking total differential, the change in $Q_{-ij}(\theta_e)$ must now be considered. Note that $Q_{-ij}(\theta_e)$ equals $Q_{ij}(\theta_e)$. Taking the total differential of equation (A15) yields:

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

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:

$$\tilde{\varepsilon}_{1-\tau_e}^{y_e}(\theta_e) = \frac{1}{\frac{1+\varepsilon_e}{\varepsilon_e}\frac{\partial \ln Q_{ij}(\theta_e)}{\partial \ln l_e(\theta_e)}\left[\frac{1}{\varepsilon_{Q_{ij}}^S(\theta_e)\frac{\partial \ln Q_{ij}(\theta_e)}{\partial \ln l_e(\theta_e)}} - \xi\right] - \left[1 - \frac{y_e(\theta_e)T_e''(y_e(\theta_e))}{1-T_e'(y_e(\theta_e))}\right]}. \tag{A21}$$

The general profit elasticity captures the extent to which a marginal profit tax reform affects firm-level profits by altering both the firm's own decisions and its competitors' outputs. In contrast, the partial elasticity reflects only the firm's direct response 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 (34) 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 unique 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,

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

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

Equation (A24) and problem (A23) implies:

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

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 $\theta_e^*(\theta_e)$ 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

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

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

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

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

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

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, (34), 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 (34) when the agent reports the true types. In conclusion, (34) 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.[52] The Lagrangian function for the planner's problem is

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

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.

[52]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

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

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

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

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:

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

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:

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

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 $\lambda N_e f_e(\theta_e)\,P(\theta_e)\,\frac{\partial Q_{ij}(\theta_e)}{\partial L_e(\theta_e)}$:

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

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

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

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

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

or

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

Combining (A52) and (A42) gives:

$$\begin{aligned}\frac{\widetilde{\tau}_e(\theta_e)}{1-\widetilde{\tau}_e(\theta_e)} &= [1-\bar{g}_e(\theta_e)]\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{1+\varepsilon_e}{\varepsilon_e}\left[\begin{array}{c}\mu(\theta_e)\frac{d\ln P_{ij}\left(Q_{ij},Q_{-ij}(\theta_e),\theta_e\right)}{d\theta_e}|_{Q_{ij}=Q_{ij}(\theta_e)} \\ +\frac{x_e'(\theta_e)}{x_e(\theta_e)}\end{array}\right] \\ &\quad -\frac{1}{1-\widetilde{\tau}_e(\theta_e)}\frac{1}{P(\theta_e)Q_{ij}(\theta_e)}\frac{\kappa'(\theta_e)}{\lambda N_e f_e(\theta_e)} \\ &\quad +\frac{1}{1-\widetilde{\tau}_e(\theta_e)}\frac{1}{P(\theta_e)Q_{ij}(\theta_e)}\frac{\varphi'(\theta_e)\varepsilon_{l_e}^{\omega}(\theta_e)}{\lambda N_e f_e(\theta_e)}.\end{aligned}$$

Using (A47) to substitute $\frac{\varphi'(\theta_e)\varepsilon_{l_e}^{\omega}(\theta_e)}{\lambda N_e f_e(\theta_e)}$ in the above equation,[53] we have:

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

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

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

$\kappa'(\theta_e)$ with (A46), we have the following equation:[54]

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

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:

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

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

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

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

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

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

we have

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

which is equivalent to

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

(b) We now try to express the right side of the above equation in terms of parameters. Using (OA30), we have:

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

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

Notice that:

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

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 $\theta_e\in\Theta_e$:

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

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., (43)), we have (A63):

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

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 (41) in the nested-CES economy.■

C Extensions and Robustness

C.1 Alternative Market and Taxation Specifications

Below, we explore the relationship between our results and four distinct issues that have been extensively studied in the literature. We frame these discussions through the lens of our model, leveraging its unique features and notation where necessary to derive new insights.

(i) Top Incomes: Market Power and Optimal Profit Tax. The taxes paid by top earners constitute the majority of income tax revenue, making the design of optimal tax rates for high-income individuals critical. Proposition 4 offers an analytic formula for determining the optimal top profit tax rate, with specific application to high-profit entrepreneurs (denoted as those with $\theta_e \geq \widehat{\theta}_e$). 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 4 *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 nested-CES economy:*

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

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

(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) extends the traditional top income tax formula, generalizing the top tax rate result of Saez (2001) (where $\xi = 0$, $\mu = 1$, $I = 1$, $\sigma \to \infty$) by incorporating more complex technology and market structures. Compared to Corollary 5 of Sachs et al. (2020) (where $\xi = 0$, $\mu = 1$, and $I = 1$), it highlights the influence of market structure and superstar effects on the optimal top tax rate. Similarly, relative to Scheuer and Werning (2017) (where $\mu = 1$, $I = 1$, and $\sigma \to \infty$), it emphasizes the interplay between superstar effects and market structure.

Proposition 4 is particularly valuable as it suggests that, under reasonable assumptions, observable statistics like $\widehat{H}_e$ can be used to derive the optimal top profit tax rate and assess whether it should be adjusted in response to changes in technology and market structure. A crucial question here is: should the government increase the top profit tax rate as market power rises? Condition (A64) indicates that whether the optimal top profit tax rate increases depends heavily on the value of $\widehat{H}_e$. Specifically, if $\widehat{g}_e \to 0$, condition (A65) becomes equivalent to $\widehat{H}_e > \frac{1}{2}$. This condition is sufficient, though not necessary, for $\widehat{\tau}e$ to increase with $\widehat{\mu}$.[55]

Empirical evidence supports this conclusion. The hazard ratio of top income in the United States was approximately 0.5 in 1992 and 1993 (Saez (2001)) and about $\frac{2}{3}$ in 2005 (Diamond and Saez (2011)). By 2007, the hazard ratios for top labor, capital, and total incomes were approximately 0.62, 0.76, and 0.71, respectively (Saez and Stantcheva (2018)). These findings suggest that condition (A65) is met in the U.S., indicating that the top profit tax rate should rise as market power increases.

(ii) Span of Control and Optimal Top Profit Tax. Similar to our model, Scheuer and Werning (2017) also explore optimal income taxation within a Lucas (1978) span of control framework. They examine the superstar effect, which arises from positive assortative matching between the abilities of entrepreneurs and the scale of their firms. In their model, the span of control reflects the magnitude of the superstar effect, representing the extent to which entrepreneurs can amplify the productivity of their workers.

In our framework, Proposition 4 provides valuable insights into how the span of control influences optimal taxation. Drawing from Scheuer and Werning (2017), an increase in ξ impacts both the elasticity of profits with respect to the tax rate and the hazard ratio of profit. Interestingly, when $I = 1$, these two effects offset each other, rendering the top profit tax rate independent of ξ. To understand this, note that under $I = 1$, the markup is uniform ($\mu = \widehat{\mu} = \frac{\sigma}{\sigma-1}$), simplifying formula (A64) 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 (40), i.e., $\frac{\xi}{\frac{\sigma}{\sigma-1}-\xi}$.

[55] In Online Appendix OC.2, we provide a more relaxed 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 bracketed term is at least 2, as $\mu \leq \frac{\sigma}{\sigma-1}$.

(iii) Market Structure, Indirect Redistribution, and Optimal Taxation. An important finding of this paper is the pivotal role of market structure in determining the optimal tax, particularly in shaping the indirect redistribution effect of taxation. In many prior studies on endogenous prices and optimal taxation, taxes have a first-order effect on relative prices, which can help ease 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 labor (wage) decreases with labor inputs, the planner can compress the wage distribution by reducing the marginal tax rate on high-skilled agents, thereby encouraging their labor supply. Saez (2004) contends that the indirect redistribution effect of taxes diminishes when agents make endogenous human capital investments, as they effectively set their own wages, reducing the tax's influence on prices to a second-order effect. In response, Naito (2004) shows that when human capital is imperfectly substitutable, the indirect redistribution effect remains relevant.

Our findings advance this discussion by showing that the extent of indirect redistribution depends critically on market structure. Even in the presence of market power, indirect redistribution persists, but its magnitude decreases as market power intensifies. In fact, the IRE vanishes entirely under monopoly, where firms independently set their prices. Conversely, under competition or oligopoly, where prices are determined fully or partially outside the firm, IRE still plays a role. This insight reconciles the seemingly conflicting conclusions of Saez (2004) and Naito (2004): Saez (2004) examines a scenario where agents act as monopolistic suppliers of their own factors, while Naito (2004) assumes a competitive labor market.

(iv) Endogenous Social Welfare Weights. In much of the analysis above, we assume exogenous social welfare weights, thereby isolating the influence of market power on optimal taxation through the four key components. This abstraction allows us to focus on the direct effects of market power.

Our analysis of the laissez-faire economy suggests that entrepreneurs' gross utility generally increases with rising market power. An increase in market power, accompanied by higher markups, redistributes income from workers to firms (primarily through lower wages) and reduces overall welfare. If similar outcomes hold under optimal taxation, the marginal social welfare weights for entrepreneurs will decline, leading to higher optimal profit tax rates. This result aligns with findings in prior studies (e.g., Kushnir and Zubrickas (2019)). Additionally, a non-linear tax system can facilitate transfers between entrepreneurs and workers, contingent on the social welfare function. Under a utilitarian framework, the burden distribution is indeterminate, but it becomes determinate under a concave welfare function, where market power plays a critical role.

Segregating the impact of endogenous social welfare weights is essential because generalized welfare weights may depend on factors beyond gross utility, such as market structure and revenue contributions (see Saez and Stantcheva (2016); Scheuer (2014)). For example, welfare weights might reflect the social versus private value of entrepreneurship. By isolating this endogeneity, Proposition 3 remains applicable. Empirically, we show that the optimal top profit tax rate increases with rising markups, supporting prior studies' conclusions.

C.2 Alternative Technology Specifications

(i) Capital Investment. Our benchmark model does not explicitly incorporate capital, but it can be adapted to do so. The framework remains valid if entrepreneurial effort is replaced with capital investment, particularly when some factor costs are non-deductible before taxation. Consider an economy where entrepreneurs choose labor inputs L_w and capital K, instead of effort:

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

where $Q_{ij}\left(K,L_w\right)$ is the firm-level production function of capital and labor inputs, r is the market price of capital,[56] and $\phi_K\left(K,\theta_e\right)$ is the unobservable cost of investment.

In practice, while the market price of capital (r) is observable, the total opportunity costs – such as fundraising or management – are often unobservable, captured by $\phi_K\left(K,\theta_e\right)$.[57] 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, $y_e = P_{ij}\left(Q_{ij}\left(K,L_w\right),Q_{-ij}\left(\theta_e\right),\theta_e\right)Q_{ij}\left(K,L_w\right)-WL_w-rK$.

In reality, incomplete deductibility of investment costs is common. For instance, while debt interest is deductible, equity investments often are 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. This distinction suggests that profits tax can effectively function as a tax on capital (see Myles (2008), Chapter 8).

Our core results remain robust under these scenarios. The fundamental issue in the incentive problem is the unobservability of certain inputs, and our model can accommodate both deductible and non-deductible factors. Consequently, the optimal profit tax formula derived here applies broadly, regardless of specific input configurations.

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

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

where s is the entrepreneur's share of the profits, and the remaining profits are distributed evenly among taxpayers (or households).

In this setup, the planner's problem and the constrained optimal allocation remain unchanged. How-

[56] In a dynamic or open economy model, r could be exogenous or determined by capital production technology, making the inclusion of K more intuitive (e.g., see Cui et al. (2021)).

[57] 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 $\phi_K\left(K,L_w,\theta_e\right)$).

ever, the optimal taxation formula is adjusted to incorporate performance pay. Specifically, the tax wedges satisfy: $\tau_s(\cdot) = t_s$, $\tau_w(\theta_w) = T_w'(y_w(\theta_w))$ and $\tau_e(\theta_e) = 1 - (1 - t_s)[1 - T_e'(y_e(\theta_e))] \cdot s$. Therefore, introducing s won't change the effective tax rate on the effort of the entrepreneur but proportionally increase $1 - T_e'(y_e(\theta_e))$. Consequently, our main results remain valid.

(iii) Monopolistic Competition with Kimball Aggregation and Endogenous Markups. In some of our results, we employ a constant elasticity of substitution (CES) technology. Although markups are endogenous, tax policies do not influence the equilibrium markup under CES. Here, we consider an alternative setup using Kimball aggregation, which introduces non-constant elasticity of substitution. This modification allows taxes to affect markups. For tractability, we focus on monopolistic competition and the second-best allocation.

The technology is described below:

$$1 = \int_{\theta_e} \chi(\theta_e) Y\left(\frac{Q(\theta_e)}{Q/N_e}\right) dF_e(\theta_e), \tag{A67}$$

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

$$P(\theta_e) = \frac{\chi(\theta_e) Y'\left(\frac{Q(\theta_e)}{Q/N_e}\right)}{\int_{\theta_e} \chi(\theta_e) Y'\left(\frac{Q(\theta_e)}{Q/N_e}\right) \frac{Q(\theta_e)}{Q/N_e} dF_{\theta_e}(\theta_e)} \tag{A68}$$

$$\text{and} \quad \mu(\theta_e) = \frac{\varepsilon(\theta_e)}{\varepsilon(\theta_e) - 1} \text{ with } \varepsilon(\theta_e) = -\frac{Y''\left(\frac{Q(\theta_e)}{Q}\right)\frac{Q(\theta_e)}{Q}}{Y'\left(\frac{Q(\theta_e)}{Q}\right)}. \tag{A69}$$

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:

$$V_e'(\theta_e) = \phi_e'(l_e(\theta_e)) l_e(\theta_e)\left[\mu(\theta_e)\frac{\chi'(\theta_e)}{\chi(\theta_e)} + \frac{x_e'(\theta_e)}{x_e(\theta_e)}\right], \forall \theta_e \in \Theta_e. \tag{A70}$$

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 (7) subject to the resource constraints (A67) and (16), the labor market clear condition (17), and the incentive conditions (A70) and (31). As a comparison to the optimal profit tax under monopolistic competition in the benchmark model, we now have the following proposition:

Proposition 5 *Under monopolistic competition with Kimball aggregation, the effective tax rate on entrepreneurial effort satisfies:*

$$\frac{1 - \frac{1-\tau_e(\theta_e)}{\mu(\theta_e)}}{\frac{1-\tau_e(\theta_e)}{\mu(\theta_e)}} = [1 - \bar{g}_e(\theta_e)] \frac{1 - F_e(\theta_e)}{f_e(\theta_e)} \left[\begin{array}{c} \frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu(\theta_e)\frac{\chi'(\theta_e)}{\chi(\theta_e)} + \frac{x_e'(\theta_e)}{x_e(\theta_e)}\right] \\ +\mu(\theta_e)\frac{\partial \ln \mu(\theta_e)}{\partial \ln Q(\theta_e)}\frac{\chi'(\theta_e)}{\chi(\theta_e)} \end{array}\right], \tag{A71}$$

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 (47)), the optimal tax here accounts for the endogeneity of markups. When $\frac{\partial \ln \mu(\theta_e)}{\partial \ln Q(\theta_e)} > 0$, the additional term in the tax formula is positive, suggesting that higher markups generally necessitate higher tax rates. Therefore, our main finding – that the top profit tax rate should increase with rising markups – remains robust under varying elasticities of substitution.

(iv) Free Entry. In this paper, we assume that the number of incumbents in a market is exogenous, meaning our analysis does not capture the impact of taxes on markups and efficiency through the extensive margin of firm entry. This omission is important because entry plays a significant role in shaping real-world policy outcomes. However, fixing the number of incumbents provides a reasonable starting point, as introducing an extensive margin complicates the analysis with an entry game that is challenging both analytically and computationally.

For example, De Loecker et al. (2019) analyze a model with endogenous markups 'a la Atkeson and Burstein (2008) and firm entry. Their entry game builds on Berry (1992) and involves complex iterations to compare all possible entry configurations. Combining such a framework with incomplete information and optimal taxation would likely render the solution intractable. Future research could extend our model to include firm entry, thereby analyzing the effects of taxation on both the intensive and extensive margins for more nuanced policy insights. There are several ways to introduce the extensive margin. Various approaches could be adopted to incorporate the extensive margin. For instance, Scheuer (2014) and Rothschild and Scheuer (2013) introduce occupational choice, while Edmond et al. (2023) consider an economy with free entry under Kimball demand. Another approach might involve adding a fringe of small businesses to each market.

(v) Uniform Income Tax. Thus far, we have considered differentiated tax policies for labor income and profits. However, in practice, governments may not perfectly distinguish between these two income sources. This raises the question of optimal policy under a uniform tax on both profit and labor income.

Introducing this policy constraint complicates the problem significantly. As an illustration, let $\theta_w(\theta_e)$ denote the ability of a worker whose income matches that of an entrepreneur with ability $y_e(\theta_e)$ (i.e., $y_e(\theta_e) = y_w(\theta_w(\theta_e))$). The first-order conditions for both types imply:

$$1 - T_w'(y_e(\theta_e)) = \frac{\phi_w'\left(\frac{y_e(\theta_e)}{W\varkappa_w(\theta_w(\theta_e))}\right)}{W\varkappa_w(\theta_w(\theta_e))}, \quad \text{and} \quad 1 - T_e'(y_e(\theta_e)) = \frac{\phi_e'(l_e(\theta_e))}{\frac{P(\theta_e)}{\mu(\theta_e)}\frac{\partial Q_{ij}(\theta_e)}{\partial l_e(\theta_e)}}. \tag{A72}$$

Thus, under a uniform tax, the following policy constraint applies:

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

where $y_e(\theta_e) = P\left(Q_{ij}(\theta_e), \theta_e\right) Q_{ij}(\theta_e) - WL_w(\theta_e)$. The planner must treat $\theta_w(\theta_e)$ as an additional decision variable (see Fu et al. (2021) for solutions to this type of problem).

Beyond the complexity, there is also a conceptual reason for differentiating labor and profit taxes in our benchmark model. Entrepreneurial income often includes corporate profits, which are distinct from labor income. Even after profits are declared as personal income, they remain conceptually different from wages. Therefore, differentiated taxation is a realistic and practical starting point, especially when analyzing the impact of rising market power.

Both differentiated and uniform taxation approaches are well-explored in the literature (see Scheuer (2014) and Rothschild and Scheuer (2013)).[58] Given the real-world tax distinctions, we prioritize differentiated taxation in our analysis to better understand its role in addressing rising market power.

[58] Scheuer (2014) consider both cases, referring to the uniform policy constraint as a "no-discrimination" constraint.

References

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 (2019): "Markups and Inequality," Tech. Rep. w25952, National Bureau of Economic Research.

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.

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.

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.

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.

Jaimovich, N. and M. Floetotto (2008): "Firm dynamics, markup variations, and the business cycle," *Journal of monetary Economics*, 55, 1238–1252.

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.

——— (2014): "A Theory of Income Taxation Under Multidimensional Skill Heterogeneity," SSRN Scholarly Paper ID 2380471, Social Science Research Network, Rochester, NY.

——— (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. (2013): "Adverse selection in credit markets and regressive profit taxation," *Journal of Economic Theory*, 148, 1333–1360.

——— (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.