DISCUSSION PAPER SERIES

DP16011
(v. 2)

Optimal Taxation and Market Power

Jan Eeckhout, Chunyang Fu, Wenjian Li and Xi Weng

INDUSTRIAL ORGANIZATION

PUBLIC ECONOMICS

ISSN 0265-8003

Optimal Taxation and Market Power

Jan Eeckhout, Chunyang Fu, Wenjian Li and Xi Weng

Discussion Paper DP16011
First Published 06 April 2021
This Revision 01 October 2021

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

Optimal Taxation and Market Power

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. As our main theoretical contribution we obtain explicit analytical expressions for the optimal tax rate as a function of market power. Our results show that optimal taxation can substantially increase welfare, but also highlight the severe constraints that the Planner faces to correct the negative externality from market power, using the income tax as a Pigouvian instrument. Pigouvian taxes compete with Mirrleesian incentive concerns, which generally leads to opposing forces in profit tax design. Overall, in our numerical analysis, we find that market power tends to lower marginal tax rates on workers, whereas it increases the marginal tax rate on entrepreneurs.

JEL Classification: D3, D4, J41

Keywords: optimal taxation, Optimal profit tax, market power, market structure, Markups

Jan Eeckhout - jan.eeckhout@upf.edu
UPF Barcelona and CEPR

Chunyang Fu - fuchunyang@pku.edu.cn
UCASS

Wenjian Li - wenjian_li@pku.edu.cn
Zhejiang University

Xi Weng - wengxi125@gsm.pku.edu.cn
Peking University

Optimal Taxation and Market Power*

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

September 30, 2021

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. As our main theoretical contribution we obtain explicit analytical expressions for the optimal tax rate as a function of market power. Our results show that optimal taxation can substantially increase welfare, but also highlight the severe constraints that the Planner faces to correct the negative externality from market power, using the income tax as a Pigouvian instrument. Pigouvian taxes compete with Mirrleesian incentive concerns, which generally leads to opposing forces in profit tax design. Overall, in our numerical analysis, we find that market power tends to lower marginal tax rates on workers, whereas it increases the marginal tax rate on entrepreneurs.

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, Aleh Tsyvinski as well as numerous colleagues and seminar participants for useful comments and insightful discussions. Renjie Bao provided superb research assistance. We gratefully acknowledges support from the China Information Economics Association, the Key Laboratory of Mathematical Economics and Quantitative Finance, the Ministry of Education and Peking University for funding support, from the National Natural Science Foundation of China (Grant No. 72003198), from the ERC, Advanced grant 882499, from AEI/FEDER, UE-PGC2018-096370-B-I00.

† jan.eeckhout@upf.edu – ICREA-GSE-CREi
‡ fuchunyang@pku.edu.cn
§ wenjian_li@zju.edu.cn
¶ wengxi125@gsm.pku.edu.cn

1 Introduction

Market power has an impact on both inequality and efficiency. As market power increases, the share of output accrues disproportionately to owners of monopolistic firms and less to workers. In addition, market power creates inefficiencies in the allocation of resources as prices are too high which leads to deadweight loss and a reduction in welfare. Therefore, we ask whether taxes should reflect the extent of market power, and if so, how? In this paper, we aim to answer the question by investigating optimal taxation in conjunction with market power. Starting with Mirrlees (1971), an extensive and influential literature on optimal taxation has analyzed what determines the properties of income tax schedules. Given that market power changes both efficiency and inequality, understanding the effect of market power on optimal tax rates is an important objective, especially in light of the rise of market power in recent years. We, therefore, contribute to the existing literature by embedding market power in an otherwise canonical setting of optimal income taxation.

The most obvious way to address the distortionary effect of market power is to eradicate the root cause of market power itself with antitrust policy. But the optimal antitrust policy may not be achievable,[1] so instead we ask what optimal policy should be when we can rely on income and goods taxation only. The Mirrleesian tax provides the correct incentives that trade-off efficient effort supply with inequality. In addition, now the optimal tax system simultaneously corrects the externalities that derive from market power in the goods market. The income tax thus also plays the role of a Pigouvian tax: a tax that corrects a market failure, whether it be pollution or in this case, market power. An important insight of our analysis is how to optimally trade-off different objectives: inequality, efficiency, and correcting externalities from market power.

Our contribution is twofold. First, in an otherwise canonical Mirrlees (1971) taxation framework, we embed endogenous market power as well as a clear distinction between wage-earning workers and profit-earning entrepreneurs. The novelty in the setup is that we add the inefficiency of market power in the hands of entrepreneurs which interacts with the unobservable effort supply of heterogeneous agents. We do this in a setting that allows for oligopolistic competition between a finite number of firms. We show that this model captures a number of empirically relevant features that link inequality to market power, in particular, how market power creates inequality. Even under Laissez-faire, our model generates novel predictions regarding the effect of market power on equilibrium allocation and inequality. Second, we derive the optimal taxation policy to implement the planner's second-best allocation. Our results advance the optimal taxation literature by deriving *explicit* analytical expressions for the tax rate as a function of market power. The nature of the optimal tax scheme now combines a Pigouvian correction of the externality due to market power with the design of the Mirrleesian incentive problem.

The main result of our analysis can be divided into two parts: the Laissez-faire economy and the economy under optimal taxation. In the absence of taxes, the Laissez-faire equilibrium predicts an increase in

[1] Antitrust policy faces many challenges, not least because the determinants of market power have multiple origins that can often not easily be corrected: those origins based in technology such as entry barriers, returns to scale and the heterogeneity in productivity between firms; and those based on the market such Mergers and Acquisitions. See amongst others Sutton (1991, 2001) and De Loecker, Eeckhout, and Mongey (2019).

inequality as aggregate markups increase. The labor income of workers decreases due to the decline in the general equilibrium wage rate as well as the decline in their hours worked in response to the lower wage rate. At the same time, entrepreneurs see an increase in their income. This is consistent with the decline in the labor share that has been documented and that coincides with the rise of market power.[2] The rise of market power also results in a decrease in output and social welfare. In addition, inequality within the pool of heterogenous entrepreneurs increases while inequality within the pool of workers remains constant. Because each entrepreneur owns a different firm, rising markups lead to higher dispersion in productivity and profits between firms, yet the inequality between workers within the firm remains unchanged. This feature of our model where inequality between firms increases is consistent with the facts on increasing between-firm inequality.

In the presence of market power, the optimal taxation policy has the following properties.[3] First, when all agents are identical, the government only needs to address the incentives to provide effort without concern for heterogeneous abilities. Then the marginal tax rate is negative and declines as markups increase. In the absence of market power, the marginal tax rate would be zero. While markups create a distortion, they also lower the incentives to work. The workers work less because the wages are lower, and the entrepreneurs work less because they price higher and thus sell and produce less. The Planner, therefore, offers incentives through negative marginal tax rates to both workers and entrepreneurs. Second, when agents are heterogeneous, the marginal tax rate for all agents now reflects the motive for redistribution and depends on the type of agent, with lower tax rates for low earners and higher tax rates for high earners. Interestingly, for any two types of entrepreneurs and workers and under monopolistic competition, the net marginal tax rate of entrepreneurs is now higher (less negative) relative to that of workers because the Planner takes into account also the effect of market power in the incentive constraint. Intuitively, raising market power expands the productivity within the pool of entrepreneurs, which requires higher marginal profit tax rates to narrow the income inequality within the group of entrepreneurs. Third, when there is oligopolistic competition, because of strategic interaction between competitors in the same market, the marginal tax rate for entrepreneurs decreases again (becomes more negative). This is because a decreasing profit tax raises the output of firms which in turn decreases the price of firm-level output and relaxes the entrepreneurs' incentive constraint. And fourth and finally, once markups are heterogeneous, the marginal tax rate of entrepreneurs changes depending on the entrepreneurs' productivity. The marginal tax rate is now lower for high-productivity entrepreneurs. The Planner wants to use the tax incentives to induce more productive entrepreneurs to produce more, which it does with lower marginal tax rates.

While our most general result gives an implicit relation between market power and the optimal tax rate, an important contribution of our paper is to obtain explicit expressions for the tax rates in terms of the markup under certain conditions. For example, for a parametric restriction on the labor supply elasticity

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

[3]We allow for taxes on the income of workers and entrepreneurs, and also on the sales of consumer goods. However, we show that we can focus attention exclusively on the tax on entrepreneurs and workers, and not on the sales tax. It is well-known in the literature (see for example Chari and Kehoe (1999) and Golosov et al. (2003)) that multiple tax policies can implement the same second-best allocation. In our setting, because the entrepreneurs are the residual claimants of output, a sales tax can without further distortions equivalently be substituted by a levy on the entrepreneurs' profits. Therefore, we assume sales taxes are zero.

and the elasticity of substitution in production in conjunctions with the returns to scale, we can analytically solve a key integral equation. Another analytical solution is for the case of elasticities of substitution in production between different markets that are common across markets. To obtain these results, we rely on the theory of integral equations to formalize the model and to solve these important cases. The setup with endogenous markups leads to a complicated form of the integral equations in which tax formulas are expressed. Despite those mathematical complexities, we are able to solve for those equations analytically. These explicit analytical solutions are a major theoretical advance of the optimal taxation literature.

In sum, the optimal marginal tax rates are a delicate balancing act trading off incentives to produce output with the distortions from market power as well as the desire to restore equity. The effect of correcting the externality from market power is, somewhat surprisingly, a decrease of marginal tax rates for both workers and firms as market power goes up. However, there are also several indirect effects from market power that affect the entrepreneur only. Two channels decrease the marginal tax rate (indirect redistribution and reallocation) while one channel generally increases the tax rate (technological differences across sectors that affect markups and hence the productivity of entrepreneurs).

Of course, the average taxes paid and the tax burden are not necessarily evolving in the same direction as the marginal tax rates. After all, we know that market power has a general equilibrium effect that lowers the wage rate, which increases inequality between workers and entrepreneurs. Whereas the marginal tax rates ensure an optimal allocation of resources and production through optimal incentive provision, the tax burden determines the optimal redistribution of income for a given social welfare function. Because we cannot analytically solve for the tax burden in our model, we simulate the economy and map the full implications of optimal taxation.

Our simulations show that as market power increases, the wage rate, output, and welfare decrease. At the same time, profits increase and the labor share declines. Both entrepreneurs and workers supply less labor as market power increases. Entrepreneurs are better off in consumption and utility, and workers are worse off. The planner's optimal taxation response sets a positive marginal tax rate on average for both workers and entrepreneurs, and it increases market power for entrepreneurs while it decreases for workers. The lump-sum tax for entrepreneurs is substantially higher than for workers. For entrepreneurs, the average tax rate and the total tax burden are positive and increase in market power, for workers it is negative and decreases. Across types within an occupation, higher-skilled agents (in both occupations) face lower marginal tax rates (as is the case for superstars, Scheuer and Werning (2017)), but the tax burden is non-monotonic. The optimal taxation allocation increases welfare relative to Laissez-faire (trivially), but output and the wage rate are lower, and markups increase. With taxes, inequality decreases substantially, especially for entrepreneurs.

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

Related Literature. There is a growing policy literature on the relation between markups and inequality (e.g., see Stiglitz (2012); Atkinson (2015); Baker and Salop (2015); Khan and Vaheesan (2017)), yet existing optimal tax papers with market power generally focus on indirect taxes, which abstracts from distribution concerns (Stern (1987); Myles (1989), Cremer and Thisse (1994); Anderson, Palma, and Kreider (2001); Colciago (2016); Atesagaoglu and Yazici (2021)). These papers generally assume that lump-sum tax is not enforceable and study how can the government raise revenue efficiently. In a recent paper, Atesagaoglu and Yazici (2021) analyze the effect of optimal taxation on the labor share in a Ramsey problem with capital. They ask a different but related question, namely whether it is optimal to tax capital rather than labor when there is pure profit and the planner cannot distinguish capital income from profits.

We embed a Mirrleesian tax problem into an economy with market power.[4] To model the economy with imperfect competition, we introduce market power in a market framework similar to Atkeson and Burstein (2008) where a finite number of oligopolistic firms have market power in their local market.[5] This setting allows us to model the influence of market structure on the optimal design of the tax system, which is often ambiguous in the literature. The technology that an entrepreneur employs is as in Lucas (1978), where a skilled entrepreneur chooses the optimal amount of labor as an input to produce output. Unlike Lucas (1978), the entrepreneur has market power and chooses prices strategically when reporting their types. Therefore the presence of market power in the principal-agent problem is notably distinct from the existing literature on optimal taxation with market power (e.g., see Kaplow (2019), Kushnir and Zubrickas (2019), Jaravel and Olivi (2019), Gürer (2021), and Boar and Midrigan (2021)).[6] In this literature, Boar and Midrigan (2021) is the only paper that also introduces entrepreneurs. They consider an alternative incentive problem between the planner and the entrepreneur where a profit tax does not affect the entrepreneur's incentive constraint. As a result, their optimal policy prescription is quantity regulation instead of a profit tax. In addition, we consider different production technologies and market structures. The source of market power in our model is the number of firms that are in oligopolistic competition, instead of preferences via the Kimball aggregator in monopolistic competition. These different modeling choices have implications for the policy conclusions. Because in our setup the markup enters the entrepreneur's incentive constraint, it affects the optimal policy not only through the Pigouvian channel (correcting the market power externality), but the optimal profit tax depends also on the markup through the Mirrleesian channel (redistribution), and the Mirrleesian channel leads to an increase of the marginal tax rate as markups in-

[4]The taxation is Mirrleesian in the sense that both labor income and profit taxes are allowed to be arbitrarily non-linear and lump-sum taxes (or transfers) are enforceable.

[5]We have a nested CES structure in inputs of production, instead of in preferences over consumption goods.

[6]Agents in Kaplow (2019), Kushnir and Zubrickas (2019) and Jaravel and Olivi (2019) treat prices and profits as given. Thus, the strategic actions of agents are not dependent on the market power in these papers. Moreover, the share of profits received by the agents in their models are either zero or determined by exogenous functions of individual ability or labor income. Thus, if profit tax is introduced in their models, it acts as lump-sum taxes. While these papers study optimal labor income taxes, they do not analyze optimal profit taxes. Specifically, Kaplow (2019) and Jaravel and Olivi (2019) abstract the problem from profit tax and Kushnir and Zubrickas (2019) considers an exogenous profit tax.

crease. This is at the heart of the role that Mirrleesian taxes play as opposing forces of Pigouvian taxes. By considering different modeling choices and the resulting differences in findings and policy prescriptions, our paper and Boar and Midrigan (2021) offer complementary insights into the problem of taxation in the presence of market power.

Our paper is also related to the literature on optimal taxation with endogenous prices or wages (e.g., see Stiglitz (1982); Naito (1999) and Naito (2004); Saez (2004); Scheuer (2014); Sachs, Tsyvinski, and Werquin (2020); Cui, Gong, and Li (2021)). This literature emphasizes the general equilibrium effect of taxes on prices of factors, which brings an indirect redistribution between agents providing different factors. While most of these papers treat agents as price takers, agents in our model have price-setting power. We show that the indirect redistribution effect of taxes now is dependent on the market structure. In particular, a reduction in the profit tax encourages entrepreneurial effort and output, thereby decreasing the price of the competitor's product which leads to redistribution indirectly. Interestingly, when there is no competitor in the submarket, i.e., under monopolistic competition, this strategic effect of taxation disappears, which, together with the entrepreneur's price-setting, eliminates the tax policy's first-order effect on prices.

The paper also contributes to the literature on optimal taxation and technology (e.g., see Ales, Kurnaz, and Sleet (2015); Ales and Sleet (2016); Scheuer and Werning (2017); Ales, Bellofatto, and Wang (2017)). Diamond and Mirrlees (1971) and Scheuer and Werning (2017) observed that the parametric optimal tax rate is not dependent on the curvature of technology. Our results extend their findings to an economy with market power: the curvature of firm-level production technology (with respect to labor inputs) does not affect the optimal labor income and profit tax rates. On the other hand, we find a novel route for the technology to affect the optimal tax rate. Since the markup is dependent on the elasticity of substitution between products and the productivity of entrepreneurs are determined by the technology as well as the markup, markup affects optimal taxation together with the technology.

Lastly, our paper belongs to the literature on optimal taxation with externality (e.g., Sandmo (1975); Ng (1980); Bovenberg and van der Ploeg (1994); Kopczuk (2003); Farhi and Gabaix (2020)). As suggested by Kopczuk (2003), one of the main results of this literature is the "additivity property":[7] optimal taxation in the presence externalities can be expressed additively by some Pigouvian taxes. However, we find that the additivity property generally does not hold in an economy with heterogeneous agents and market power. Not only the Pigouvian tax – the tax used to restore efficiency – changes with the externality induced by market power, but also the redistributive tax does. This is because social welfare weights change with the extent of market power. And even if the social welfare weights are exogenous, the redistributive tax changes because of output is endogenous changes with market power.

2 The Model Setup

Environment. The economy is static. Production of the final consumption good needs the composite input of an intermediate good produced by an entrepreneur (idea), and the effort of workers.

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

Agents and Preferences. Agents belong to one of two occupations $o \in \{e, w\}$, entrepreneur or worker. The occupational types are fixed. Within each occupation, agents are heterogeneous in their productivity. Denote the type of an agent by $\theta_o \in \Theta_o \subset \mathbb{R}_+$, distributed according to the cumulative density function $F_o(\theta_o)$ with density $f_o(\theta_o)$. The measure of entrepreneurs is N; the measure of workers is normalized to one. There is a representative firm producing final goods in a competitive market and making zero profits.

Both worker and entrepreneur have a preference over consumption and effort. We denote by $U_o(\theta_o) = c_o - \phi_o(l_o)$ the utility function of an agent of type o (worker or entrepreneur), where l_o refers to working hours.[8] The cost of effort functions $-\phi_o(\cdot)$ are twice continuously differentiable and strictly concave. To make the analysis transparent and in the simulations, we will consider utility function with constant elasticity of labor supply, i.e., $\varepsilon_o \equiv \frac{\phi_o'(l_o)}{l_o \phi_o''(l_o)}$ is constant. We denote by $V_o(\theta_o)$ the optimal utility of agent θ_o.

Market Structure. The labor and final good markets are perfectly competitive. Instead, the intermediate goods market exhibits market power. There are two levels of production: intermediate inputs and final goods. The market structure in the intermediate goods market is a variation of the structure in Atkeson and Burstein (2008), but with product differentiation in production rather than in preferences.

At the intermediate goods level, identical entrepreneurs of type θ_e compete producing differentiated inputs, that consist of a small number of close substitutes (say Coke and Pepsi, or Toyota and Ford), and a continuum of less substitutable input goods (say soft drinks and cars). The most granular market is small, where a finite number of I entrepreneurs (with $I \geq 1$) of equal type θ_e produce a differentiated input good under imperfect competition. In this market, an entrepreneur $i = 1, ..., I$ Cournot competes against $I - 1$ competitors. The number of competitors I determines the degree of market power. The output produced within this market is differentiated with a common elasticity of substitution $\eta(\theta_e)$ across all I goods. There are a continuum of these imperfectly competitive markets, denoted by j with measure $J(\theta_e) = \frac{N f(\theta_e)}{I}$. Each of those markets j has I goods (with elasticity of substitution within the market $\eta(\theta_e)$ and produced by identical entrepreneurs θ_e), and the elasticity of substitution σ across markets (between soft drinks and cars) is smaller than within markets (between Coke and Pepsi): $\sigma < \eta(\theta_e)$. In order to rule out abnormal markup, throughout this paper we assume that σ is greater than 1.[9]

At the final goods level, the inputs produced in markets i, j by heterogeneous entrepreneurs θ_e is aggregated to a final output good with the same elasticity of substitution σ. Thus, one individual firm i in a market j that produces an intermediate with entrepreneurs θ_e is fully identified by the triple (i, j, θ_e).

Technology. Heterogeneous agents supply efficiency units of labor: an agent of type θ_o who works l_o hours supplies $x_o(\theta_o) l_o$ efficiency units of labor.[10] Because in general, the equilibrium labor inputs depends

[8] Our utility function is separable between consumption and labor, and we eliminate income effects. The assumption is crucial for the tractability of the optimal tax problem, and it is not crucial to the economic implication of this paper. The empirical literature using detailed micro data sets has typically not rejected a zero income elasticity on labor supply or found very small effects (e.g., see Gruber and Saez (2002); Kleven and Schultz (2014)). Readers who are interested in how the complementarity and substitution between consumption and labor can refer to Atkinson and Stiglitz (1976), Mirrlees (1976) and Christiansen (1984).

[9] See equation 22 below for details.

[10] The assumption of efficiency units drastically simplifies the solution of the model but it is not innocuous. The efficiency units assumption rules out sorting because firms are indifferent across worker types as long as they provide exactly the same efficiency

on the firm (i, j, θ_e), we denote the efficiency units of labor demand and entrepreneurial effort by $L_{w,ij}(\theta_e)$ and $x_e(\theta_e)l_{e,ij}(\theta_e)$ respectively, where $l_{e,ij}(\theta_e)$ is the working hours of entrepreneur.

The firm level production technology of the intermediate good is as in Lucas (1978), with one heterogeneous entrepreneur hiring an endogenous number of workers to maximize profits. Because the productivity of entrepreneurs and workers is expressed in efficiency units, the technology takes efficiency units as inputs instead of bodies. The quantity of output of a θ_e entrepreneur is therefore:

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

where $L_{w,ij}(\theta_e)$ is the *quantity* of labor in efficiency units the entrepreneur hires to work in the firm and $0 < \xi \leq 1$.[11] Note that because of the efficiency units assumption, output $Q_{ij}(\theta_e)$ does not depend on the worker types θ_w that are employed. There is no capital in our model. Therefore we assume that, as in Lucas (1978) or Prescott and Visscher (1980), the entrepreneur is the residual claimant of output, i.e., they "own" the technology θ_e. Therefore, the entrepreneur hires labor to maximize profits.

Given the technology, we can aggregate the firm-level output first within the market with I close substitutes (with elasticity $\eta(\theta_e)$) to $Q_j(\theta_e)$, then across all $J(\theta_e)$ markets (with elasticity σ) to $Q(\theta_e)$, and finally from aggregated inputs (with the same elasticity σ)[12] to output goods Q:

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

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

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

where $\widetilde{\chi}(\theta_e)$ is a distribution parameter. As illustrated by Ales et al. (2015), variations in $\chi(\theta_e)$ captures the technological or preference-based variations in demand for different skills and intermediate goods. To abstract from the love-of-variety effect related to I, we normalize the firm-level and market-level output by $I^{-\frac{1}{\eta(\theta_e)}}$ and $J(\theta_e)^{-\frac{1}{\sigma}}$. Then we introduce $\chi(\theta_e) = \widetilde{\chi}(\theta_e) f_{\theta_e}(\theta_e)^{-\frac{1}{\sigma}}$ as a modified distribution parameter (note that $J(\theta_e) = \frac{Nf(\theta_e)}{I}$).

units. See amongst others Sattinger (1975a), Sattinger (1993) and Eeckhout and Kircher (2018) how the assumption of efficiency implies an absence of sorting. To date, we know of no way how to solve the optimal taxation problem with market power in the presence of sorting.

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

[12]For notational simplicity and without loss, we assume the elasticity of substitution between intermediate inputs θ_e is the same as the the elasticity of substitution between inputs in different markets j as there is no market power at both levels of aggregation. The Key is that the elasticity within the small markets $\eta(\theta_e)$ where firms have market power is different from the elasticity across markets where there is a continuum of other products and hence 0 market power.

Prices, Wages and Market Clearing. Denote the price of intermediate goods produced by firm (i, j, θ_e) by $P_{ij}(\theta_e)$ and the income of a worker by $y(\theta_w)$ and the profits of an entrepreneur by $\pi(\theta_e)$. The profits of the entrepreneurs are determined by the fact that the entrepreneur is the residual claimant of revenue after paying for wages to the workers. The workers' wages are determined in a competitive labor market, subject to market clearing. Denote by W the competitive wage any firm pays for an efficient unit of labor.

Because of the efficiency wage assumption and competitive labor markets, there is a unique equilibrium wage W that solves market clearing for workers, given optimal labor supply l_w and optimal labor demand $L_{ij}(\theta_e)$. Obviously, labor supply increases in W and labor demand decreases in W. In the next equation we equating aggregate labor demand (left hand side) and aggregate labor supply (right hand side) to solve for equilibrium wages W:

$$\int_{\theta_e} \int_j \sum_{i=1}^{I} L_{w,ij}(\theta_e; W) dj d\theta_e = \int_{\theta_w} x_w(\theta_w) \, l_w(\theta_w; W) f_w(\theta_w) d\theta_w \tag{5}$$

Note that in equilibrium $l_{w,ij}(\theta_e) = l_w(\theta_w)$ for all θ_w because we assume labor markets are perfectly competitive and all firms pay the same W for one efficiency unit. Therefore, $y(\theta_w) = W x_w(\theta_w) \, l_w(\theta_w)$.[13]

Information. In the tradition of Mirrleesian taxation, we assume that types θ_o are not observable, while labor income $y(\theta_w) \in \mathbb{R}_+$ and profits of the entrepreneur $\pi(\theta_e) \in \mathbb{R}_+$ are observable. This assumption is equivalent to say that direct taxes can only depend on labor incomes and entrepreneurial profits.

We are interested in analyzing the role of wage and profit taxes in the presence of market power. To that effect we assume that markups are unobservable and the planner does not condition the optimal tax scheme on them. There are sound empirical reasons for this assumption. Markups are not easily obtained because they are the ratio of prices and marginal costs. While in some settings we have price information, the problem is obtaining information on marginal costs which are not directly observable. In the light of the academic research that is behind the estimation of markups, it is unlikely that the government will be able to easily generate markup estimates on which to condition taxes for all firms. It is important to stress that our assumption of unobservable markups does not imply that if markups were observable, the planner's solution would be first best. The reason is that markups, like profits, are endogenous and are vary with effort and output.[14] Moreover, in practice, trade between related parties can be used to change the reported markup (see Chari et al. (2012), where in their terminology, the markup would be a manipulable signal).

Policy, Taxation, and the Planner's Objective. We now specify how the government intervenes in the economy. Government uses taxation as an instrument to affect the equilibrium allocation in this economy. Following the literature, we assume the government levies taxes to collect an exogenous amount of revenue

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

[14] It is true that in equilibrium, there may be a one-to-one mapping between markups and types (though the monotonicity of the mapping is not guaranteed). But even then, if the planner bases taxes on the realized markup, then entrepreneurs will take the tax on markups into consideration when choosing their quantity or effort. Thus the planner can not achieve the first best with a tax on markups.

R. Given R, the government objective is to choose tax policies to maximize the social welfare:

$$\sum_{o\in\{w,e\}} N_o \int_{\theta^o} G\left(V_o(\theta_o)\right) \widetilde{f}_o\left(\theta_o\right) d\theta_o, \tag{6}$$

where $G : \mathbb{R}_+ \longmapsto \mathbb{R}_+$ is a twice differentiable social welfare function. We assume that both $G(\cdot)$ and $G'(\cdot)$ are strictly positive and $G''(\cdot) \leq 0$. The PDF $\widetilde{f}_\theta\left(\cdot\right)$ is a Pareto weights schedule, which is assumed to be continuous (e.g., see Saez and Stantcheva (2016)).

We assume that the government levies a linear sales tax, or a linear tax on labor inputs (such as salary tax), which is typically used in the real economy. In the Discussion section below, we consider non-linear sales tax.

Because types are not directly observable, the planner solves for the *constrained optimal allocation, or second best*. The first best allocation is unattainable given workers and entrepreneurs have private information over their type θ_o. Specifically, we consider that the government can use profit and labor income taxes $T_e : \pi \mapsto \mathbb{R}$ and $T_w : y \mapsto \mathbb{R}$ to be arbitrarily non-linear in the Mirrlees tradition. These direct taxes together with a sales tax $t_s \in \mathbb{R}$ compose the tax policy system $\mathcal{T} \equiv \{T_e, T_w, t_s\}$ that we consider in our benchmark model.[15]

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

3 Solution

3.1 The Cournot Competitive Tax Equilibrium

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

[15] Both the linear sales tax and linear tax on the salary pay act as tax wedges between the marginal cost and income of labor inputs $L_{w,ij}$. Since the prediction of optimal taxation is about tax wedges while not about specific tax policies (e.g., see Chari and Kehoe (1999); Golosov et al. (2003); Salanié (2003), pages 64-66), there is no need to introduce both of these indirect taxes. To see this, consider equation (10) below, where if we levy an additional tax t_l on the labor cost of the firm, the ratio of the marginal income of $L_{w,ij}$ to the marginal cost of $L_{w,ij}$ is $\frac{1+t_l}{1-t_s}$, which means the role of τ_l as a tax wedge can be replaced by t_s. Later in section 3.2, we will introduce the tax wedges considered in this paper.

The demand $Q_{ij}(\theta_e)$ for the intermediate input solves:

$$\Pi = \max_{Q_{ij}(\theta_e)} Q - \int_{\theta_e}\int_j \left[\sum_{i=1}^{I} Q_{ij}^{D}(\theta_e) P_{ij}(\theta_e)\right] dj d\theta_e, \tag{7}$$

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

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

All firms treat others' outputs as given. We denote by $P_{ij}\left(Q_{ij}(\theta_e), \theta_e\right)$ the inverse-demand function faced by the entrepreneur with firm (i, j, θ_e), whose problem is:

$$V_e(\theta_e) \equiv \max_{l_e, L_{w,ij}} c_e - \phi_e(l_e) \tag{8}$$

$$\text{s.t. } c_e = \pi_{ij} - T_e\left(\pi_{ij}\right) \tag{9}$$

$$\pi_{ij} = P_{ij}\left(Q_{ij}(\theta_e), \theta_e\right) Q_{ij}(\theta_e)(1 - t_s) - W L_{w,ij}(\theta_e), \tag{10}$$

where $Q_{ij}(\theta_e)$ is the quantity supplied of the intermediate good as defined in equation (1). Denote by $l_{e,ij}(\theta_e), c_{e,ij}(\theta_e), \pi_{ij}(\theta_e)$, and $L_{w,ij}(\theta_e)$ the solution to the above problem.

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

$$V_w(\theta_w) \equiv \max_{l_w} c_w - \phi_w(l_w) \tag{11}$$

$$\text{s.t. } c_w = W x_w(\theta_w) l_w - T_w\left(W x_w(\theta_w) l_w\right). \tag{12}$$

We denote by $c_w(\theta_w)$, and $l_w(\theta_w)$ the solution to (11). Besides, we denote by $y_w(\theta_w) = W x_w(\theta_w) l_w(\theta_w)$ the labor income of θ_w-type worker.

Market Clearing. Commodity and labor markets clearing require that for any (i, j, θ_e), the quantity demanded in the output sector $Q_{ij}^{D}(\theta_e)$ from equation (7) equals the quantity supplied $Q_{ij}^{S}(\theta_e)$ from equation (8):

$$Q_{ij}^{D}(\theta_e) = Q_{ij}^{S}(\theta_e) \tag{13}$$

and

$$Q = \int_{\theta_w} c_w(\theta_w) f_w(\theta_w) d\theta_w + \int_{\theta_e}\int_j \left[\sum_{i=1}^{I} c_{e,ij}(\theta_e)\right] dj d\theta_e + R, \tag{14}$$

and

$$\int_{\theta_w} x_w(\theta_w)\, l_w(\theta_w)\, f_w(\theta_w) d\theta_w = \int_{\theta_e}\int_j \left[\sum_{i=1}^{I} L_{w,ij}(\theta_e)\right] dj d\theta_e, \tag{15}$$

where R is the exogenous government revenue.

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

$$P_{ij}(\theta_e) = \frac{\partial Q}{\partial Q_{ij}(\theta_e)}, \tag{16}$$

and

$$\frac{W}{1-t_s} = \frac{\partial\left[P_{ij}\left(Q_{ij}(\theta_e), \theta_e\right) Q_{ij}(\theta_e)\right]}{\partial L_{w,ij}(\theta_e)}, \tag{17}$$

and

$$W\varkappa_w(\theta_w)\left[1 - T_w'\left(W\varkappa_w(\theta_w)\, l_w(\theta_w)\right)\right] = \phi_w'\left(l_w(\theta_w)\right), \tag{18}$$

and

$$\frac{P_{ij}(\theta_e)}{\mu_{ij}(\theta_e)}\frac{\partial Q_{ij}(\theta_e)}{\partial l_{e,ij}(\theta_e)}(1-t_s)\left[1 - T_e'\left(\pi_{ij}(\theta_e)\right)\right] = \phi_e'\left(l_{e,ij}(\theta_e)\right). \tag{19}$$

When first-order conditions are both necessary and sufficient to individuals' and final good producer's problems, the equilibrium allocations are determined by (13) to (19) and individuals' budget constraints.

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

Definition 1 *A Symmetric Cournot Competitive Tax Equilibrium (SCCTE) is a combination of tax system $\mathcal{T}$, symmetric allocation $\mathcal{A}$, and symmetric price system $\mathcal{P}$, such that given the policy and price system, the resulting allocation maximize the final good producer's profit (7); maximize entrepreneurs' utilities (8) subject to the budget constraint (9); maximize workers' utilities (11) subject to the budget constraint (12); the price system satisfies (17) and (16); and labor and commodity markets are cleared, i.e., (13) to (15) are satisfied.*

Note that we do not need to impose the government's budget constraint in our definition of SCCTE, since under the Walras's law, given the agent's budget constraints, and commodity market clear condition, the government's budget constraint must be satisfied.

We now make some common restrictions on the equilibria that we consider throughout the paper. First, we assume that the mechanisms (tax policies) are sufficiently differentiable. Second, we assume that:

Assumption 1 *In a Symmetric Cournot Competitive Tax Equilibrium:*

(i) $y(\theta_w)$ is differentiable, strictly positive, and strictly increasing in θ_w;

(ii) $\pi(\theta_e)$ *is differentiable, strictly positive, and positive increasing in* θ_e*;*

(iii) $\mu_{ij}(\theta_e)\frac{\partial \ln P\left(Q_{ij}(\theta_e),\theta_e\right)}{\partial\theta_e}+\frac{x_e'(\theta_e)}{x_e(\theta_e)}$ *is strictly positive.*

An incentive compatible allocation under the Spence-Mirrlees condition requires labor income to be non-decreasing in wage.[16] For simplicity, we assume that $y(\theta_w)$ is strictly increasing in θ_w, which in turn implies $x_w'(\cdot)>0$. With Assumption 1, we can define $F_y(y(\theta_w))=f_w(\theta_w)$ and $f_y(y(\theta_w))=F_y'(y(\theta_w))$ as the CDF and PDF of labor incomes. Besides, Assumption 1 excludes cases with mass points. Similar to the assumption on monotonicity of labor income, we assume monotonicity on $\pi(\theta_e)$. We define the distribution function of profits as $F_\pi(\pi(\theta_e))=F_e(\theta_e)$ with PDF $f_\pi(\pi(\theta_e))=F_\pi'(\pi(\theta_e))$.

Part three of Assumption 1 is used to guarantee that higher-skilled entrepreneur has a higher gross utility.[17] Such an assumption is needed to identify individuals of heterogeneous types when prices or wages of factors are endogenous (e.g., see Sachs et al. (2020) and Cui et al. (2021)).

Notation. In what follows, where there is no confusion, we will drop the subscript ij. For example, in the symmetric equilibrium the markup in each market $\{i,j,\theta_e\}$ is the same for all entrepreneurs with types θ_e. Therefore, we often denote the markup $\mu_{ij}(\theta_e)$ by $\mu(\theta_e)$ and the labor demand $L_{w,ij}(\theta_e)$ by $L_w(\theta_e)$.

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

$$\mu_{ij}(\theta_e)\equiv\frac{P_{ij}(\theta_e)}{MC_{ij}(\theta_e)}=\frac{P_{ij}(\theta_e)}{\frac{W}{\frac{\partial Q_{ij}(\theta_e)}{\partial L_{w,ij}(\theta_e)}(1-t_s)}}. \tag{20}$$

The firm's first order condition delivers a relationship, known as the Lerner Rule, between the inverse-demand elasticity $\varepsilon_{Q_{ij}}(\theta_e)\equiv\frac{\partial\ln P\left(Q_{ij}(\theta_e),\theta_e\right)}{\partial\ln Q_{ij}(\theta_e)}$ and markups $\mu(\theta_e)$.[18] As in Atkeson and Burstein (2008), under the benchmark technology with nested CES preferences, the inverse-demand elasticity can be written in weighted form:[19]

$$\varepsilon_{Q_{ij}}(\theta_e)=-\left[\frac{1}{\eta(\theta_e)}\left(1-s_{ij}\right)+\frac{1}{\sigma}s_{ij}\right]\geq-\frac{1}{\sigma}, \tag{21}$$

where s_{ij} is the sales share of firm i in market j. The markup is thus related to the demand elasticity:

$$\mu_{ij}(\theta_e)=\frac{1}{1+\varepsilon_{Q_{ij}}(\theta_e)}\leq\frac{\sigma}{\sigma-1}. \tag{22}$$

The higher the demand elasticity (the lower the inverse demand elasticity), the higher the markup. Therefore, the markup depends on the weighted sum of the elasticity of substitution between intermediate

[16] See e.g., see Salanié (2003), p. 87. When the Spence-Mirrlees condition is not satisfied, the analysis becomes much more complicated as local incentive compatibility becomes insufficient for global incentive compatibility (see, e.g., Schottmüller (2015)). These assumptions can be relaxed when considering free entry in the intermediate goods market, where individuals choose their occupations.

[17] See (B5) in Lemma B.1 for details.

[18] This follows from profit maximization, in equation (17), which implies $W=P_{ij}\left(Q_{ij}(\theta_e),\theta_e\right)[1+\gamma(\theta_e)]\frac{\partial Q_{ij}(\theta_e)}{\partial L_{w,ij}(\theta_e)}(1-t_s)$.

[19] See Appendix A.6.2 for details.

goods, and the intensity of competition in the submarket. The lower the $\eta(\theta_e)$ and σ, the less substitutable the goods are within and between markets, and the higher the markup. Most crucially, the markup increases as the sales share s_{ij}, and hence the number of competitors I, decreases within a market. The smaller the number of competitors I, the smaller the weight on the within market elasticity higher the weight on $\frac{1}{\eta(\theta_e)}$ and the higher the weight on $\frac{1}{\sigma}$. Firms that face little competition face little substitution and hence markups.[20]

In our results under the SCCTE, we will also use the economy-wide aggregate markup defined as:

$$\mu \equiv \frac{\int_{\theta_e} \mu\left(\theta_e\right) L_w\left(\theta_e\right) f_e\left(\theta_e\right) d\theta_e}{\int_{\theta_e} L_w\left(\theta_e\right) f_e\left(\theta_e\right) d\theta_e}. \tag{24}$$

It is the employment weighted (by $L_w(\theta_e)$) sum of the firm level markups.

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

$$\nu_{ij}(\theta_e) \equiv \frac{W L_{w,ij}(\theta_e)}{P_{ij}\left(\theta_e\right) Q_{ij}\left(\theta_e\right)\left(1-t_s\right)}. \tag{25}$$

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

$$\nu_{ij}(\theta_e) = \frac{\xi}{\mu_{ij}(\theta_e)}. \tag{26}$$

Although the firm-level labor share is exogenous, the aggregate labor share is endogenous. Denote the aggregate labor share by

$$\nu \equiv \frac{W \int x_w\left(\theta_w\right) l_w\left(\theta_w\right) f_w\left(\theta_w\right) d\theta_w}{Q}. \tag{27}$$

Then we summarize the results on the equilibrium labor share in the following Proposition 1:

Proposition 1 *(i) The firm labor share $\nu_{ij}(\theta_e)$ is independent of taxes and is decreasing in the markup $\mu_{ij}(\theta_e)$;*

(ii) In the Laissez-faire economy,[21] *the aggregate labor share ν is decreasing in market power (decrease in I) when*

$$\frac{1+\varepsilon_w}{\varepsilon_w} - \xi\left(\varepsilon_e+1\right) > 0, \tag{28}$$

[20] We can derive the equivalent inverse demand elasticity under Bertrand competition which is different from the residual demand elasticity under Cournot:

$$\varepsilon_{Q_{ij}}(\theta_e) = -\left[\left(1-s_{ij}\right)\eta\left(\theta_e\right) + s_{ij}\sigma\right]^{-1}. \tag{23}$$

In fact, all our results go through under Bertrand and are similar to Cournot once we adjust equation (21).

[21] See the following section 3.1 for details about the Laissez-faire economy.

and

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

Proof. See Appendix A.3. ■

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

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

Part two of Proposition 1 is strong in the sense that it is not dependent on the assumptions on $\eta(\theta_e)$. The two restrictions on the parameters are weak and are generally satisfied for the range of parameter values used in the quantitative literature.[22] In addition, the parameter restrictions have intuitive economic interpretations. Condition (28) guarantees that the equilibrium wage is increasing in TFP (e.g., see (A22)), while condition (29) ensures the labor demand is decreasing in W (e.g., see (A12)).

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

Proposition 2 *Let conditions (28) and (29) hold and let $\eta(\theta_e)$ be constant. When the number of firms I decreases in all markets, the markup $\mu_{ij}(\theta_e)$ increases in all markets. Then:*

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

[22] The literature typically uses parameters in the range $\eta \in [3, 10], \sigma \in (1, 4], \xi \in [0.7, 1]\varepsilon_w, \varepsilon_e \in [0.1, 0.5]$. See amongst others Atkeson and Burstein (2008), Hendel and Nevo (2006), Broda and Weinstein (2006), Lucas (1978), Chetty et al. (2011).

(ii) At the aggregate level, the wage rate W, the aggregate labor share ν, output Q, aggregate worker consumption C_w and aggregate worker utility V_w decline. The effects on aggregate entrepreneur profits Π and aggregate entrepreneur utility V_e are ambiguous.

Individual and aggregate entrepreneur profits increase if and only if

$$\mu_{ij}(\theta_e) \leq \frac{\xi}{\frac{\varepsilon_e}{1+\varepsilon_e} + \frac{\varepsilon_w}{\varepsilon_w+1}\xi}, \tag{30}$$

and individual and aggregate entrepreneur utility increase if and only if

$$\mu_{ij}(\theta_e) \leq \frac{\xi + \frac{\varepsilon_e}{\varepsilon_e+1}}{\frac{\varepsilon_e}{\varepsilon_e+1} + \frac{\varepsilon_w}{1+\varepsilon_w}\xi}. \tag{31}$$

Proof. See Appendix A.4. ■

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

The restrictions for increasing profits (30) and increasing utility (31) are satisfied for typical values used in quantitative studies. For example, with $\varepsilon_e = \varepsilon_w = 0.25$ and $\xi = 0.85$, the condition for increasing profits is satisfied for all firms with markup $\mu_{ij}(\theta_e) < 2.3$ and the second is condition for increasing utility is satisfied for $\mu_{ij}(\theta_e) < 2.8$.

3.2 The Planner's Problem

The planner's problem can be treated in a number of different ways. In the heuristic argument that follows, the planner adopts feasible direct truthful mechanisms $\{c_w(\theta_w), y(\theta_w)\}$ for workers and similarly adopts $\{c_e(\theta_e), \pi(\theta_e)\}$ for entrepreneurs to implement allocation rules that maximize social welfare under other information and resource constraints. Specifically, the planner asks each of the entrepreneurs and workers to report their types and assigns a reward contingent based on the announced type. A worker who reports θ'_w obtains $y(\theta'_w)$ in labor income, which results in $c_w(\theta'_w)$ of after-tax income. Similarly, an entrepreneur who reports θ'_e obtains $\pi(\theta'_e)$ in profit and $c_e(\theta'_e)$ in after-tax profit.

Incentive Compatibility of the Worker. Workers are atomistic and take the offered mechanisms as given. They report their types to maximize their gross utility $V_w(\theta_w)$:

$$V_w(\theta_w) \equiv \max_{\theta'_w \in \Theta_w} c_w(\theta'_w) - \phi_w\left(\frac{y(\theta'_w)}{x_w(\theta_w) W}\right). \tag{32}$$

Denote $V_w(\theta'_w|\theta_w) = c_w(\theta'_w) - \phi_w\left(\frac{y(\theta'_w)}{x_w(\theta_w)W}\right)$ as the utility of θ_w worker who reports θ'_w. Using envelope theory, we obtain

$$V'_w(\theta_w) = l_w(\theta_w)\phi'_w(l_w(\theta_w))\frac{x'_w(\theta_w)}{x_w(\theta_w)}. \tag{33}$$

Under our monotonicity assumption on $y(\theta_w)$, (33) is not only a necessary but also a sufficient condition to the worker's problem (see Mirrlees (1971)).

Incentive Compatibility of the Entrepreneur. Entrepreneurs report a type θ'_e to maximize their gross utility

$$V_e(\theta_e) = \max_{\theta' \in \Theta_e} V_e(\theta'_e|\theta_e), \tag{34}$$

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, which is given by

$$l_e(\theta'_e|\theta_e) = \min_{L_w, l_e} l_e$$
$$\text{s.t. } P(Q_{ij}, \theta_e) Q_{ij}(1 - t_s) - WL_w = \pi(\theta'_e).$$

The first-order necessary incentive condition requires the following: $\frac{\partial V_e(\theta'_e|\theta_e)}{\partial \theta'_e}|_{\theta'_e=\theta_e} = 0$. In Appendix B.1, we prove that

Lemma 1 *Under Assumption 1, we have the following results:*

(i) The first-order necessary incentive condition $\frac{\partial V_e(\theta'_e|\theta_e)}{\partial \theta'_e}|_{\theta'_e=\theta_e} = 0$ *is not only a necessary but also a sufficient condition to the entrepreneur's problem;*

(ii) Given $V_e(\theta) = c_e(\theta_e) - \phi_e(l_e(\theta_e))$ *and the inverse demand function in the SCCTE, the first-order necessary incentive condition is equivalent to*

$$V'_e(\theta_e) = \phi'_e(l_e(\theta_e)) l_e(\theta_e)\left[\mu(\theta_e)\frac{\partial \ln P(Q_{ij}(\theta_e), \theta_e)}{\partial \theta_e} + \frac{x'_e(\theta_e)}{x_e(\theta_e)}\right], \tag{35}$$

where

$$\frac{\partial \ln P(Q_{ij}(\theta_e), \theta_e)}{\partial \theta_e} = \frac{\chi'(\theta_e)}{\chi(\theta_e)} + \varepsilon_{Q_{-ij}}(\theta_e)\frac{Q'_{ij}(\theta_e)}{Q_{ij}(\theta_e)}, \tag{36}$$

and $\varepsilon_{Q_{-ij}}(\theta_e) = \left[\frac{1}{\eta(\theta_e)} - \frac{1}{\sigma}\right]\frac{I-1}{I} < 0$ *is the cross inverse-demand elasticity,* $\theta_e \in \Theta_e$.[23]

Proof. See Appendix B.1. ■

Lemma 1 is useful because it demonstrates that the incentive- compatible constraint of the entrepreneur boils down to condition (35), which has an intuitive economic explanation. Because of its critical importance, we name $\frac{x'_w(\theta_w)}{x_w(\theta_w)} = \frac{d \ln x_w(\theta_w)}{d\theta_w}$ as θ_w-type worker's productivity premium, which is the percentage

[23]The cross inverse-demand elasticity is the elasticity of an entrepreneur's inverse demand function with respect to her competitor's output. See Appendix A.6.2 for details about this elasticity.

change of the individual productivity with respect to skill θ_w. The worker's productivity premium together with the skill distribution determines the distribution of wage. More importantly, it determines $V_w'(\theta_w)$ given labor supply $l_w(\theta_w)$. Analogously, we call $\mu(\theta_e)\frac{\partial \ln P\left(Q_{ij}(\theta_e),\theta_e\right)}{\partial\theta_e}+\frac{x_e'(\theta_e)}{x_e(\theta_e)}$ the θ_e-type entrepreneur's productivity since it determines $V_e'(\theta_e)$ given entrepreneurial effort $l_e(\theta_e)$. Our incentive constraints thus highlight the interaction between technology and market power in determining productivity premiums and differences in gross utility. Now we explain the incentive condition in two different situations: (1) When $I=1$, the entrepreneurial productivity is $\mu(\theta_e)\frac{\chi'(\theta_e)}{\chi(\theta_e)}+\frac{x_e'(\theta_e)}{x_e(\theta_e)}$. It is increasing in market power $\mu(\theta_e)$ when $\frac{\chi'(\theta_e)}{\chi(\theta_e)}>0$. This is because, other things being equal, the disutility induced by obtaining one extra unit of profit is lower for entrepreneurs with higher production efficiency or higher pricing powers.

A key feature of the incentive condition is that the price component is multiplied by the markup, which suggests that $x_e(\theta_e)$ and $\chi(\theta_e)$ affect the gross utility in different ways. More specifically, given inputs, raising $\chi(\theta_e)$ increases the price directly. With the increase of price, the effort needed to finish a task becomes lower, and since the price goes up with the reduction of effort, there is a multiplier effect from raising $\chi(\theta_e)$, which is in terms of markup. On the other hand, increasing $x_e(\theta_e)$ raises output and lowers the price. Therefore, in order to obtain the same growth in profit, $x_e(\theta_e)$ should be increased at a higher rate compared to $\chi(\theta_e)$.

(2) When $I>1$, the sign of $\frac{\partial \ln P\left(Q_{ij}(\theta_e),\theta_e\right)}{\partial\theta_e}$ is ambiguous. In particular, if $\frac{\chi'(\theta_e)}{\chi(\theta_e)}=0$, $\frac{\partial \ln P\left(Q_{ij}(\theta_e),\theta_e\right)}{\partial\theta_e}$ is generally negative, because $Q_{ij}(\theta_e)$ is generally increasing in the skill of the entrepreneur. One may now think that when $\frac{\chi'(\theta_e)}{\chi(\theta_e)}=0$, rising markup loosens the incentive constraint instead of tightening it. This is not necessarily true, because $\varepsilon_{Q_{-ij}}(\theta_e)\frac{Q_{ij}'(\theta_e)}{Q_{ij}(\theta_e)}$ also changes when the markup increases. In Appendix A.2, we show that in the Laissez-faire economy with constant markup and satisfies condition (29), θ_e-type entrepreneur's productivity increases in markup when $\frac{d\ln X(\theta_e)}{d\theta_e}>0$, where

$$X(\theta_e)\equiv x_e(\theta_e)^{\frac{\sigma-1}{\sigma}}\chi(\theta_e)$$

is a *composite productivity* of the θ_e-type entrepreneurs. Moreover, we show that the entrepreneurial productivity and $\frac{d\ln V_e(\theta_e)}{d\theta_e}$ grows with the introduce of markup inequality, even if $\frac{\chi'(\theta_e)}{\chi(\theta_e)}=0$.

One interesting feature of the incentive condition is that it depends on $\frac{\partial \ln P\left(Q_{ij}(\theta_e),\theta_e\right)}{\partial\theta_e}$ instead of $\frac{d\ln P(\theta_e)}{d\theta_e}$. This is because entrepreneurs can change the price by changing their own output Q_{ij}. As a result, a tax reform has no first-order effect on the relative price through its effect on a firm's own output Q_{ij}. There are two interesting findings with the incentive condition:

(1) Taking $I=1$ as an illustration, one can see that the indirect redistribution route present in a competitive economy is closed in our economy since $\frac{\partial \ln P\left(Q_{ij}(\theta_e),\theta_e\right)}{\partial\theta_e}$ is exogenous. Intuitively, when tax policy is changed, entrepreneurs react to the tax reform by changing prices until any marginal change in price has no first-order effect on entrepreneurial gross utilities. Thus, tax has no first-order effect on the gross utility of entrepreneurs through its effect on the prices of products.

(2) When $I>1$, the indirect redistribution effect emerges. It is dependent on the strategic interaction between competitors in a submarket, which is shown by the last term on the right side of equation (36).

Specifically, an increase in the competitors' outputs decreases the price of goods in the submarket, which is caught in the incentive constraint by the cross inverse-demand elasticity $\varepsilon_{Q_{-ij}}(\theta_e)$. Without strategic interaction between competitors in a submarket, the inverse demand of the firm is not dependent on the outputs of other firms in the same submarket (because $\varepsilon_{Q_{-ij}}(\theta_e) = 0$), and the indirect redistribution route is closed.

Tax Wedges. In the second-best allocation, marginal distortions in agents' choices can be described with wedges. Entrepreneurs have three possible choices (consumption, working hours, and hiring workers), while workers have two possible choices (consumption and working hours). In total, there are three tax wedges: (i) the tax wedge $\tau_s(\theta_e)$ between the marginal cost and marginal income of labor inputs $L_w(\theta_e)$, (ii) the tax wedge $\tau_w(\theta_w)$ between the marginal disutility and income of the labor supply l_w, and (iii) the tax wedge $\tau_e(\theta_e)$ between the marginal disutility and income of the entrepreneur's labor supply l_e. Specifically, we shall define the three types of tax wedges as

$$\tau_s(\theta_e) = 1 - \frac{W}{\varpi(\theta_e)}, \quad \tau_w(\theta_w) = 1 - \frac{\phi_w'(l_w(\theta_w))}{W x_w(\theta_w)}, \quad \text{and} \quad \tau_e(\theta_e) = 1 - \frac{\phi_e'(l_e(\theta_e)) / [1 - \tau_s(\theta_e)]}{\frac{\partial\left[P\left(Q_{ij}(\theta_e),\theta_e\right)Q_{ij}(\theta_e)\right]}{\partial l_e(\theta_e)}}, \tag{37}$$

where

$$\varpi(\theta_e) \equiv \frac{\partial\left[P\left(Q_{ij}(\theta_e), \theta_e\right) Q_{ij}(\theta_e)\right]}{\partial L_w(\theta_e)} \tag{38}$$

and $\frac{\partial\left[P\left(Q_{ij}(\theta_e),\theta_e\right)Q_{ij}(\theta_e)\right]}{\partial l_e(\theta_e)}$ are the marginal revenue of labor inputs and entrepreneurial effort, respectively.

Due to the policy constraint, the government cannot levy firm-specific or non-linear sales tax, which means $\tau_s(\theta_e)$ is restricted to be uniform. Then these tax wedges can be implemented by the tax system previously introduced (i.e., $\mathcal{T}$). From the FOCs of the workers, the entrepreneurs, and the final good producer, we obtain $\tau_s = t_s$, $\tau_w(\theta_w) = 1 - [1 - T_w'(y(\theta_w))]$ and $\tau_e(\theta_e) = 1 - [1 - T_e'(\pi(\theta_e))]$.[24] Observe that the sales tax enforces a uniform tax on both labor factors. Thus, the effective tax rates on labor factors are captured by $1 - [1 - T_w'(y(\theta_w))](1 - t_s)$ and $1 - [1 - T_e'(\pi(\theta_e))](1 - t_s)$.

As is known from the optimal tax literature, generally there are multiple tax systems that can implement the second-best allocation (e.g., see Chari and Kehoe (1999); Golosov et al. (2003)). In our model, as long as $\tau_s(\theta_e)$ is restricted to be uniform and income taxes are free, there is no need to enforce a sales tax. Hence, in the following analysis, we will assume $t_s = 0$, where $\tau_w(\theta_w)$ and $\tau_e(\theta_e)$ are the effective tax rates on labor factors. In the model extension, we loosen the policy constraint and provide the optimal tax wedges including $\tau_s(\theta_e)$.

Implementability. In this subsection we show how the second-best allocation can be implemented by the tax system studied in this paper. In addition, we demonstrate that t_s is redundant.

[24] The FOCs imply $t_s = 1 - \frac{W}{\frac{P(\theta_e)}{\mu(\theta_e)}\frac{\partial Q_{ij}(\theta_e)}{\partial L_w(\theta_e)}}$, $T_w'(y(\theta_w)) = 1 - \frac{\phi_w'(l_w(\theta_w))}{W x_w(\theta_w)}$ and $T_e'(\pi(\theta_e)) = 1 - \frac{\phi_e'(l_e(\theta_e))}{\frac{P(\theta_e)}{\mu(\theta_e)}\frac{\partial Q_{ij}(\theta_e)}{\partial l_e(\theta_e)}(1-t_s)}$.

Lemma 2 *Suppose that the FOCs of the agents and the final good producer are both necessary and sufficient. Suppose* $t_s = 0$*. A symmetric Cournot competitive tax equilibrium* $\{\mathcal{A}, \mathcal{T}, \mathcal{P}\}$ *with* $t_s = 0$ *satisfies the following conditions jointly:*

(i) Incentive conditions (33) and (35) are satisfied;

(ii) Prices and wage satisfy (16) and (17);

(iii) Market clearing conditions (13) to (15) are satisfied.

Conversely, suppose $t_s = 0$ *and the allocation* $\mathcal{A}$ *and price* $\mathcal{P}$ *satisfy the properties in parts 1 to 3 above. Then there exists a tax system* $\mathcal{T}$ *with* $t_s = 0$ *such that the allocation* $\mathcal{A}$ *can be implemented at the prices* $\mathcal{P}$ *by the tax system* $\mathcal{T}$*.*

Proof. See Appendix B.2. ■

Lemma 2 establishes that if sales tax is restricted to be uniform, we can focus on a tax system where sales tax is zero. Under this tax system, $\tau_o(\theta_o)$ captures the effective tax rate on the labor factor.[25] Intuitively, the effective tax rates on the labor factors can be manipulated by the labor and profit tax rates. Thus, the sales tax is redundant when labor income and profit taxes are free.[26]

Reformulating the Planner's Problem. We can treat the planner's problem in a number of different ways. In the heuristic argument that follows, the planner adopts feasible direct, truthful mechanisms $\{\pi(\theta_e), c_e(\theta_e)\}_{\theta_e \in \Theta_e}$ and $\{y(\theta_w), c_w(\theta_w)\}_{\theta_w \in \Theta_w}$ to implement an allocation that maximizes the social welfare function under the feasibility conditions and information constraints.

This turns out to be easier if we take as the planner's control variables $V_o(\theta_o)$ instead of $c_o(\theta_o)$. To this end, the planner now chooses the variables $\{V_w(\theta_w), l_w(\theta_w), V_e(\theta_e), L_w(\theta_e), l_e(\theta_e), W\}_{\theta_o \in \Theta_o}$ to maximize the social welfare (6), subject to the incentive conditions (33) and (35); the feasibility conditions (13) to (15); and condition (17) with $t_s = 0$. Condition (17) can be treated as a policy constraint in the planner's problem. Since the planner cannot levy a firm-specific sales tax or differential tax on the labor inputs of firms, the marginal revenue of labor inputs must be equal for firms. Since W is controllable, the uniform-sales-tax policy constraint can be rewritten as $\frac{d \ln \varpi(\theta_e)}{d\theta_e} = 0$. Accordingly, we can save a variable (i.e., W) in the planner's problem.

In this reformulated planner's problem, we can now introduce some shorthand notation for the social welfare weights and the elasticities that appear in the solution to the planner's problem. We denote $g_o(\theta_o)$

[25]In our model, we allow profit and labor income tax to be different, which governs the wage rate, so that there is no need to use the sales tax to manipulate W to achieve indirect redistribution between the entrepreneur and worker. However, the sales tax is needed if income taxes are restricted to be uniform (e.g., see Scheuer (2014)).

[26]To see this, suppose that $\{T_w(y), T_e(\pi), t_s\}$ is the optimal tax that implements the second-best allocation and that there exists another optimal tax system $\{T_w^{\#}(y), T_e^{\#}(\pi), t_s^{\#}\}$ that can implement the second-best allocation with $t_s^{\#} = 0$. Then the tax system can be constructed such that $1 - T_o^{\#}(x) = [1 - T_o'(x)](1 - t_s), x \in \mathbb{R}_+$.

and $\bar{g}_o(\theta_o)$ as the marginal and weighted social welfare weights, respectively:

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

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

We denote $\varepsilon^{\varpi}_{L_w}(\theta_e)$, $\varepsilon^{\varpi}_{l_e}(\theta)$ and $\varepsilon_{Q_{-ij}}(\theta_e)$ as the own elasticities of wage with respect to labor inputs and effort, and cross inverse-demand elasticity, respectively. These elasticities under our technology are given by:[27]

$$\varepsilon^{\varpi}_{L_w}(\theta_e) = \xi\frac{\sigma-1}{\sigma} - 1,\ \varepsilon^{\varpi}_{l_e}(\theta_e) = \frac{\sigma-1}{\sigma} \text{ and } \varepsilon_{Q_{-ij}}(\theta_e) = -\frac{1}{\mu(\theta_e)} + \frac{\sigma-1}{\sigma}. \tag{39}$$

We define a linear elasticity of profit with respect to net-tax income rate as:

$$\varepsilon^{\pi}_{1-\tau_e}(\theta_e) \equiv \frac{1}{\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu(\theta_e)-\xi\right]-1}. \tag{40}$$

This elasticity captures θ_e firm's reaction to the net-tax income rate when it is taking others' actions as given and a linear profit tax is in place. Notably, the elasticity is decreasing in the markup. See Appendix A.6 for details about these elasticities.

4 Main Results

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

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

$$\frac{1}{1-\tau_w(\theta_w)} = \frac{1+\left[1-\bar{g}_w(\theta_w)\right]\frac{1+\varepsilon_w}{\varepsilon_w}\frac{1-F_w(\theta_w)}{f_w(\theta_w)}\frac{x'_w(\theta_w)}{x_w(\theta_w)}}{\mu}, \tag{41}$$

$$\frac{1}{1-\tau_e(\theta_e)} = \frac{\frac{1+\left[1-\bar{g}_e(\theta_e)\right]H(\theta_e)\frac{1}{\varepsilon^{\pi}_{1-\tau_e}(\theta_e)}}{\mu(\theta_e)} + \frac{\frac{\sigma}{\sigma-1}}{\frac{\sigma}{\sigma-1}-\xi}IRE(\theta_e)}{1-\frac{\xi}{\frac{\sigma}{\sigma-1}-\xi}RE(\theta_e)}. \tag{42}$$

The Reallocation Effect $RE(\theta_e)$, Indirect Redistribution Effect $IRE(\theta_e)$, and Virtual Hazard Ratio of Profit $H(\theta_e)$

[27] An alternative expression for $\varepsilon_{Q_{-ij}}(\theta_e)$ is $\varepsilon_{Q_{-ij}}(\theta_e) = \left[\frac{1}{\eta(\theta_e)} - \frac{1}{\sigma}\right]\frac{I-1}{I}$.

are defined as

$$RE\left(\theta_e\right) \equiv \frac{\mu}{\mu\left(\theta_e\right)} - 1, \tag{43}$$

$$IRE\left(\theta_e\right) \equiv \varepsilon_{Q_{-ij}}(\theta_e)\left\{\left[1 - g_e(\theta_e)\right] - \left[1 - \bar{g}_e(\theta_e)\right] H(\theta_e)\right\}, \tag{44}$$

$$H(\theta_e) \equiv \frac{1 - F_e(\theta_e)}{f_e\left(\theta_e\right)} \left\{ \frac{\frac{\sigma-1}{\sigma}\frac{1+\varepsilon_e}{\varepsilon_e}\frac{d}{d\theta_e}\left[\ln\frac{X(\theta_e)}{\mu(\theta_e)}\right]}{\frac{1+\varepsilon_e}{\varepsilon_e}\left(\frac{\sigma}{\sigma-1} - \xi\right) - 1} + \frac{d\ln\left[\mu\left(\theta_e\right) - \xi\right]}{d\theta_e} \right\}. \tag{45}$$

Last, the average markup satisfies

$$\mu = \int_{\theta_e} \mu\left(\theta_e\right)\omega\left(\theta_e\right)d\theta_e \tag{46}$$

where $\mu\left(\theta_e\right) = \frac{1}{1-\left[\frac{1}{\eta(\theta_e)}\frac{I-1}{I} + \frac{1}{\sigma}\frac{1}{I}\right]}$, *and*

$$\omega\left(\theta_e\right) = \frac{\left[\left[1-\tau_e\left(\theta_e\right)\right]\left(\frac{X(\theta_e)}{\mu(\theta_e)}\right)^{\frac{\varepsilon_e+1}{\varepsilon_e}\frac{\sigma}{\sigma-1}}\right]^{\frac{1}{\frac{1+\varepsilon_e}{\varepsilon_e}\left(\frac{\sigma}{\sigma-1}-\xi\right)-1}} f_e\left(\theta_e\right)}{\int_{\theta_e}\left[\left[1-\tau_e\left(\theta_e\right)\right]\left(\frac{X(\theta_e)}{\mu(\theta_e)}\right)^{\frac{\varepsilon_e+1}{\varepsilon_e}\frac{\sigma}{\sigma-1}}\right]^{\frac{1}{\frac{1+\varepsilon_e}{\varepsilon_e}\left(\frac{\sigma}{\sigma-1}-\xi\right)-1}} f_e\left(\theta_e\right)d\theta_e}. \tag{47}$$

Proof. See Appendix C.2. ■

The formulas (41), (42), and the average markup (46) determine the optimal tax rates. The term $RE\left(\theta_e\right)$ captures the reallocation effect of taxes. When the markup of firm θ_e is higher than the modified average markup μ, the $RE\left(\theta_e\right)$ of tax decreases $\tau_e\left(\theta_e\right)$, and it increases $\tau_e\left(\theta_e\right)$ otherwise. This is because the labor demand of a high-markup firm is inefficiently lower than that of a low-markup firm. Thus, interventions in the product market should reallocate labor factors to the high-markup firms.

The term $IRE\left(\theta_e\right)$ captures the indirect redistribution effect of the profit tax through prices. $IRE\left(\theta_e\right)$ contains two redistribution effects caused by a percentage change in $Q_{-ij}(\theta_e)$: a local redistribution effect captured by $\varepsilon_{Q_{-ij}}(\theta_e)\left[1 - g_e(\theta_e)\right]$ and a cumulative redistribution effect $\varepsilon_{Q_{-ij}}(\theta_e)\left[1 - \bar{g}_e(\theta_e)\right]\frac{1-F_\pi(\pi(\theta_e))}{f_\pi(\pi(\theta_e))\pi(\theta_e)}$. Intuitively, decreasing $\tau_e\left(\theta_e\right)$ increases the output of firms in θ_e submarket (i.e., $Q_{-ij}(\theta_e)$), which in turn decreases the price of products in θ_e submarket ($P\left(\theta_e\right)$). Meanwhile, a decrease of price $P\left(\theta_e\right)$ reduces the after-tax income of θ_e entrepreneur, which promotes equality and social welfare if and only if $g_e(\theta_e) < 1$. This decrease in price triggers an incentive-compatible redistribution between the government and all the entrepreneurs with skills higher than θ_e. Since $\bar{g}_e(\theta_e) \leq 1$, this cumulative indirect redistribution effect induced by the decrease of price always requires a lower $\tau_e\left(\theta_e\right)$.

Following Saez (2001)'s definition of virtual density of income, we define the *virtual hazard ratio of profit* $H(\theta_e)$. $H(\theta_e)$ is exogenous and equals $\frac{1-F_\pi(\pi(\theta_e))}{\pi(\theta_e)f_\pi(\pi(\theta_e))}$ when $\tau_e'\left(\theta_e\right) = 0$.[28] Thus, it equals the empirical profit hazard ratio if the profit tax in the real economy is linear. Under condition (29), we have $\frac{1+\varepsilon_e}{\varepsilon_e}\left(\frac{\sigma}{\sigma-1} - \xi\right) - 1 > 0$ and thus $H(\theta_e)$ increases in $\frac{d\ln X(\theta_e)}{d\theta_e}$. Last, $\frac{\xi}{\frac{\sigma}{\sigma-1}-\xi} = -\xi\frac{\varepsilon^{\omega}_{l_e}(\theta_e)}{\varepsilon^{\omega}_{L_w}(\theta_e)} > 0$ and $\frac{\frac{\sigma}{\sigma-1}}{\frac{\sigma}{\sigma-1}-\xi} = 1 - \xi\frac{\varepsilon^{\omega}_{l_e}(\theta_e)}{\varepsilon^{\omega}_{L_w}(\theta_e)} > 0$, where ξ is the elasticity of firm-level output with respect to the labor inputs.

[28] See the proof of Theorem 1. In particular, equation (C42).

Theorem 1 fully describes the optimal tax wedges, but the tax rate for the entrepreneurs $\tau_e(\theta_e)$ cannot be written explicitly because the weights $\omega(\theta_e)$ in equation (47) and hence the average markup μ is a function of the tax rate $\tau_e(\theta_e)$. In what follows, we can write the tax rate explicitly under a particular parameter configuration. Than we can solve explicitly for the weights $\omega(\theta_e)$ and therefore we can write the average markup μ explicitly in the following Corollary:

Corollary 1 *When* $\frac{1+\varepsilon_e}{\varepsilon_e}\left(\frac{\sigma}{\sigma-1}-\xi\right)=2$, *we have*

$$\mu=\frac{\left(\frac{\sigma}{\sigma-1}+\xi\right)\int_{\theta_e}\overline{\gamma}(\theta_e)\mu\left(\theta_e\right)d\theta_e}{2\xi}-\frac{\sqrt{\Delta}}{2\xi},$$

where

$$\begin{aligned}
\overline{\gamma}(\theta_e) &= \frac{\gamma(\theta_e)}{\int_{\theta_e}\gamma(\theta_e)d\theta_e},\\
\gamma(\theta_e) &= \frac{\left(\frac{X(\theta_e)}{\mu(\theta_e)}\right)^{\frac{\varepsilon_e+1}{\varepsilon_e}\frac{\sigma}{\sigma-1}}f_e\left(\theta_e\right)}{1+\left[1-\bar{g}_e(\theta_e)\right]H(\theta_e)\left[\frac{1}{\varepsilon_e}\mu\left(\theta_e\right)-\frac{1+\varepsilon_e}{\varepsilon_e}\xi+\frac{1}{\sigma-1}\right]+\left[\mu\left(\theta_e\right)-\frac{\sigma}{\sigma-1}\right]\left[1-g_e(\theta_e)\right]},\\
\Delta &= \left(\frac{\sigma}{\sigma-1}-\xi\right)^2\left[\int_{\theta_e}\overline{\gamma}(\theta_e)\mu\left(\theta_e\right)d\theta_e\right]^2-4\frac{\sigma}{\sigma-1}\xi\left[\int_{\theta_e}\mu\left(\theta_e\right)^2\overline{\gamma}(\theta_e)d\theta_e-\left(\int_{\theta_e}\overline{\gamma}(\theta_e)\mu\left(\theta_e\right)d\theta_e\right)^2\right].
\end{aligned}$$

Proof. See Appendix C.3. ■

Corollary 1 provides a special case where we can provide explicit optimal tax formulas with given social welfare weights. Note that $\frac{1+\varepsilon_e}{\varepsilon_e}\left(\frac{\sigma}{\sigma-1}-\xi\right)=2$ is consistent with condition (29). Corollary 1 thus suggests that under some reasonable parameter values, there is a unique and well-defined solution to the equation system (42) and (46).

(i) Homogeneous Agents

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

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

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

Proof. When worker and entrepreneur types are homogeneous, $\bar{g}_o=g_o=1$, and the optimal tax formulas (41) and (42) can be simplified to (48). ■

A few aspects of this finding deserve mention. First, note that since all entrepreneurs are identical, the firm-level markup $\mu(\theta_e)$ is equal to the average weighted markup μ, and the optimal tax wedge is independent of the firm type. Second, this result holds irrespective of the number of competitors I in each market and includes cases of monopolistic competition and oligopolistic competition. Third, not only the profit tax rate but also the labor income tax rate are *negative* and decrease in the markup.

The interpretation of the optimal tax formula is straightforward. When agents are identical, the incentive constraints are muted because they are trivially satisfied. As a result, the optimal tax wedge exactly offsets the distortion due to the markup. This may seem surprising, but it affects both workers and entrepreneurs equally because the planner can only impact output by affecting their incentives to produce and provide effort.[29]

The tax rate is negative in order to incentivize workers and entrepreneurs to supply labor in order to offset the distortion from market power. Because there is no heterogeneity within groups (workers and entrepreneurs), the only role the planner bestows on the tax wedges is to correct the externality or markup distortion. As a result, we can think of the tax wedge here playing the role of a Pigouvian tax. In fact, when the output market is competitive and markups are equal to one, the marginal tax rate is zero, and there is no role for efficiency-enhancing taxes. In that case, the economy is Pareto efficient.

Even though the tax wedges are the same, that does not mean workers and entrepreneurs will face the same tax burdens. non-linear tax system facilitates a transfer between entrepreneurs and workers. The tax burden of each occupation thus depends on the social welfare function. Under a utilitarian social welfare function, the burden is indeterminate, whereas it is determinate under a concave social welfare function. Here market power and the level of the markup μ play a key role in determining the tax burden. As we have seen from Proposition 2, a rise in market power accompanied by an increase in μ leads to a redistribution of income from workers to firms (mainly through a lower wage rate W) as well as a decrease in welfare. Therefore, with non-linear welfare weights, the tax burden will change as markups change.

(ii) Monopolistic Competition and Uniform Markups

We now turn to an economy without strategic interaction within each market j, that is, with a monopolistic producer in each market where $I = 1$. In addition, we assume that markups are uniform; that is, the residual demand elasticity is constant, $\eta(\theta_e) = \eta$. Entrepreneurs are heterogeneous in productivity θ_e, but their markets all face the same demand. Under monopolistic competition and uniform markups, the solution to the planner's problem yields the following optimal taxation policy:

Proposition 4 *When $I = 1$, optimal labor income tax satisfies (41) and optimal profit tax can be simplifies as:*

$$\frac{1}{1-\tau_e(\theta_e)} = \frac{1 + [1-\bar{g}_e(\theta_e)]\,\frac{1+\varepsilon_e}{\varepsilon_e}\,\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\left[\mu\frac{\chi'(\theta_e)}{\chi(\theta_e)} + \frac{x_e'(\theta_e)}{x_e(\theta_e)}\right]}{\mu}, \forall \theta_e \in \Theta_e. \tag{49}$$

Proof. When $I = 1$, $\mu(\theta_e) = \mu$ and $\varepsilon_{Q_{-ij}}(\theta_e) = 0$. Then (C35) can be reduced to (49). Alternatively, we can derive the above equation by reducing (42). ■

[29] The planner would also be able to achieve this outcome with a sales tax wedge τ_s, but as we have shown above, the outcome of a sales tax can always be mimicked with appropriate income taxes. Because at the margin the contribution of effort from the workers and entrepreneurs to the output is the same, the income tax wedges τ_w and τ_e are identical.

Comparing the optimal tax rates under heterogeneous and homogeneous agents establishes that the heterogeneity between agents calls for a higher tax wedge. The numerators on the right sides of the optimal tax formulas immediately stem from the incentive constraint for both the worker and the entrepreneur. The numerators are larger than one, and thus, the marginal tax is lower than in the case of homogeneous agents and identical markups. When the government has a preference for equality, it will raise the tax rate to generate tax revenue from high-income individuals and transfer the revenue to low-income individuals. However, the government should also take tax's distortion on effort into consideration. In the end the trade-off between deadweight loss and the redistribution of benefits is captured by the numerator.

For both workers and entrepreneurs, the extent to which the marginal tax rate is higher depends on the Pareto weight $\bar{g}_o(\theta_o)$, the elasticity of labor supply ε_o, and the hazard ratio of the skills $\frac{1-F_o(\theta_o)}{f_o(\theta_o)}\frac{x_o'(\theta_o)}{x_o(\theta_o)}$, which captures the trade-off between efficiency and equality. As the product of productivity ($x_o\left(\theta_o\right)$) and the population ($\frac{f_o(\theta_o)}{x_o'(\theta_o)}$) of θ_o-type agents increase, the tax wedge decreases in order to reduce distortion. As the population of agents with skills higher than $x_o\left(\theta_o\right)$ (i.e., $1-F_o(\theta_o)$) increases, the tax wedge increases to enhance the redistribution.

While there is a lot of similarity in the expression of the tax wedge for the workers and the entrepreneurs, market power induces one marked difference between the two. The tax wedge between the worker and the entrepreneur differs through the term $\mu\frac{\partial \ln P(Q(\theta_e),\theta_e)}{\partial\theta_e}$, which is equal to $\mu\frac{\chi'(\theta_e)}{\chi(\theta_e)}$ when $I=1$. This term is only present for entrepreneurs, not workers. It captures the two key forces. First, it depends on the technology and how different sectors differ in productivity $\chi(\theta_e)$, and second, it is influenced by how big the markup is. This difference in the numerator on the right side of (49) goes to the heart of one of the main findings of this paper. It highlights the interaction between market power and technology and how this interaction affects optimal taxation.

The two components ($\chi\left(\theta_e\right)$ and $x_e\left(\theta_e\right)$) that play a crucial role in pinning down the optimal tax formulas are both determinants of the productivity of a firm. In Appendix A.5, we show that the equilibrium labor supply and sales income only depend on the composite productivity term $X(\theta_e)$. However, $\chi\left(\theta_e\right)$ and $x_e\left(\theta_e\right)$ are not perfect substitutes in the sense that the equilibrium prices depend on the specific values of $\chi\left(\theta_e\right)$ and $x_e\left(\theta_e\right)$. This comes from the fact that $\chi\left(\theta_e\right)$ directly enters the demand function and thus interacts with the markup while $x_e\left(\theta_e\right)$ does not.

(iii) Oligopolistic Competition with Uniform Markups

We now consider cases with $I>1$ but still restrict the markup to be uniform; that is, $\eta\left(\theta_e\right)$ is constant. This setting introduces interfirm strategic action but still abstracts from the effect of markup inequality between firms. A planner who intends to take advantage of the general equilibrium price effect and ease the incentive constraint would like to decrease the relative price of goods produced by high-skilled entrepreneurs. However, whether the planner should encourage the factor inputs of high-skilled entrepreneurs remains ambiguous because there are two opposing forces. On the one hand, raising the labor inputs of competitors in the same submarket reduces the relative price of goods in the submarket; on the other hand, raising the labor inputs increases entrepreneurial effort's marginal productivity.

Proposition 5 *Let $\eta(\theta) = \eta$ be constant then we have the following result (i). In addition, let the social welfare weights be exogenous, then we have the following result (ii) and (iii):*

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

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

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

$$g_e(\theta_e) < \frac{\xi(\sigma-1)}{\sigma} \left\{ 1 + [1 - \bar{g}_e(\theta_e)] H(\theta_e) \left[\frac{\varepsilon_e + 1}{\varepsilon_e} \left(\frac{\sigma}{\sigma - 1} - \xi \right) - 1 \right] \right\}; \tag{51}$$

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

$$g_e(\theta_e) < 1 + [1 - \bar{g}_e(\theta_e)] H(\theta_e) \left[\frac{\varepsilon_e + 1}{\varepsilon_e} \left(\frac{\sigma}{\sigma - 1} - \xi \right) - 1 \right]; \tag{52}$$

(iii) When $g_e(\theta_e) = \bar{g}_e(\theta_e)$ and $H(\theta_e) > 0$, $\tau_e(\theta_e)$ increases in μ if $g_e(\theta_e) \leq \frac{\xi(\sigma-1)}{\sigma}$; $\frac{1-\tau_w(\theta_w)}{1-\tau_e(\theta_e)}$ increases in μ if $g_e(\theta_e) < 1$.

Proof. See Appendix C.4. ■

Part one of Proposition 5 provides an explicit optimal profit tax formula when all firms have identical markups and the social welfare weights are exogenous. Compared to the tax formula under monopolistic competition (49), there is now an additional term, $IRE(\theta_e) \frac{\frac{\sigma}{\sigma-1}}{\frac{\sigma}{\sigma-1}-\xi}$, that captures the indirect redistribution effect of profit tax.

The term $\frac{\frac{\sigma}{\sigma-1}}{\frac{\sigma}{\sigma-1}-\xi} = 1 - \xi \frac{\varepsilon^{\omega}_{l_e}(\theta_e)}{\varepsilon^{\omega}_{L_w}(\theta_e)}$ is the percentage change of $Q_{ij}(\theta_e)$ with one percentage increase of $l_e(\theta_e)$. To see this, note that one percentage increase of $l_e(\theta_e)$ induces one percentage increase of $Q_{ij}(\theta_e)$ directly. In addition, the labor demand $L_w(\theta_e)$ will increase by $-\xi \frac{\varepsilon^{\omega}_{l_e}(\theta_e)}{\varepsilon^{\omega}_{L_w}(\theta_e)} = \frac{\xi}{\frac{\sigma}{\sigma-1}-\xi}$ percent as a result of a one percent increase in $l_e(\theta_e)$. This change of $L_w(\theta_e)$ ensures that the marginal productivity of L_w is uniform between firms. This crowding in effect of $l_e(\theta_e)$ on $L_w(\theta_e)$ induces a $\frac{\xi}{\frac{\sigma}{\sigma-1}-\xi}$ percentage increase in $Q_{ij}(\theta_e)$. In sum, under general equilibrium, a one percentage increase of $l_e(\theta_e)$ triggers a $1 + \frac{\xi}{\frac{\sigma}{\sigma-1}-\xi} > 1$ percentage increase of $Q_{ij}(\theta_e)$. On the other hand, as explained before, $IRE(\theta_e)$ is a marginal redistribution effect of $Q_{ij}(\theta_e)$. Thus, $\frac{\frac{\sigma}{\sigma-1}}{\frac{\sigma}{\sigma-1}-\xi} IRE(\theta_e)$ is the indirect redistribution effect induced by one percentage increase of $l_e(\theta_e)$.

Our optimal tax formula suggests that the optimal profit tax rate increases in the $IRE(\theta_e)$. Notice that $\frac{\frac{\sigma}{\sigma-1}}{\frac{\sigma}{\sigma-1}-\xi} \varepsilon_{Q_{-ij}}(\theta_e) \leq 0$;[30] $g_e(\theta_e)$ and $\bar{g}_e(\theta_e)$ approach zero, and thus $IRE(\theta_e)$ approaches $\varepsilon_{Q_{-ij}}(\theta_e) H(\theta_e)$, when the planner has a preference for equality and θ_e is large enough. Introducing the indirect redistribution effect generally requires a lower profit tax rate for high-skilled entrepreneurs if $H(\theta_e) < 1$.[31] Intuitively,

[30] Equation (22) implies $\mu \leq \frac{\sigma}{\sigma-1}$.

[31] The hazard ratios of top income in the United States is around 0.5 in 1992 and 1993 (Saez (2001)) and around $\frac{2}{3}$ in 2005 (Diamond and Saez (2011)). In 2007, the hazard ratio of top labor, capital and total incomes in the United States are around 0.62, 0.76 and 0.71, respectively (e.g., see Saez and Stantcheva (2018)).

decreasing the profit tax rate on high-skilled entrepreneurs can enhance the output of intermediate goods in the submarket, which in turn reduces the price of intermediate goods and improves the predistribution.[32]

While IRE requires a lower profit tax rate on high-skilled entrepreneurs, it also suggests that the optimal top profit tax rate may increase with the rise of markups, because the indirect redistribution effect become weaker with the rise of market power (i.e., $\left|\varepsilon_{Q_{-ij}}(\theta_e)\right|$ become smaller with the rise of $\mu(\theta_e)$). Part two of Proposition 5 provides a general criteria for when should the profit tax rate increases with the markup. It suggests that (a) given $H(\theta_e)$, when $g_e(\theta_e)$ (the social welfare weight for θ_e-type entrepreneur) is low enough as relative to $\bar{g}_e(\theta_e)$ (the average social welfare weight for entrepreneurs with types higher than θ_e), $\tau_e(\theta_e)$ should increase with μ; and (b) when $\frac{\varepsilon_e+1}{\varepsilon_e}\left(\frac{\sigma}{\sigma-1}-\xi\right) > 1$ (which is true under condition (29)), the larger the $H(\theta_e)$, the larger the possibility that $\tau_e(\theta_e)$ increase with μ. This is because $H(\theta_e)$, as an approximation of hazard ratio of profit,[33] is a statistic reflecting the trade-off between redistribution and efficiency. The larger the $H(\theta_e)$, the larger the redistribution benefit of marginal tax as relative to the efficiency cost of marginal tax.[34] To see this, consider the following experiments: when $H(\theta_e) \to 0$, condition (51) is equivalent to $g_e(\theta_e) < \frac{\xi(\sigma-1)}{\sigma}$; when $H(\theta_e) \to 1$ and $g_e(\theta_e) = \bar{g}_e(\theta_e)$, condition (51) is equivalent to $g_e(\theta_e) < \frac{\frac{\varepsilon_e+1}{\varepsilon_e}\xi}{1+\frac{\varepsilon_e+1}{\varepsilon_e}\xi}$. In both cases, $\tau_e(\theta_e)$ increases in μ when $g_e(\theta_e)$ is low enough or equivalently θ_e is large enough (when the planner has a preference for equality).

Part two of Proposition 5 also gives a condition under which $\frac{1-\tau_w(\theta_w)}{1-\tau_e(\theta_e)}$ increases in μ. It can be seen that such a condition is weaker than the condition for $\tau_e(\theta_e)$ to be increasing in μ. Moreover, since the right sides of condition (51) and (52) are positive there generally exists θ_e satisfies the condition. Interestingly, under condition (29), condition (51) as well as condition (52) are satisfied only if $g_e(\theta_e) < 1$.

Part three of Proposition 5 suggests that under reasonable $H(\theta_e)$, the top profit tax rate will increase with the rise of market power when the social welfare weights $g_e(\theta_e)$ for the top is constant (such that for the top types $g_e(\theta_e) = \bar{g}_e(\theta_e)$) and low enough. Moreover, we find that $\frac{1-\tau_w(\theta_w)}{1-\tau_e(\theta_e)}$ increases in μ if $g_e(\theta_e) < 1$, where 1 is the shadow price of government revenue.

Note that in the above analysis, we just assume that the social welfare weights are exogenous. Doing so, we get abstract from the influence of market power on changing social welfare weights and thus on the optimal taxation. This allows us to highlight the incentive elements. Our analysis of the Laissez-faire economy suggests that the gross utility of entrepreneurs generally increases with market power. If this is also the case under optimal taxation the marginal social welfare weights for entrepreneurs will decrease which in turn increases the optimal profit tax rates.[35] This finding has been emphasized by previous studies

[32] In Theorem 2, we illustrate that the indirect redistribution effect can be split into the local and cumulative indirect redistribution effects. Actually, the local effect generally dominates the cumulative effect (e.g., see the numerical analysis in Cui et al. (2021)). Then one can see that the sign of $IRE(\theta_e)$ is mainly determined by $\varepsilon_{Q_{-ij}}(\theta_e)\left[1-g_e(\theta_e)\right]$, which is positive when $g_e(\theta_e) > 1$.

[33] $H(\theta_e)$ equals the hazard ratio of profit at $\pi(\theta_e)$ when $\frac{d\tau_e(\theta_e)}{d\theta_e} = 0$.

[34] To see this, consider a marginal rise of $\tau_e(\theta_e)$ (see Saez (2001) for a more rigorous analysis). The larger the numerator of the hazard ratio ($\frac{1-F_\pi(\pi(\theta_e))}{\pi(\theta_e)f_\pi(\theta_e)}$), the larger the population of entrepreneurs with profit larger than $\pi(\theta_e)$, and thus the more the additional tax collected from individuals with profit larger than $\pi(\theta_e)$. On the other hand, due the suppression on labor supply of θ_e entrepreneurs, the tax revenue is reduced, and the larger the $\pi(\theta_e) f_\pi(\theta_e)$, the more the reduced tax revenue.

[35] Note that since the government can make a lump-sum transfer between the entrepreneurs and workers, the gross utility of an entrepreneur increases only if the benefit from rising market power exceeds the rise of taxes including the lump-sum part. One possible scenario is that the gross utility of some enterprises increase, while others decrease, because the lump-sum transfer from the entrepreneurs to the workers probably increases with the market power.

(see e.g., Kushnir and Zubrickas (2019)). A reason for segregating the effect of endogeneity of the social welfare weight is that a generalized social welfare weight may depend on factors other than the gross utility and those factors may also change with the market structure (see Saez and Stantcheva (2016) for generalized social welfare weight). For example, a generalized social welfare weight may depend on the revenue contribution of an entrepreneur relative to that of a worker (see Scheuer (2014)). More generally, it depends on the gap between the social and private value of being an entrepreneur. By segregating the endogeneity of social welfare weight, the criterion becomes more applicable. Besides, we demonstrate that the optimal profit tax rate may increase with the markup even if the social welfare weights are exogenous.

Market Structure, Indirect Redistribution, and Optimal Tax. One interesting finding of this paper is that the market structure is crucial to the indirect redistribution effect (or supply-side effect) of taxation. Many previous studies on endogenous prices and optimal taxation, taxes have a first-order effect on relative prices and thus can be used to ease the incentive constraints and improve income distribution (e.g., see Naito (1999); Stiglitz (2018); Sachs et al. (2020); Cui et al. (2021)). Specifically, when the marginal productivity of the labor factor (wage) decreases with labor inputs, the planner can compress the wage distribution by reducing the marginal tax rate of high-skilled agents and enhance the high-skilled agents' labor supply. Saez (2004) argued that tax's indirect redistribution effect collapse when agents make endogenous human capital investments. Under that case, the agents determine their wages, and the tax's effect on prices becomes second order. As a response to Saez (2004), Naito (2004) defended by showing that when agents have heterogeneous comparative advantages in accumulating different human capital, relative wages are still endogenous to taxes. Our finding contributes to the debate by demonstrating that the indirect redistribution of tax depends on the market structure. If the agents monopolize the market of the goods they provide, the indirect redistribution of taxes does fail. As long as the agent can not determine the price alone, taxes play a role in indirect redistribution. In the end, the strength of indirect redistribution depends on the agent's market power on the factor or goods provided.

(iv) Oligopolistic Competition with Heterogeneous Markups

Finally, we get to the full-blown tax formulas with both oligopolistic competition and heterogeneous markups from Theorem 1. Now the planner faces firms with heterogeneous markups and hence can use taxes to implement an efficiency-enhancing reallocation of factors.

For the workers, the tax formula (41) remains unchanged compared to the case with uniform markups. The introduction of heterogenous markups introduces the last change in the tax formula for the entrepreneurs (42), which is captured by the denominator on the right side of (42):

$$1 - RE\left(\theta_e\right) \frac{\xi}{\frac{\sigma}{\sigma-1} - \xi} = 1 + \frac{1}{\mu\left(\theta_e\right)} \frac{\xi\left[\mu\left(\theta_e\right) - \mu\right]}{\frac{\sigma}{\sigma-1} - \xi}.$$

Notice that

$$\mu\left(\theta_e\right)W = P\left(\theta_e\right)\frac{\partial Q_{ij}\left(\theta_e\right)}{\partial L_w\left(\theta_e\right)} \quad \text{and} \quad \mu W = \int_{\theta_e} P\left(\theta_e'\right)\frac{\partial Q_{ij}\left(\theta_e'\right)}{\partial L_w\left(\theta_e'\right)}\frac{L_w\left(\theta_e'\right)f_e\left(\theta_e'\right)}{\int L_w\left(\theta_e\right)f_e\left(\theta_e\right)d\theta_e}d\theta_e'.$$

One can see that $\left[\mu\left(\theta_e\right)-\mu\right]W$ is the increase in the total output of transferring $\frac{L_w(\theta_e')}{\int L_w(\theta_e)f_e(\theta_e)d\theta_e}$ units of labor factors from a type θ_e' firm to a type θ_e firm. As a result, the labor input in each type of firm is decreased by $\frac{1}{\int L_w(\theta_e)f_e(\theta_e)d\theta_e}$ percent, and the marginal productivity of the labor inputs of different firms are still uniform. On the other hand, $\frac{\xi}{\frac{\sigma}{\sigma-1}-\xi} = -\xi\frac{\varepsilon^{\varpi}_{l_e}(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)}$ is the percentage increase of labor demand $L_w(\theta_e)$ that ensures the marginal productivity of labor inputs are uniform between firms when $l_e(\theta_e)$ is increased by one percent.

In conclusion, $\frac{\xi[\mu(\theta_e)-\mu]}{\frac{\sigma}{\sigma-1}-\xi}$ captures the aggregate output that increases with a one percent increase of $l_e(\theta_e)$ and the resulting interfirm reallocation of workers' labor inputs. Our optimal tax formula thus suggests that the reallocation effect requires a lower (higher) tax rate on firms with a markup higher (lower) than the average markup because the labor inputs to these firms are relatively inefficiently low.[36] The above findings provide a novel explanation (i.e., markup inequality) for why profit tax in the real economy is less progressive than labor income tax (e.g., see Scheuer (2014)).

Technology, Monopoly Power, and Optimal Tax. It is of our interest to see how technology affects the optimal tax rate. As an illustration, we present the top tax rate. Top tax rate is crucial because top earners account for the vast majority of income. Moreover, current changes in technology are biased toward top-income individuals; thus, it's important to see how technologies affect the top tax rates. Specifically, we will focus on the impact of $\mu\left(\cdot\right)$ and ξ, the concavity of firm-level production technology with respect to the labor inputs, e.g., see equation (1). Note that $\mu\left(\cdot\right)$ is determined by the market structure as well as the technology. Also note that ξ is a mirror of the superstar effect considered by Scheuer and Werning (2017). We can establish Corollary 2:

Corollary 2 *Assume that there exists* $\theta_e^* \in \left(\underline{\theta}_e,\overline{\theta}_e\right)$ *such that* $\mu\left(\theta_e\right) = \widehat{\mu}$ *and* $\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{d\ln X(\theta_e)}{d\theta_e}$ *are constant in* $\left(\theta_e^*,\overline{\theta}_e\right)$.[37] *Then the virtual hazard ratio* $H\left(\theta_e\right) = \widehat{H}$ *is also constant in* $\theta_e^* \in \left(\underline{\theta}_e,\overline{\theta}_e\right)$. *Assume that there exists* $\theta_e^\star \in \left(\underline{\theta}_e,\overline{\theta}_e\right)$ *such that for any* $\theta_e \geq \theta_e^\star$, $g_e(\theta_e) = \widehat{g}_e < 1$ *is constant. We call entrepreneurs with a skill* $\theta_e \geq \max\left\{\theta_e^*,\theta_e^\star\right\}$ *the top entrepreneurs, and profit tax rates faced by the top entrepreneurs (i.e.,* $\tau_e\left(\theta_e\right)$ *for* $\theta_e \geq \max\left\{\theta_e^*,\theta_e^\star\right\}$*) the top profit tax rates. Then:*

[36] Since there is no use to set a marginal tax rate larger than one, the right side of (42) is positive. Supposing $\tau_e\left(\theta_e\right) < 1$, the numerator of the right side of (42) is positive if the denominator $1-\frac{\xi}{\frac{\sigma}{\sigma-1}-\xi}RE\left(\theta_e\right) > 0$, which is true because $\mu < \frac{\sigma}{\sigma-1}$.

[37] When the firm-level markup is constant, $H(\theta_e) = \frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{d\ln X(\theta_e)}{d\theta_e}\frac{\frac{1+\varepsilon_e}{\varepsilon_e}}{\frac{1+\varepsilon_e}{\varepsilon_e}\left(1-\xi\frac{\sigma-1}{\sigma}\right)-\frac{\sigma-1}{\sigma}}$ (e.g., see (45)). Therefore, given that $\mu\left(\theta_e\right) = \widehat{\mu}$ is constant on $\left(\theta_e^*,\overline{\theta}_e\right)$, $\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{d\ln X(\theta_e)}{d\theta_e}$ is constant on $\left(\theta_e^*,\overline{\theta}_e\right)$ if and only if the virtual profit hazard ratio $H(\theta_e)$ is constant on $\left(\theta_e^*,\overline{\theta}_e\right)$. This finding suggests that the assumption that there exists $\theta_e^* \in \left(\underline{\theta}_e,\overline{\theta}_e\right)$ such that $\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{d\ln X(\theta_e)}{d\theta_e}$ is constant on $\left(\theta_e^*,\overline{\theta}_e\right)$ is not a strong assumption if the hazard ratio of profit is constant in the real economy.

(i) The top profit tax rate is constant. Denote it as $\widehat{\tau}_e$. We have:

$$\frac{1}{1-\widehat{\tau}_e} = \frac{\frac{1}{\widehat{\mu}}\left[1+(1-\widehat{g}_e)\,\widehat{H}\frac{1}{\widehat{\varepsilon}^{\pi}_{1-\tau_e}}\right] + \frac{1-\frac{\sigma}{\sigma-1}\frac{1}{\widehat{\mu}}}{\frac{\sigma}{\sigma-1}-\xi}\,(1-\widehat{g}_e)\left(1-\widehat{H}\right)}{1-\frac{\xi}{\frac{\sigma}{\sigma-1}-\xi}\frac{\mu-\widehat{\mu}}{\widehat{\mu}}}, \tag{53}$$

where $\widehat{\varepsilon}^{\pi}_{1-\tau_e} \equiv \frac{1}{\frac{1+\varepsilon_e}{\varepsilon_e}(\widehat{\mu}-\xi)-1}$ is the elasticity of profit and hazard ratio for the top entrepreneurs;

(ii) Supposing $\widehat{g}_e$ is exogenous, one sufficient condition for $\widehat{\tau}_e$ to be increasing with the decrease of I is:

$$\widehat{g}_e < 1 - \frac{1}{2\cdot\widehat{H}}. \tag{54}$$

Proof. See Appendix C.5. ■

Note that if the top profit tax in the real economy is linear, $\widehat{H}$ equals the empirical top profit hazard ratio.[38] Corollary 2 is powerful in the sense that it suggests that under reasonable assumptions, one can use original observable statistics including $\widehat{H}$ to derive the optimal top profit tax rate and judge whether to increase the top profit tax. It is worth noting that $\widehat{H}$ won't change with the tax policies. However, $\widehat{H}$ changes with the markup. As a special case, when $I = 1$ and markup is uniform, for any $\theta_e \geq \max\{\theta_e^*, \theta_e^\star\}$:

$$\frac{1}{1-\widehat{\tau}_e} = \frac{1+(1-\widehat{g}_e)\,\widehat{H}\frac{1}{\widehat{\varepsilon}^{\pi}_{1-\tau_e}}}{\mu}. \tag{55}$$

Formula (55) generalizes the familiar top tax rate result of Saez (2001) (in which $\xi = 0$ and $\mu = 1$) to a CES production function under a monopoly competitive economy. Comparing the above result to Corollary 5 of Sachs et al. (2020), we see how market structure and technology affect the top tax rate given the hazard ratio of profit. While this statistics-based optimal tax formulas facilitate tax design (they are robust tax formulas in the sense of their independence from technology), one should note that profit distribution is endogenous to the markup, and when analyzing how profit tax changes with the technology and market structure, we see that both the effects of markups on the elasticity of profit and on profit distribution should be taken into consideration.

Combining Corollary 2 and Proposition 4 delivers interesting insights in the light of the findings by Scheuer and Werning (2017). When we look into the elasticity of profit and hazard ratio, one can see that an increase in ξ (a superstar effect) changes the elasticity of profit (40) as well as the hazard ratio. Moreover, when $I = 1$ and markup is uniform, these two effects will cancel each other out such that the relative net-tax income rate is unchanged with ξ. However, this is not true under general cases, where ξ affects the top tax rate through the indirect redistribution effect and reallocation effect. Rising ξ enlarges the influence of reallocation effect (because $\frac{\xi}{\frac{\sigma}{\sigma-1}-\xi}$ increases with ξ) while it effect on indirect redistribution effect is more ambiguous.

Proposition 4 has demonstrated how optimal profit tax changes with market power is dependent on the

[38] See the proof of Theorem 1. In particular, equation (C42).

technology. In addition, Proposition 5 provides useful criteria which help to judge the change of optimal profit tax. There we demonstrate that without markup inequality, the optimal profit tax rate for the top entrepreneurs generally increases with the rising markup. Comparing to (50) in Proposition 5, (53) suggests how may introducing markup inequality affects optimal profit taxation. A key question here is, considering the reallocation effect, should the government increase the top profit tax rate with the rise of market power? Condition (53) suggests that whether the optimal top profit tax rate increases crucially depends on the value of $\widehat{H}$. Specifically, if $\widehat{g}_e \to 0$, condition (54) is equivalent to $\widehat{H} > \frac{1}{2}$. (54) is a sufficient but not necessary condition for $\widehat{\tau}_e$ to be increasing with the marginal in I.[39] At this stage, it is important to remind that the hazard ratios of top income in the United States is around 0.5 in 1992 and 1993 (Saez (2001)) and around $\frac{2}{3}$ in 2005 (Diamond and Saez (2011)). In 2007, the hazard ratio of top labor, capital and total incomes in the United States are around 0.62, 0.76 and 0.71, respectively (e.g., see Saez and Stantcheva (2018)).

In conclusion, our optimal tax formulas deliver three stark findings regarding the profit tax: (i) while optimal profit tax rate may decrease with the rise of markup, it generally increases relative to the labor income tax rate; (ii) while the optimal profit tax rate in an economy with market power may be lower than that in a competitive economy, it may at the same time increase with the rise of markup; (iii) whether the optimal profit tax rate increases with the market power depend heavily on the relative value of social welfare weight to the virtual hazard ratio. We believe these are novel insights in the optimal taxation literature in the presence of market power.

5 Discussion and Robustness

We consider three alternative specifications of our benchmark model and show that our main findings are still robust in these settings.[40]

5.1 Non-linear Sales Taxes

In our benchmark model, we consider an environment with uniform linear sales tax, which restricts $\tau_s(\theta_e)$ to be constant. In this section we remove this policy constraint and allow for non-linear sales tax as considered by Ales et al. (2017). To do this, we allow the planner to contract with entrepreneurs on sales income $S(\theta_e) \equiv P(\theta_e) Q_{ij}(\theta_e)$ in addition to $\pi(\theta_e)$. An entrepreneur reporting θ_e' should obtain $\pi(\theta_e')$ in profit, $S(\theta_e')$ in sales income, and receive $c_e(\theta_e')$ in after-tax profit. Thus, a θ_e-type entrepreneur reporting θ_e' should choose L_w and l_e to satisfy the following two promise-keeping constraints:

$$P\left(Q_{ij}, \theta_e\right) Q_{ij} = S\left(\theta_e'\right) \quad \text{and} \quad P\left(Q_{ij}, \theta_e\right) Q_{ij} - W L_w = \pi\left(\theta_e'\right).$$

The two promise-keeping constraints determine the combination of L_w and l_e that are needed to com-

[39] In Appendix C.3, we also provide a looser sufficient condition: $\widehat{g}_e < 1 - \frac{1}{\left[\xi\left(1-\mu\frac{\sigma-1}{\sigma}\right)\frac{1+\varepsilon_e}{\varepsilon_e}+1+\frac{\frac{\sigma}{\sigma-1}-\xi\mu\frac{\sigma-1}{\sigma}}{\frac{\sigma}{\sigma-1}-\xi}\right]\widehat{H}}$, where the term in the bracket of the right side of the above inequality is not less than 2, because $\mu \leq \frac{\sigma}{\sigma-1}$.

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

plete the tasks of sales income and profit. Denote $L_w(\theta_e'|\theta_e)$ and $l_e(\theta_e'|\theta_e)$ as the labor input and effort needed to complete the tasks, respectively. Denote $Q_{ij}(x_e(\theta_e)l_e, L_w)$ as the firm-level production function. Combining these two constraints we immediately have

$$L_w(\theta_e'|\theta_e) = \frac{S(\theta_e') - \pi(\theta_e')}{W},$$

and $l_e(\theta_e'|\theta_e)$ satisfies

$$P\left(Q_{ij}\left(x_e(\theta_e)l_e(\theta_e'|\theta_e), \frac{S(\theta_e')-\pi(\theta_e')}{W}\right), \theta_e\right) Q_{ij}\left(x_e(\theta_e)l_e(\theta_e'|\theta_e), \frac{S(\theta_e')-\pi(\theta_e')}{W}\right) = S(\theta_e').$$

Two things worth noting here. First, by enforcing tasks of sales income and profit, the planner can directly control the amount of labor inputs. That is to say, $L_w(\theta_e'|\theta_e)$ is independent of θ_e. Second, notice that $\frac{\partial\left[P\left(Q_{ij},\theta_e\right)Q_{ij}\right]}{\partial Q_{ij}} = P\left(Q_{ij},\theta_e\right)\left[1+\varepsilon_{Q_{ij}}(\theta_e)\right]$ and $\varepsilon_{Q_{ij}}(\theta_e) > -1$. As long as Q_{ij} strictly increases in $l_e(\theta_e'|\theta_e)$ with $P\left(Q_{ij}(0,\cdot),\theta_e\right)Q_{ij}(0,\cdot) = 0$, there exists a unique solution $l_e(\theta_e'|\theta_e)$ for any $S(\theta_e') \geq 0$.

Therefore, under out setup, we can reformulate the entrepreneur's problem as:

$$V_e(\theta_e) \equiv \max_{\theta_e'} \; c_e(\theta_e') - \phi_e\left(l_e(\theta_e'|\theta_e)\right). \tag{56}$$

Solving the above problem, as in the benchmark model, we have

$$\frac{\partial l_e(\theta_e'|\theta_e)}{\partial\theta_e} = -\frac{\frac{\partial P\left(Q_{ij},\theta_e\right)}{\partial\theta_e}Q_{ij} + P\left(Q_{ij},\theta_e\right)\left[1+\varepsilon_{Q_{ij}}(\theta_e)\right]\frac{\partial Q_{ij}}{\partial x_e(\theta_e)}x_e'(\theta_e)}{P\left(Q_{ij},\theta_e\right)\left[1+\varepsilon_{Q_{ij}}(\theta_e)\right]\frac{\partial Q_{ij}}{\partial l_e(\theta_e'|\theta_e)}} \tag{57}$$

and

$$V_e'(\theta_e) = \frac{\phi_e'\left(l_e(\theta_e)\right)Q_{ij}(\theta_e)}{\frac{\partial Q_{ij}(\theta_e)}{\partial l_e(\theta_e)}}\left[\frac{\partial\ln P\left(Q_{ij}(\theta_e),\theta_e\right)}{\partial\theta_e}\frac{1}{1+\varepsilon_{Q_{ij}}(\theta_e)} + \frac{\partial\ln Q_{ij}(\theta_e)}{\partial x_e(\theta_e)}x_e'(\theta_e)\right] \tag{58}$$

which is equivalent to the benchmark incentive-compatible constraint (35), because in the benchmark model $\mu(\theta) = \frac{1}{1+\varepsilon_{Q_{ij}}(\theta_e)}$, $\frac{Q_{ij}(\theta_e)}{\frac{\partial Q_{ij}(\theta_e)}{\partial l_e(\theta_e)}} = l_e(\theta_e)$ and $\frac{\partial\ln Q_{ij}(\theta_e)}{\partial x_e(\theta_e)}x_e'(\theta_e) = \frac{x_e'(\theta_e)}{x_e(\theta_e)}$.

The worker's problem remains the same as before. Therefore, all incentive-compatible allocations satisfying (33), (35) and the resource constraints are feasible. The planner's problem is similar to the one in the benchmark model, except that the policy constraint $\frac{d\ln\varpi(\theta_e)}{d\theta_e} = 0$ is now relaxed.

Denote by $\tau_w^E(\theta_w)$, $\tau_e^E(\theta_e)$ and $\tau_s^E(\theta_e)$ the tax wedges in this extension. See Appendix D.2 for additional explicit expressions for the tax wedges. Analogous to Theorem 1, Theorem 2 provides the most general result on the optimal tax formula in this extension.

Theorem 2 *The optimal tax wedges satisfy the following:*

$$\frac{\tau_w^E(\theta_w)}{1-\tau_w^E(\theta_w)} = \frac{1-\widetilde{\mu}+[1-\bar{g}_w(\theta_w)]\frac{1-F_w(\theta_w)}{f_w(\theta_w)}\frac{x_w'(\theta_w)}{x_w(\theta_w)}\frac{1+\varepsilon_w}{\varepsilon_w}}{\widetilde{\mu}}, \tag{59}$$

$$\frac{\tau_e^E(\theta_e)}{1-\tau_e^E(\theta_e)} = \frac{1-\widetilde{\mu}+[1-\bar{g}_e(\theta_e)]\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu(\theta_e)\frac{\partial\ln P(Q(\theta_e),\theta_e)}{\partial\theta_e}+\frac{x_e'(\theta_e)}{x_e(\theta_e)}\right]}{\widetilde{\mu}}, \tag{60}$$

$$\frac{\tau_s^E(\theta_e)}{1-\tau_s^E(\theta_e)} = \overbrace{\left[\frac{\widetilde{\mu}}{\mu(\theta_e)}-1\right]}^{\widetilde{RE}(\theta_e)} + \overbrace{\left[1-\tau_e^E(\theta_e)\right]\varepsilon_{Q_{-ij}}(\theta_e)\left[\begin{array}{c}[1-g_e(\theta_e)]-\frac{[1-\bar{g}_e(\theta_e)][1-F_e(\theta_e)]}{f_e(\theta_e)} \\ \times\left[\frac{1+\varepsilon_e}{\varepsilon_e}\frac{l_e'(\theta_e)}{l_e(\theta_e)}+\frac{d\ln\left[\mu(\theta_e)\varepsilon_{Q_{-ij}}(\theta_e)\right]}{d\theta_e}\right]\end{array}\right]}^{\widetilde{IRE}(\theta_e)}, \tag{61}$$

where

$$\widetilde{\mu} \equiv \frac{\int_{\theta_e}\frac{\mu(\theta_e)}{1-\tau_s^E(\theta_e)}L_w(\theta_e)f_e(\theta_e)\,d\theta_e}{\int_{\theta_e}L_w(\theta_e)f_e(\theta_e)\,d\theta_e} \tag{62}$$

is a modified average markup.

Proof. See Appendix D.1. ■

In line with our benchmark model, $\widetilde{\mu}$ is reduced to μ when the sales tax is zero. $\widetilde{RE}(\theta_e)$ and $\widetilde{IRE}(\theta_e)$ are analogous to $RE(\theta_e)$ and $IRE(\theta_e)$ defined in Theorem 1. In particular, $\widetilde{RE}(\theta_e)$ and $RE(\theta_e)$ ($\widetilde{IRE}(\theta_e)$ and $IRE(\theta_e)$) are equivalent to each other under linear taxes.

Comparing Theorems 1 and 2, we see that allowing for a non-linear sales tax can indeed improve welfare. However, the first best still cannot be achieved. Since an entrepreneur controls both efforts and labor inputs, contracting on both sales income and profit cannot fully reveal the type. In the benchmark model, labor income and profit taxes mimic the role of a non-linear sales tax. By assuming away non-linear sales tax, our benchmark model offers a clean way to investigate the effects of the profit tax, which is the effective tax rate on factors of production, as well as to conduct numerical analysis.

5.2 Conditioning Taxes on Markups

In our setup so far, the planner can not condition the tax on 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, but it exists for some markets. 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 time and resources to do this estimation.

Nonetheless, we now derive the solution even if the planner has the ability to obtain these markup estimates. We show that in our model setup of production functions, the optimal solution where taxes condition on markups as well as profits is equivalent to the solution in Section 5.1 with non-linear sales tax schedules.[41] This equivalence leads us to conclude that the first-best can not be achieved when taxes condition 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 firms based on their market power can levy a profit tax and punish the firms based on their markups (a tax on markups for example). In particular, it 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 $\pi\left(\theta_e'\right)$ units of profit. Then the entrepreneur will receive $c_e\left(\theta_e'\right)$ units of consumption. Supposing there is a unique solution to the promise-keeping constraints, the bundle $L_w\left(\theta_e'|\theta_e\right)$ and $l_e\left(\theta_e'|\theta_e\right)$ satisfies:

$$\frac{P\left(Q_{ij},\theta_e\right)\frac{\partial Q_{ij}}{\partial L_w}}{W}=\mu\left(\theta_e'\right) \quad \text{and} \quad P\left(Q_{ij},\theta_e\right)Q_{ij}-WL_w=\pi\left(\theta_e'\right). \tag{63}$$

The entrepreneur's problem can again be formulated as in equation (56). Thus, the incentive condition of entrepreneur is the same as that in Section 5.1, if and only if (57) holds here too. A sufficient condition for (57) is that $L_w\left(\theta_e'|\theta_e\right)$ is independent of θ_e, which is true under our benchmark model. To see this, combine the promise-keeping constraints to derive $L_w\left(\theta_e'|\theta_e\right)=\frac{\pi(\theta_e')}{W[\mu(\theta_e')/\xi-1]}$, where we substitute $\frac{\partial Q_{ij}}{\partial L_w}\frac{1}{Q_{ij}}$ by $\frac{\xi}{L_w}$. Now that the incentive condition remains the same, we know from the solution to the non-linear sales tax problem that the first best is not achievable. The main insight of this robustness exercise is that we highlight an important new friction in the optimal taxation analysis: market power, even if observable,

With more general firm-level technology however – for example when there is sorting and a worker's labor is not supplied in efficiency units –, $L_w\left(\theta_e'|\theta_e\right)$ may depend on θ_e. Then,

$$\frac{\partial l_e\left(\theta_e'|\theta_e\right)}{\partial\theta_e}=-\frac{\frac{\partial P\left(Q_{ij},\theta_e\right)}{\partial\theta_e}Q_{ij}+P\left(Q_{ij},\theta_e\right)\left[1+\varepsilon_{Q_{ij}}(\theta_e)\right]\left[\frac{\partial Q_{ij}}{\partial x_e(\theta_e)}x_e'\left(\theta_e\right)+\frac{\partial Q_{ij}}{\partial L_w(\theta_e'|\theta_e)}\frac{\partial L_w(\theta_e'|\theta_e)}{\partial\theta_e}\right]-W\frac{\partial L_w(\theta_e'|\theta_e)}{\partial\theta_e}}{P\left(Q_{ij},\theta_e\right)\left[1+\varepsilon_{Q_{ij}}(\theta_e)\right]\frac{\partial Q_{ij}}{\partial l_e(\theta_e'|\theta_e)}},$$

and (57) no longer holds, unless $P\left(Q_{ij},\theta_e\right)\left[1+\varepsilon_{Q_{ij}}(\theta_e)\right]\frac{\partial Q_{ij}}{\partial L_w(\theta_e'|\theta_e)}=W$. Since the latter equation does not generally hold everywhere, the incentive condition of entrepreneur is different from that in subsection 5.1. The optimal markup tax as well as all other taxes therefore depend on the details of the production technology.

5.3 Capital Investment

We do not explicitly model capital in our benchmark model. However, the the problem can be modeled parallel with capital in place of entrepreneurial effort. The most relevant assumption is that part of the cost (or benefit) from factors cannot be deducted before the profit tax (either because the cost is unobservable or legally excluded from the deductible costs).

[41] In general, for different production technologies, this equivalence result may not hold.

Formally, consider an economy where the entrepreneur chooses labor inputs L_w and capital investment K, instead of effort, to maximize the utility:

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

$Q_{ij}\left(K,L_w\right)$ is a firm-level production function of capital and labor inputs, and r is the market price of capital,[42] and $\phi_K\left(K,\theta_e\right)$ is the unobservable cost of investment, which may be dependent on the type of entrepreneur.

In the real economy, although the market price of capital (i.e., r) can be observed, the total opportunity costs of investments generally cannot be observed. The unobservable part of cost is captured by $\phi_K\left(K,\theta_e\right)$, which may include the cost of raising and managing funds.[43] An alternative explanation for $\phi_K\left(K,\theta_e\right)$ is the preference for asset (wealth). Under this case, $\phi_K\left(K,\theta_e\right)$ can be negative, which means investment directly generates positive utility. The common ground of the above situations is that the elasticity of investment may be finite, which is the key point of Saez and Stantcheva (2018). In the above cases, $\pi = P\left(Q_{ij}\left(K,L_w\right),\theta_e\right)Q_{ij}\left(K,L_w\right) - WL_w - rK$.

In the real economy, not all costs of investment are deductible before the profit tax. For example, the interest of debt can is deductible before tax, but the equity investment is not. Although equity investments occupy the case flow of shareholders and thus also generate costs. Under this case, even if $\phi_K\left(K,\theta_e\right) = 0$, there are capital costs that cannot be deducted before the tax. Moreover, $\pi = P\left(Q_{ij}\left(K,L_w\right),\theta_e\right)Q_{ij}\left(K,L_w\right) - WL_w$. It is worth noting that in all the cases above, our main results will still hold.

6 Numerical Analysis

Our general results depend on the social preferences for redistribution. To see the overall impact of market power on optimal taxation, we numerically analyze an economy with concave social welfare functions with $G\left(V\right) = \frac{V^{1-k}}{1-k}$. The parameter k governs the concavity of the social welfare function and, therefore, the desire for redistribution by the planner. We provide the optimal tax rates for $k = 1$ (as is in Sachs, Tsyvinski, and Werquin (2020)). Our objective is to measure the variation in the equilibrium allocation and the optimal tax policy as market power, measured by the number of competitors I within each market. The fewer competitors I, the more market power firms have.

We maintain the following assumptions for the numerical analysis as follow. We treat θ_e and θ_w as the

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

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

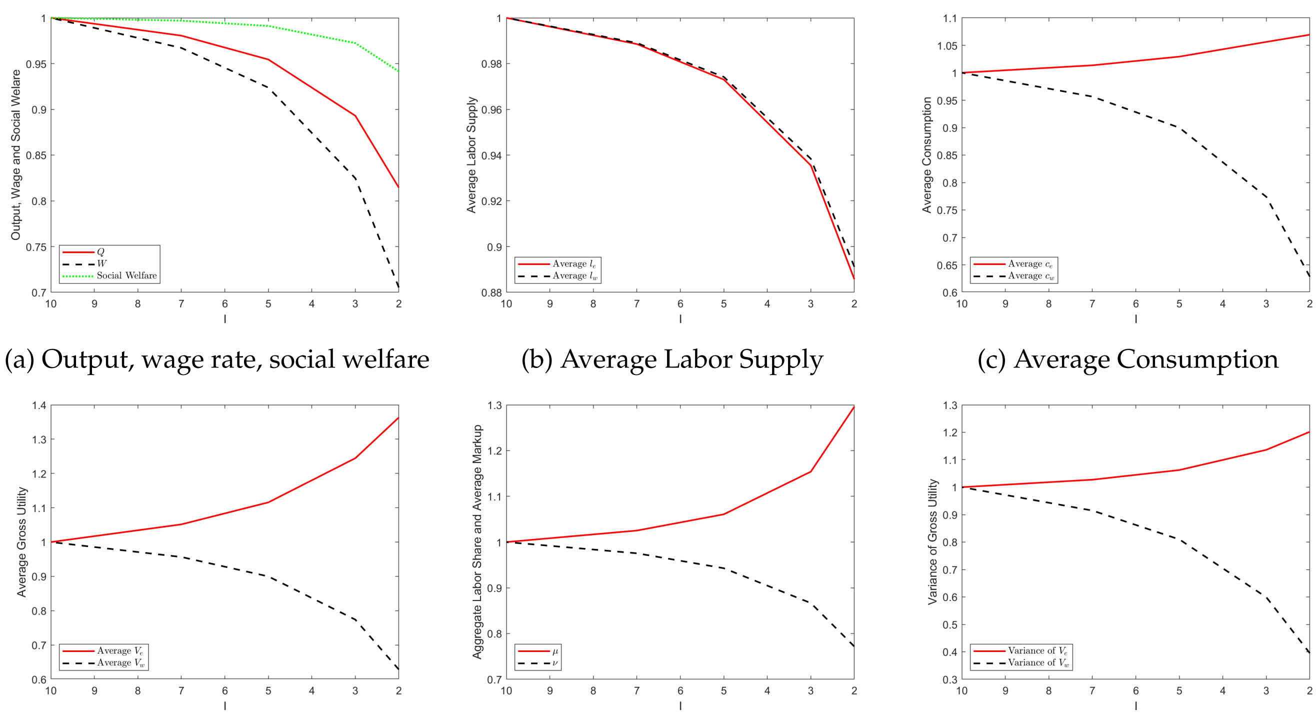
where $K\left(\theta_e\right)$ is the investment of θ_e firms.

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

quantiles of $\pi(\theta_e)$ and $y(\theta_w)$, which means $f_o = 1$ is uniform on $\Theta_o = [0,1]$. Since the functions $x_o(\theta_o)$ and $\chi(\theta_e)$ are used to govern the heterogeneity, there is no loss to assume that the distribution is uniform. The full parameterization is detailed in Table 1.

Table 1: Parameterization

$G(V) = \frac{V^{1-k}}{1-k}$	social welfare function
$k \in \{1,3\}$	concavity of the social welfare function; $k = 1$ is benchmark
$f_o(\theta_o) = 1$	PDF of skills
$N_e = 0.2$	measure of entrepreneurs
$A = 10^4$	the TFP of final good production technology Q
$\xi = 0.85$	concavity of technology Q_{ij}
$\sigma = 1.5$	elasticity of substitution between submarkets
$\eta(\theta_e) = 10 - 8\theta_e$	elasticity of substitution within submarkets
$x_o(\theta_o) = \theta_o$	individual-level productivity
$\chi(\theta_e) = \theta_e$	distribution parameter
$\varepsilon_o = 0.33$	the elasticity of labor supply (Chetty (2012))



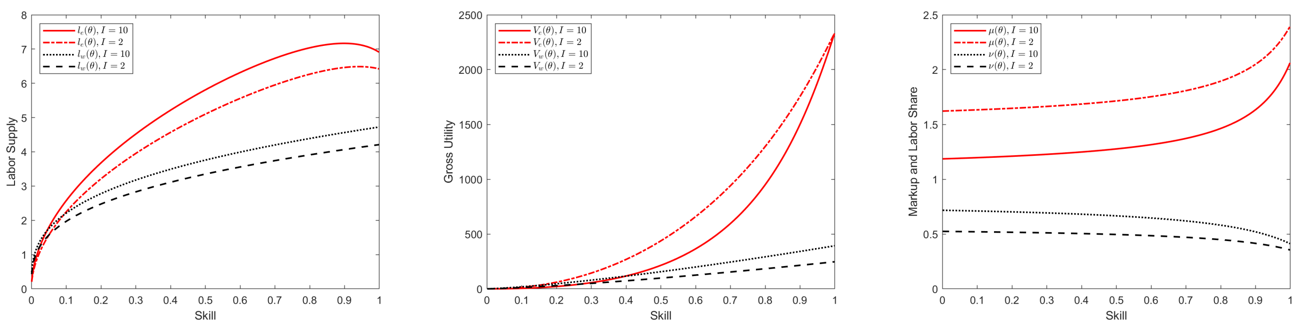
(a) Output, wage rate, social welfare (b) Average Labor Supply (c) Average Consumption

(d) Average Utility (e) Average markup and labor share (f) Variance of Utility

Figure 1: Laissez-faire economy: Effect of market Power (number of competitors I); normalize to 1 when $I = 10$

Laissez-faire Economy. To benchmark our taxation results, we first summarize the properties of the competitive equilibrium allocation without taxation. Figure 1 summarizes the effect of a change in market power in *all* submarkets. We plot the number of competitors on the horizontal axis in a decreasing order, to indicate increasing market power. The number of competitors within a market varies between $I = 10$ (competitive) and $I = 2$, duopoly. Most striking is the massive decline in the wage rate W by 70% (Figure 1a). Output drops by 18% and welfare by 6%. The welfare effect is mitigated due to the decline in labor supply by 11% (Figure 1b). Also, entrepreneurs decrease their labor supply despite the fact that they get higher profits and higher consumption. The reason is that with the Lucas (1978) span-of-control technology, the effort of entrepreneurs and workers are complements. Consumption (Figure 1c) and utility (Figure 1d) is increasing for entrepreneurs and decreasing for workers. This is the main inequality generating force of markups: the division of output between profits and labor income. This is consistent with the increase in the aggregate markup and the decrease in the average labor share (Figure 1e). Finally, inequality within entrepreneurs is increasing while inequality within workers is decreasing (Figure 1f). The latter stems from the labor supply response of the workers to a lower wage rate W.

In Figure 2, we report how the equilibrium outcomes vary by skill. The labor supply of all agents is increasing in skill for most types, except at the very top of the entrepreneur's type distribution. Finally, markups are increasing and the labor shares are decreasing in entrepreneur type.



(a) Labor Supply (b) Utility (c) Markups and Labor Share

Figure 2: Laissez-faire economy: Variation by skill θ

Optimal Taxation. Next, we analyze how optimal taxation policy varies with market power. To set the stage, in Table 2 we summarize the different tax measures that we use in the numerical analysis.

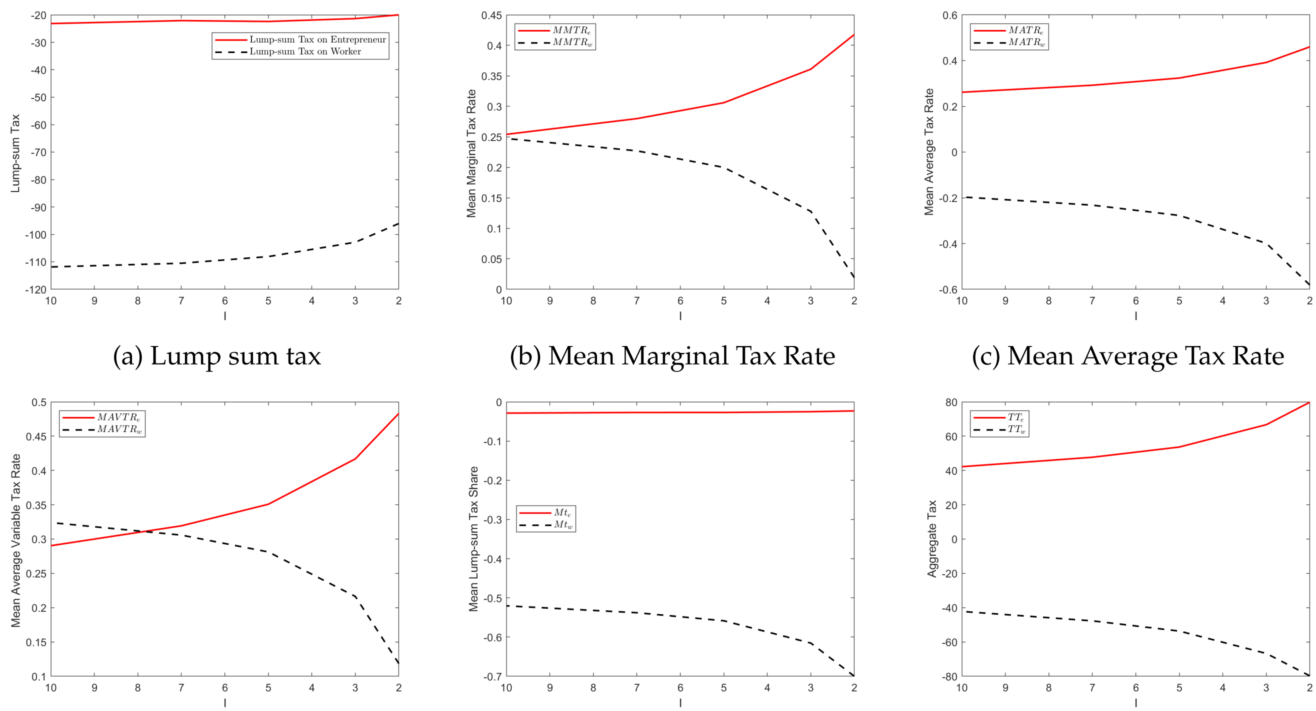
Figure 3 graphically represents how optimal taxation changes as market power increases. In this exercise, we set the tax revenue to be collected by the government equal to zero: $R = 0$. First, we find that the lump sum taxes increase in market power for both workers and entrepreneurs (Figure 3a). Lump-sum taxes are negative because the marginal tax rate is on average positive. Over the entire distribution, Figure 3b shows that the mean of the marginal tax rate is increasing in market power for the entrepreneurs and decreasing for workers. The same is true for the mean of the average tax rate (Figure 3c). This tells us that the optimal tax acts as a Pigouvian tax to correct the inefficiency (externality) due to market power: the higher profits that the entrepreneurs earn and the lower labor income that the workers earn are due to an inefficiency that the tax system corrects. The variable component of the average tax rate is increasing for entrepreneurs and decreasing for workers (Figure 3d), while the lump-sum component is constant for entrepreneurs and decreasing for workers (Figure 3e). In line with this, the total tax burden for the

Table 2: Summary of Tax Measures

t_o	Lump-sum tax (depends on occupation, not on incomes)
$\tau_o(\theta_o)$	Marginal tax rate
$T_o(y(\theta_o))$	Tax burden
$ATR_o(\theta_o) = \frac{T_o(y(\theta_o))}{y(\theta_o)}$	Average tax rate
$AVTR_o(\theta_o) = \frac{T_o(y(\theta_o))-t_o}{y(\theta_o)}$	Average variable tax rate
$TT_o = N_o \int_{\theta_o} T_o(y(\theta_o)) f_{\theta_o}(\theta_o) d\theta_o$	Total tax burden
$MMTR_o = \int_{\theta_o} \tau_o(\theta_o) f_{\theta_o}(\theta_o) d\theta_o$	Mean marginal tax rate
$MATR_o = \frac{TT_o}{N_o \int_{\theta_o} y(\theta_o) f_{\theta_o}(\theta_o) d\theta_o}$	Mean average tax rate
$MAVTR_o = \frac{TT_o - t_o * N_o}{N_o \int_{\theta_o} y(\theta_o) f_{\theta_o}(\theta_o) d\theta_o}$	Mean average variable tax rate
$Mt_o = \frac{t_o}{\int_{\theta_o} y(\theta_o) f_{\theta_o}(\theta_o) d\theta_o}$	Mean lump-sum tax share

Note: we denote profits by $\pi(\theta_e) = y(\theta_e)$.

entrepreneurs is increasing while it is decreasing for the workers (Figure 3f).

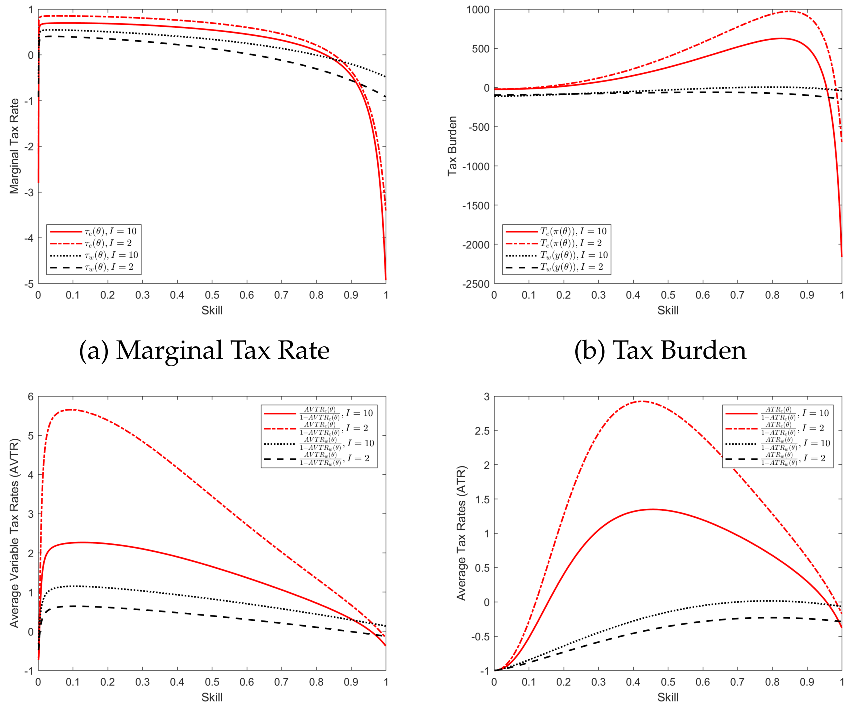


(a) Lump sum tax (b) Mean Marginal Tax Rate (c) Mean Average Tax Rate

(d) Mean Average Variable Tax Rate (e) Mean Lump-sum Tax Share (f) Total Tax Burden

Figure 3: Optimal Taxation: Effect of market Power (number of competitors I)

Next, in Figure 4, we report how optimal taxes vary by skill. The marginal tax rate is decreasing in skill (except for those with very low skills), in order to provide incentives to exert effort (Figure 4a). This is the standard Mirrleesian incentive property. To provide incentives to the top entrepreneurs, a sharp decline until negative in the marginal rate is needed. Because the incomes are large, the total tax burden is mostly increasing, but it decreases at the top entrepreneurs (Figure 4b). The average variable tax rate and



(a) Marginal Tax Rate

(b) Tax Burden

(c) Average Variable Tax Rate

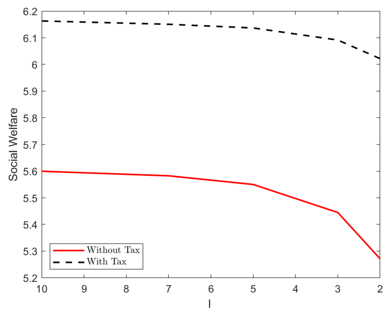
(d) Average Tax Rate

Figure 4: Optimal Taxation: Variation by skill θ

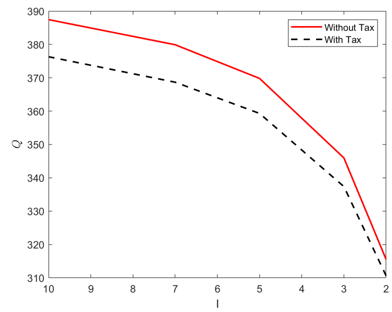
the average tax rates are inverted U-shaped because the marginal tax rate is generally decreasing in skills, which is higher than the average tax rate at the beginning and lower than the average tax after a certain income level.

Comparing economies with and without taxes. Next, in Figure 5 we compare the equilibrium outcome of the Laissez-faire economy with the optimal taxation economy. Not surprisingly, social welfare is higher under optimal taxation than in Laissez-faire (Figure 5a). However, the output is lower (Figure 5b), where the output decline is lowest under high market power. This is due to the fact that under optimal taxes, there is a decline in effort (Figure 5c). Note that although the average labor supply decreases after the tax, the labor supplies of high-skill entrepreneurs actually increase. Also, the firm-level labor supply suggests that although before tax labor supplies of entrepreneurs may decrease with the skill, it is increasing with the skill after the tax.

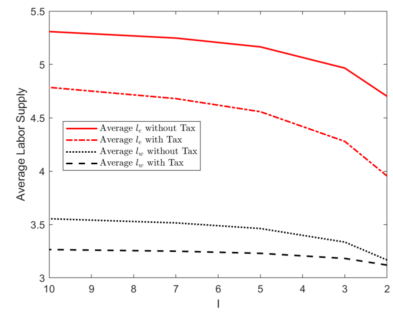
As a result of the lower output produced and hence the lower aggregate demand for labor, there is a decline in the equilibrium wage rate W (Figure 5d). In equilibrium, optimal taxation has an adverse effect. It increases the markup and decreases the labor share (Figure 5e). Though the effect is small, it is important to see how taxation of income has adverse effects on market power and labor share. This adverse effect also shows up in the before-tax profit rates that are higher under optimal taxation. Due to the Reallocation Effect, the regressive tax reallocates factors from the low-markup firms to the high-markup firms. Finally, taxes sharply reduce inequality. The variance of gross utility is lower for both entrepreneurs and workers (Figure 5f), but remarkably more so for entrepreneurs.



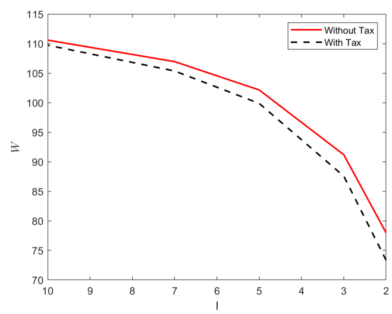
(a) Social Welfare



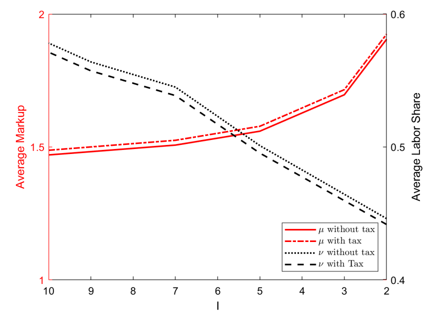
(b) Output



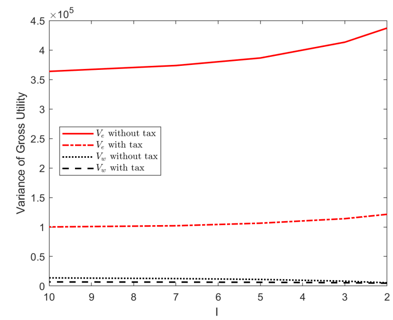
(c) Effort



(d) Equilibrium wage rate W



(e) Markup and Labor Share



(f) Variance of Utility

Figure 5: Comparing variables under zero tax and optimal tax

7 Conclusion

The best way to address market power is to cut out the root cause with an antitrust policy. In its absence, we ask what the role is for income taxation to address the inefficiency and inequality that market power creates. In a standard partial equilibrium setting, taxing profits redistributes resources but does not affect optimal production. In a Mirrleesian setting, income and profit taxes do affect optimal production due to the incentive constraint and endogenous labor supply and general equilibrium wages.

We show in the Laissez-faire economy that market power increases profits, lowers the equilibrium wage rate and that it leads to lower effort, output and welfare. In response, optimal taxation can help correct the externality caused by market power, and the income tax plays a Pigouvian role. Typically, higher market power leads to higher marginal tax rates on entrepreneurs and lower marginal rates on workers.

APPENDIX

A Environment

A.1 The Cournot Competitive Tax Equilibrium

When first-order conditions are both necessary and sufficient to both the individuals' and final good producer's problems, the equilibrium allocations are determined by (13) to (19) and the individuals' budget constraints. Under the technology considered in this paper and $\phi_o(l_o) = \frac{l_o^{1+\frac{1}{\varepsilon_o}}}{1+\frac{1}{\varepsilon_o}}$, we have the following conditions in the symmetric equilibrium:

1. First-order conditions

$$P(\theta_e) = [Nf_e(\theta_e)]^{-\frac{1}{\sigma}} \widetilde{\chi}(\theta_e) A^{\frac{\sigma-1}{\sigma}} Q_{ij}(\theta_e)^{-\frac{1}{\sigma}} Q^{\frac{1}{\sigma}}, \tag{A1}$$

and

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

and

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

and

$$\left[1 + \frac{\partial \ln P\left(Q_{ij}(\theta_e), \theta_e\right)}{\partial \ln Q_{ij}(\theta_e)}\right] P(\theta_e) Q_{ij}(\theta_e)\,(1-t_s)\left[1 - T_e'(\pi(\theta_e))\right] = l_e(\theta_e)^{1+\frac{1}{\varepsilon_e}}, \theta_o \in \Theta_o. \tag{A4}$$

Combination of (A2) and (20) (i.e., $\mu(\theta_e) = \frac{P(\theta_e)}{W/\left[\frac{\partial Q_{ij}(\theta_e)}{\partial L_w(\theta_e)}(1-t_s)\right]} = \frac{\xi P(\theta_e) Q_{ij}(\theta_e)(1-t_s)}{WL_w(\theta_e)}$) delivers (22). Substituting $1 + \frac{\partial \ln P\left(Q_{ij}(\theta_e), \theta_e\right)}{\partial \ln Q_{ij}(\theta_e)}$ by (22), we have

$$WL_w(\theta_e) = \frac{\xi(1-t_s)}{\mu(\theta_e)} P(\theta_e) Q_{ij}(\theta_e), \tag{A5}$$

and

$$\frac{P(\theta_e) Q_{ij}(\theta_e)(1-t_s)}{\mu(\theta_e)}\left[1 - T_e'(\pi(\theta_e))\right] = l_e(\theta_e)^{1+\frac{1}{\varepsilon_e}}, \theta_e \in \Theta_e. \tag{A6}$$

2. Inverse demand function

$$P(Q_{ij}, \theta_e) = \chi(\theta_e) A^{\frac{\sigma-1}{\sigma}} Q_{ij}^{-\frac{1}{\eta(\theta_e)}} I^{-\left[\frac{1}{\eta(\theta_e)} - \frac{1}{\sigma}\right]\frac{\eta(\theta_e)}{\eta(\theta_e)-1}} \begin{bmatrix} (I-1)\, Q_{ij}(\theta_e)^{\frac{\eta(\theta_e)-1}{\eta(\theta_e)}} \\ +Q_{ij}^{\frac{\eta(\theta_e)-1}{\eta(\theta_e)}} \end{bmatrix}^{\left[\frac{1}{\eta(\theta_e)} - \frac{1}{\sigma}\right]\frac{\eta(\theta_e)}{\eta(\theta_e)-1}} \left[\frac{Q}{N}\right]^{\frac{1}{\sigma}}, \tag{A7}$$

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

3. Labor market clear condition

$$\int_{\theta_w} x_w(\theta_w)\, l_w(\theta_w)\, f_w(\theta_w) d\theta_w = W^{\varepsilon_w} \int_{\theta_w} [\varkappa(\theta_w)]^{\varepsilon_w+1} [1 - t_w(\theta_w)]^{\varepsilon_w} f_w(\theta_w) d\theta_w \tag{A8}$$

4. Meanwhile, in the equilibrium, we have

$$Q = \int_{\theta_e} N f_e(\theta_e) \left[P(\theta_e) Q_{ij}(\theta_e)\right] d\theta_e. \tag{A9}$$

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. Lastly, one can derive other allocations with individuals' budget constraints.

A.2 Laissez-faire Economy

Combining (A5), (A6), and (A1) gives

$$l_e(\theta_e) = \left[\frac{A^{\frac{\sigma-1}{\sigma}} x_e(\theta_e)^{\frac{\sigma-1}{\sigma}} \chi(\theta_e)}{\mu(\theta_e)} \left(\frac{Q}{N}\right)^{\frac{1}{\sigma}}\right]^{\frac{\sigma\varepsilon_e}{\varepsilon_e+\sigma}} L_w(\theta_e)^{\frac{\xi(\sigma-1)\varepsilon_e}{\varepsilon_e+\sigma}} \tag{A10}$$

Substituting $P(\theta_e)$ and $Q_{ij}(\theta_e)$ in (A5) with (A1) and $Q_{ij}(\theta_e) = [x_e(\theta_e)\, l_e(\theta_e)]\, L_w(\theta_e)^{\xi}$, respectively, we have

$$\begin{aligned} L_w(\theta_e) &= \frac{\xi}{W\mu(\theta_e)} \chi(\theta_e) \left[x_e(\theta_e) l_e(\theta_e) L_w(\theta_e)^{\xi}\right]^{\frac{\sigma-1}{\sigma}} \left(\frac{Q}{N}\right)^{\frac{1}{\sigma}} \\ &= \frac{\xi}{W} \left[\frac{A^{\frac{\sigma-1}{\sigma}} x_e(\theta_e)^{\frac{\sigma-1}{\sigma}} \chi(\theta_e)}{\mu(\theta_e)} \left(\frac{Q}{N}\right)^{\frac{1}{\sigma}}\right]^{\frac{\sigma(\varepsilon_e+1)}{\varepsilon_e+\sigma}} L_w(\theta_e)^{\frac{\xi(\sigma-1)(\varepsilon_e+1)}{\sigma+\varepsilon_e}}, \end{aligned} \tag{A11}$$

where we substitute $l_e(\theta_e)$ with (A10) in the second equation.

Rearranging the above equation gives

$$L_w(\theta_e) = \left(\frac{\xi}{W}\right)^{\frac{\sigma+\varepsilon_e}{\sigma+\varepsilon_e-\xi(\sigma-1)(\varepsilon_e+1)}} \left[\frac{A^{\frac{\sigma-1}{\sigma}} x_e(\theta_e)^{\frac{\sigma-1}{\sigma}} \chi(\theta_e)}{\mu(\theta_e)} \left(\frac{Q}{N}\right)^{\frac{1}{\sigma}}\right]^{\frac{\sigma(\varepsilon_e+1)}{\sigma+\varepsilon_e-\xi(\sigma-1)(\varepsilon_e+1)}}. \tag{A12}$$

Substituting the above equation into (A10), we have

$$P(\theta_e) Q_{ij}(\theta_e) = \mu(\theta_e) \left(\frac{\xi}{W}\right)^{\frac{\xi(\sigma-1)(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}} \left[\frac{A^{\frac{\sigma-1}{\sigma}} x_e(\theta_e)^{\frac{\sigma-1}{\sigma}} \chi(\theta_e)}{\mu(\theta_e)} \left(\frac{Q}{N}\right)^{\frac{1}{\sigma}}\right]^{\frac{\sigma(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}} \tag{A13}$$

and

$$l_e(\theta_e) = \left(\frac{\xi}{W}\right)^{\frac{\xi(\sigma-1)\varepsilon_e}{\sigma+\varepsilon_e-\xi(\sigma-1)(\varepsilon_e+1)}} \left[\frac{A^{\frac{\sigma-1}{\sigma}} x_e(\theta_e)^{\frac{\sigma-1}{\sigma}} \chi(\theta_e)}{\mu(\theta_e)} \left(\frac{Q}{N}\right)^{\frac{1}{\sigma}}\right]^{\frac{\sigma\varepsilon_e}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}}. \tag{A14}$$

Equation (A3) gives

$$l_w(\theta_w) = [W x_w(\theta_w)]^{\varepsilon_w}. \tag{A15}$$

The three equations above together with (A8) and (A9) solve the symmetric equilibrium allocation $\{L_w(\theta_e), l_e(\theta_e), l_w(\theta_w)\}$, price system $\{P(\theta_e), W\}$, and total output Q. Lastly, one can derive other allocations with individuals' budget constraints. See below for details.

For later use, we define

$$\begin{aligned} A_1 &= \int_{\theta_e} N f_e(\theta_e)\,\mu(\theta_e)\left[\frac{A^{\frac{\sigma-1}{\sigma}} x_e(\theta_e)^{\frac{\sigma-1}{\sigma}}\chi(\theta_e)}{\mu(\theta_e)\,N^{\frac{1}{\sigma}}}\right]^{\frac{\sigma(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}} d\theta_e, \\ A_2 &= \int N f_e(\theta_e)\left[\frac{A^{\frac{\sigma-1}{\sigma}} x_e(\theta_e)^{\frac{\sigma-1}{\sigma}}\chi(\theta_e)}{\mu(\theta_e)\,N^{\frac{1}{\sigma}}}\right]^{\frac{\sigma(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}} d\theta_e, \\ A_3 &= N_w \xi^{\varepsilon_w}\int_{\theta_w} x(\theta_w)^{\varepsilon_w+1} f_w(\theta_w)d\theta_w. \end{aligned} \tag{A16}$$

Substituting $L_w(\theta_e)$ in (A5) with (A12), we have

$$P(\theta_e)Q_{ij}(\theta_e) = \mu(\theta_e)\left(\frac{\xi}{W}\right)^{\frac{\xi(\sigma-1)(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}}\left[\frac{A^{\frac{\sigma-1}{\sigma}} x_e(\theta_e)^{\frac{\sigma-1}{\sigma}}\chi(\theta_e)}{\mu(\theta_e)}\left(\frac{Q}{N}\right)^{\frac{1}{\sigma}}\right]^{\frac{\sigma(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}} \tag{A17}$$

Substituting $P(\theta_e)Q_{ij}(\theta_e)$ in (A9) with (A17), we have

$$Q = \int_{\theta_e} N f_e(\theta_e)\,\mu(\theta_e)\left(\frac{\xi}{W}\right)^{\frac{\xi(\sigma-1)(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}}\left[\frac{A^{\frac{\sigma-1}{\sigma}} x_e(\theta_e)^{\frac{\sigma-1}{\sigma}}\chi(\theta_e)}{\mu(\theta_e)}\left(\frac{Q}{N}\right)^{\frac{1}{\sigma}}\right]^{\frac{\sigma(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}} d\theta_e,$$

which gives the following equation by the definition of A_1:

$$Q = \left(\frac{\xi}{W}\right)^{\frac{\xi(\varepsilon_e+1)}{1-\xi(\varepsilon_e+1)}} A_1^{\frac{\sigma+\varepsilon_e-\xi(\sigma-1)(\varepsilon_e+1)}{1-\xi(\varepsilon_e+1)}\frac{1}{\sigma-1}}. \tag{A18}$$

Similarly, substituting $L_w(\theta_e)$ in (A8) with (A12), we have the aggregate labor demand

$$L^D \equiv \left(\frac{\xi}{W}\right)^{\frac{1}{1-\xi(\varepsilon_e+1)}}(A_1)^{\frac{\varepsilon_e+1}{(\sigma-1)[1-\xi(\varepsilon_e+1)]}} A_2. \tag{A19}$$

On the other hand, according to (A8) and (A15), we have the aggregate labor supply

$$L^S \equiv N_w[W]^{\varepsilon_w}\int_{\theta_w} x(\theta_w)^{\varepsilon_w+1} f_w(\theta_w)d\theta_w = \left[\frac{W}{\xi}\right]^{\varepsilon_w} A_3. \tag{A20}$$

Combining (A19) and (A20) gives

$$\left[\frac{W}{\xi}\right]^{\varepsilon_w+\frac{1}{1-\xi(\varepsilon_e+1)}} = (A_1)^{\frac{\varepsilon_e+1}{(\sigma-1)[1-\xi(\varepsilon_e+1)]}}\,\frac{A_2}{A_3}, \tag{A21}$$

that is

$$W = \xi\left[(A_1)^{\frac{\varepsilon_e+1}{(\sigma-1)[1-\xi(\varepsilon_e+1)]}}\,\frac{A_2}{A_3}\right]^{\frac{1}{\varepsilon_w+\frac{1}{1-\xi(\varepsilon_e+1)}}}. \tag{A22}$$

Lastly, substituting W in (A18) with (A22), we have

$$Q = \left[\frac{A_3}{A_2 A_1^{\frac{\varepsilon_e+1}{(\sigma-1)[1-\xi(\varepsilon_e+1)]}}}\right]^{\frac{1}{\varepsilon_w+\frac{1}{1-\xi(\varepsilon_e+1)}}\frac{\xi(\varepsilon_e+1)}{1-\xi(\varepsilon_e+1)}} A_1^{\frac{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}{(\sigma-1)[1-\xi(\varepsilon_e+1)]}}. \tag{A23}$$

Then we can derive $l_w(\theta_w)$, $L_w(\theta_e)$, and $l_e(\theta_e)$ by substituting Q and W into (A12), (A14), and (A15). Moreover, by definition, we have

$$c_e(\theta_e) = [\mu(\theta_e)-\xi]\left(\frac{\xi}{W}\right)^{\frac{\xi(\sigma-1)(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}}\left[\frac{A^{\frac{\sigma-1}{\sigma}}x_e(\theta_e)^{\frac{\sigma-1}{\sigma}}\chi(\theta_e)}{\mu(\theta_e)}\left(\frac{Q}{N}\right)^{\frac{1}{\sigma}}\right]^{\frac{\sigma(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}}, \tag{A24}$$

and

$$Q_{ij}(\theta_e) = x_e(\theta_e)\left(\frac{\xi}{W}\right)^{\frac{\xi\sigma(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}}\left[\frac{A^{\frac{\sigma-1}{\sigma}}x_e(\theta_e)^{\frac{\sigma-1}{\sigma}}\chi(\theta_e)}{\mu(\theta_e)}\left(\frac{Q}{N}\right)^{\frac{1}{\sigma}}\right]^{\frac{\sigma(\varepsilon_e+1)\xi+\sigma\varepsilon_e}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}}, \tag{A25}$$

and

$$P(\theta_e) = \frac{\mu(\theta_e)}{x_e(\theta_e)}\left(\frac{\xi}{W}\right)^{\frac{-\xi(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}}\left[\frac{A^{\frac{\sigma-1}{\sigma}}x_e(\theta_e)^{\frac{\sigma-1}{\sigma}}\chi(\theta_e)}{\mu(\theta_e)}\left(\frac{Q}{N}\right)^{\frac{1}{\sigma}}\right]^{\frac{\sigma-\sigma(\varepsilon_e+1)\xi}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}}. \tag{A26}$$

In addition, we have

$$V(\theta_e) = c_e(\theta_e) - l_e(\theta_e)^{\frac{\varepsilon_e+1}{\varepsilon_e}} = \left[\mu(\theta_e)-\xi-\frac{\varepsilon_e}{\varepsilon_e+1}\right] l_e(\theta_e)^{\frac{\varepsilon_e+1}{\varepsilon_e}}. \tag{A27}$$

According to the above results, we have

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

Note that $\frac{\varepsilon_e+1}{\varepsilon_e}-\frac{\sigma-1}{\sigma}\left(1+\frac{\varepsilon_e+1}{\varepsilon_e}\xi\right)$, $\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)$ and $\frac{\sigma}{\sigma-1}-\left(\frac{\varepsilon_e}{\varepsilon_e+1}+\xi\right)$ are positive under condition (29). Under such a condition, whether $l_e(\theta_e)$, $L_w(\theta_e)$, $Q_{ij}(\theta_e)$, $Q_{ij}(\theta_e)P(\theta_e)$ and $V(\theta_e)$ increases with θ_e is determined by the relative change of $X(\theta_e)$ to $\mu(\theta_e)$. In addition, we have

$$
\begin{aligned}
\frac{d\ln V(\theta_e)}{d\theta_e} &= \left[\frac{\mu(\theta_e)}{\mu(\theta_e)-\left(\xi+\frac{\varepsilon_e}{\varepsilon_e+1}\right)}-\frac{\frac{\sigma}{\sigma-1}}{\frac{\sigma}{\sigma-1}-\left(\xi+\frac{\varepsilon_e}{\varepsilon_e+1}\right)}\right]\frac{\mu'(\theta_e)}{\mu(\theta_e)}+\frac{\frac{\sigma}{\sigma-1}\frac{d\ln X(\theta_e)}{d\theta_e}}{\frac{\sigma}{\sigma-1}-\left(\frac{\varepsilon_e}{\varepsilon_e+1}+\xi\right)}\\
&\geq \frac{\frac{\sigma}{\sigma-1}\frac{d\ln X(\theta_e)}{d\theta_e}}{\frac{\sigma}{\sigma-1}-\left(\frac{\varepsilon_e}{\varepsilon_e+1}+\xi\right)},
\end{aligned}
$$

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

Also,

$$
V_e'(\theta_e)=l_e(\theta_e)\phi'(l_e(\theta_e))\left[\mu(\theta_e)\frac{\partial\ln P(Q_{ij}(\theta_e),\theta_e)}{\partial\theta_e}+\frac{x_e'(\theta_e)}{x_e(\theta_e)}\right], \tag{A28}
$$

and

$$
\begin{aligned}
&\mu(\theta_e)\frac{\partial\ln P(Q_{ij},\theta_e)}{\partial\theta_e}+\frac{x_e'(\theta_e)}{x_e(\theta_e)} \qquad\qquad (A29)\\
=\ &\mu(\theta_e)\left\{\frac{\chi'(\theta_e)}{\chi(\theta_e)}+\left[\frac{\sigma-1}{\sigma}-\frac{1}{\mu(\theta_e)}\right]\frac{d\ln Q_{ij}(\theta_e)}{d\theta_e}\right\}+\frac{x_e'(\theta_e)}{x_e(\theta_e)}\\
=\ &\frac{\sigma(\varepsilon_e+1)\left[\mu(\theta_e)-\xi\right]-\sigma\varepsilon_e}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}\frac{d\ln X(\theta_e)}{d\theta_e}\\
&+\frac{(\sigma-1)\left[\varepsilon_e+\xi(\varepsilon_e+1)\right]\left[\frac{\sigma}{\sigma-1}-\mu(\theta_e)\right]}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}\frac{d\ln\mu(\theta_e)}{d\theta_e}.
\end{aligned}
$$

Notice that $\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)$ is positive under condition (29). The entrepreneurial skill premium increases with $\frac{d\ln X(\theta_e)}{d\theta_e}$. Moreover, since $\mu(\theta_e)\leq\frac{\sigma}{\sigma-1}$, the second term on the right side of (A29) is positive under condition (29), which suggests that the θ_e-type entrepreneur's skill premium increases with the introduce of $\frac{d\ln\mu(\theta_e)}{d\theta_e}$.■

A.3 Proof of Proposition 1

Part 1 of Proposition 1 can be derived by (17) and (22). We now prove part 2. By (A23), (A22), and (A20), we have

$$\nu(I) \triangleq \frac{WL}{\xi Q} = \frac{A_2}{A_1}, \tag{A30}$$

where L is the aggregate labor inputs. Substituting A_1 and A_2 by (A16), we have

$$\nu(I) = \frac{\int f_e(\theta_e)\left[\frac{x_e(\theta_e)^{\frac{\sigma-1}{\sigma}}\chi(\theta_e)}{\mu(\theta_e)}\right]^{\frac{\sigma(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}} d\theta_e}{\int_{\theta_e} f_e\left(\theta_e\right)\mu\left(\theta_e\right)\left[\frac{x_e(\theta_e)^{\frac{\sigma-1}{\sigma}}\chi(\theta_e)}{\mu(\theta_e)}\right]^{\frac{\sigma(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}} d\theta_e}. \tag{A31}$$

For the convenience of analysis, define

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

Then we have

$$\nu\left(I\right) = \int_{\theta_e}\frac{f_s(\theta_e,I)}{\mu\left(\theta_e\right)}d\theta_e. \tag{A32}$$

In addition,

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

Since $\left(\frac{\sigma}{\sigma-1}-\frac{\varepsilon_e}{\varepsilon_e+1}+\xi\right)>0$ and $\frac{d\ln\left[\frac{1}{\mu(\theta_e)}\right]}{d\ln I}>0$, we have

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

On the other hand, notice that

$$\frac{d\ln\left[\frac{1}{\mu(\theta_e)}\right]}{d\ln I} = \left[1 - \frac{\sigma-1}{\sigma}\mu\left(\theta_e\right)\right]\frac{I}{I-1}$$

is decreasing in $\mu\left(\theta_e\right)$, so we now try to prove that

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

To do this, note that by (A32), we have

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

where $f_s(\theta_e, I)\left[\frac{1}{\mu(\theta_e)} - \nu\left(I\right)\right]$ is positive if and only if $\frac{1}{\mu(\theta_e)} - \nu\left(I\right)$ is positive. Define

$$\Omega \equiv \left\{\theta_e | \mu\left(\theta_e\right) < \frac{1}{\nu\left(I\right)}\right\}$$

such that $f_s(\theta_e, I)\left[\frac{1}{\mu(\theta_e)} - \nu\left(I\right)\right] > 0$ if and only if $\theta_e \in \Omega$.

Notice that

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

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

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

Thus,

$$\begin{aligned}
&\int_{\theta_e\in\Omega_*} f_s(\theta_e, I)\left[\frac{1}{\mu\left(\theta_e\right)} - \nu\left(I\right)\right]\frac{d\ln\left[\frac{1}{\mu(\theta_e)}\right]}{d\ln I}d\theta_e \\
\geq\ &\left[1 - \frac{\sigma-1}{\sigma}\nu\left(I\right)\right]\frac{I}{I-1}\int_{\theta_e\in\Omega_*} f_s(\theta_e, I)\left[\frac{1}{\mu\left(\theta_e\right)} - \nu\left(I\right)\right]d\theta_e.
\end{aligned} \tag{A33}$$

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

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

Thus,

$$\int_{\theta_e \notin \Omega_*} f_s(\theta_e, I) \left[\frac{1}{\mu(\theta_e)} - \nu(I)\right] \frac{d \ln\left[\frac{1}{\mu(\theta_e)}\right]}{d \ln I} d\theta_e \tag{A34}$$
$$\geq \left[1 - \frac{\sigma - 1}{\sigma}\nu(I)\right] \frac{I}{I-1} \int_{\theta_e \notin \Omega_*} f_s(\theta_e, I) \left[\frac{1}{\mu(\theta_e)} - \nu(I)\right] d\theta_e.$$

Combining (A33) and (A34) gives

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

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

A.4 Proof of Proposition 2

Assume that markups are constant. (A22) and (A23) give

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

Substituting W and Q in (A12) with the above equations, we have

$$L_w(\theta_e) \propto \left[\frac{1}{\mu}\right]^{\frac{(\sigma-1)(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}} \left[\frac{1}{\mu}\right]^{\frac{\varepsilon_w(\varepsilon_e+1)-\sigma+1}{\varepsilon_w[1-\xi(\varepsilon_e+1)]+1}\frac{\varepsilon_e+1}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}} = \left[\frac{1}{\mu}\right]^{\frac{(\varepsilon_e+1)\varepsilon_w}{\varepsilon_w[1-\xi(\varepsilon_e+1)]+1}}, \forall \theta_e \in \Theta_e.$$

Similarly, we have

$$
\begin{aligned}
S(\theta_e) &\propto \left[\frac{1}{\mu}\right]^{\frac{\varepsilon_e(\varepsilon_w+1)+\varepsilon_w\xi(\varepsilon_e+1)}{\varepsilon_w[1-\xi(\varepsilon_e+1)]+1}}, \\
l_e(\theta_e) &\propto \left[\frac{1}{\mu}\right]^{\frac{(\varepsilon_w+1)\varepsilon_e}{\varepsilon_w[1-\xi(\varepsilon_e+1)]+1}}, \\
Q_{ij}(\theta_e) &\propto \left[\frac{1}{\mu}\right]^{\frac{(\varepsilon_e+1)\varepsilon_w\xi+(\varepsilon_w+1)\varepsilon_e}{\varepsilon_w[1-\xi(\varepsilon_e+1)]+1}}, \\
P(\theta_e) &\propto \left[\frac{1}{\mu}\right]^{0}, \\
c_e(\theta_e) &\propto [\mu-\xi]\left[\frac{1}{\mu}\right]^{\frac{(\varepsilon_w+1)(\varepsilon_e+1)}{\varepsilon_w[1-\xi(\varepsilon_e+1)]+1}}, \\
V_e(\theta_e) &\propto \left[\mu-\xi-\frac{\varepsilon_e}{\varepsilon_e+1}\right]\left[\frac{1}{\mu}\right]^{\frac{(\varepsilon_w+1)(\varepsilon_e+1)}{\varepsilon_w[1-\xi(\varepsilon_e+1)]+1}}, \forall\theta_e \in \Theta_e.
\end{aligned}
$$

It's easy to see that under the conditions (28) and (29), $L_w(\theta_e)$, $S(\theta_e)$, $l_e(\theta_e)$, $Q_{ij}(\theta_e)$, and $P(\theta_e)$ go down with the decrease of I. Moreover, since the markup is uniform, firm-level labor shares must go down too. Changes of $c_e(\theta_e)$ and $V_e(\theta_e)$ are ambiguous.

Notice that

$$\frac{d\ln c(\theta_e)}{d\ln\mu} \geq 0 \Leftrightarrow \frac{\mu-\xi}{\mu} \leq \frac{\varepsilon_w+1-\varepsilon_w\xi\left(\varepsilon_e+1\right)}{\left(\varepsilon_w+1\right)\left(\varepsilon_e+1\right)}.$$

One can see that

$$\mu \leq \frac{\xi}{\frac{\varepsilon_e}{\varepsilon_e+1}+\frac{\varepsilon_w}{\varepsilon_w+1}\xi}$$

is the condition for $\frac{d\ln c_{ij}(\theta_e)}{d\ln\mu} \geq 0$.

On the other hand,

$$\frac{dV_e(\theta_e)}{d\ln\mu} \propto \left[\xi+\frac{\varepsilon_e}{\varepsilon_e+1}\right]-\left[\frac{\varepsilon_w}{1+\varepsilon_w}\xi+\frac{\varepsilon_e}{\varepsilon_e+1}\right]\mu,$$

thus,

$$\mu \leq \frac{\xi+\frac{\varepsilon_e}{\varepsilon_e+1}}{\frac{\varepsilon_e}{\varepsilon_e+1}+\frac{\varepsilon_w}{1+\varepsilon_w}\xi}$$

is a condition for $\frac{dV_e(\theta_e)}{d\ln\mu} \geq 0$.■

A.5 Technology and Equilibrium

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

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

$x_e(\theta_e)^{\frac{\sigma-1}{\sigma}}\chi(\theta_e)$. Specifically, according to (A25) and (A26), we have

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

and

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

On the other hand, given $X(\theta_e)$, we see that $P(\theta_e)Q_{ij}(\theta_e)$, $L_w(\theta_e)$, $l_e(\theta_e)$, and $V_e(\theta_e)$ are independent of the specific values of $\chi(\theta_e)$ and $x_e(\theta_e)$. According to (A12) to (A14), and (A27), we have the following results:

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

and

$$\begin{aligned}
V(\theta_e) &= \left[\mu(\theta_e)-\xi-\frac{\varepsilon_e}{\varepsilon_e+1}\right]\left(\frac{\xi}{W}\right)^{\frac{\xi(\sigma-1)(\varepsilon_e+1)}{\sigma+\varepsilon_e-\xi(\sigma-1)(\varepsilon_e+1)}} \\
&\quad\times\left[\frac{A^{\frac{\sigma-1}{\sigma}}X(\theta_e)}{\mu(\theta_e)}\left(\frac{Q}{N}\right)^{\frac{1}{\sigma}}\right]^{\frac{\sigma(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}}.
\end{aligned}$$

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

Lastly, we find that given $\frac{d\ln X(\theta_e)}{d\theta_e}$, $\frac{V'(\theta_e)}{V(\theta_e)}$ is independent of the specific values of $\chi(\theta_e)$ and $x_e(\theta_e)$. Combining (A28) and (A29) gives

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

Combining (A27) and (A28) gives

$$\begin{aligned}
\frac{V'(\theta_e)}{V(\theta_e)} &= \frac{\mu(\theta_e)\frac{\partial\ln P(Q_{ij}(\theta_e),\theta_e)}{\partial\theta_e}+\frac{x_e'(\theta_e)}{x_e(\theta_e)}}{\mu(\theta_e)-\xi-\frac{\varepsilon_e}{\varepsilon_e+1}} \\
&= \frac{\frac{\sigma(\varepsilon_e+1)[\mu(\theta_e)-\xi]-\sigma\varepsilon_e}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}\frac{d\ln X(\theta_e)}{d\theta_e}+\frac{(\sigma-1)[\varepsilon_e+\xi(\varepsilon_e+1)]\left[\frac{\sigma}{\sigma-1}-\mu(\theta_e)\right]}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}\frac{d\ln\mu(\theta_e)}{d\theta_e}}{\mu(\theta_e)-\xi-\frac{\varepsilon_e}{\varepsilon_e+1}},
\end{aligned} \tag{A35}$$

where the second equation is derived by (A29) and (A28). Specially, when markup is constant, we have

$$\frac{V'(\theta_e)}{V(\theta_e)} = \frac{\sigma\left(\varepsilon_e+1\right)}{\varepsilon_e+\sigma-\xi\left(\sigma-1\right)\left(\varepsilon_e+1\right)}\frac{d\ln X(\theta_e)}{d\theta_e}, \forall\theta_e \in \Theta_e. \tag{A36}$$

A.6 Elasticities

A.6.1 Profit Elasticity

Note that we have assumed that the linear elasticity $\varepsilon_e = \frac{\phi_e'(l_e(\theta))}{l_e(\theta)\phi_e''(l_e(\theta))}$ is constant in the model setup to simplify the notation. We define the non-linear elasticity of profit with respect to the net-tax income rate as (A41). To understand the elasticity, consider the following tax reform.

Consider a small increase (i.e., $d\tau$) in the marginal tax rate faced by the θ_e-type entrepreneur, and suppose that the tax reform has no first-order effects on the aggregate values and the actions of other types. Then, based on the entrepreneur's problem, the optimal choice of the θ_e-type entrepreneur (i.e., L_w and l_e) satisfy the following first-order conditions:

$$WL_w = P\left(Q_{ij},\theta_e\right)Q_{ij}\frac{\xi}{\mu\left(\theta_e\right)},$$

and

$$\begin{aligned}
\phi_e'\left(l_e\right) &= \left[1-T_e'(P\left(Q_{ij},\theta_e\right)Q_{ij}-WL_w)-d\tau\right]\frac{P\left(Q_{ij},\theta_e\right)Q_{ij}}{\mu\left(\theta_e\right)}\frac{1}{l_e(\theta_e)} \\
&= \left[1-T_e'\left(\left(\frac{\mu\left(\theta_e\right)}{\xi}-1\right)WL_w\right)-d\tau\right]\frac{WL_w}{\xi}\frac{1}{l_e},
\end{aligned}$$

where the second equation is derived by $WL_w = P\left(Q_{ij},\theta_e\right)Q_{ij}\frac{\xi}{\mu(\theta_e)}$. Assumption 1 ensures that the first-order condition corresponds to a unique global maximum; thus, we can apply the implicit function group theorem to derive the elasticities of effort and factor demand according to the net profit tax rate.

The θ_e-type entrepreneur treats $\mu\left(\theta_e\right)$, W, and other firms' outputs (i.e., outputs other than $Q_{ij}\left(\theta\right)$) as given. In such a scenario, its reaction to the tax reform can be described by differential equations of the first-order conditions. On the one hand,

$$\begin{aligned}
\phi_e''\left(l_e\right)dl_e &= \left[1-T_e'(\pi)\right]\left[\frac{WdL_w}{\xi}\frac{1}{l_e}-\frac{WL_w}{\xi}\frac{dl_e}{l_e}\frac{1}{l_e}\right] \\
&\quad -\left[T_e''(\pi)\left(\frac{\mu\left(\theta_e\right)}{\xi}-1\right)WdL_w\right]\frac{WL_w}{\xi}\frac{1}{l_e}-d\tau\frac{WL_w}{\xi}\frac{1}{l_e}.
\end{aligned}$$

Divide both sides of the above equation by $\phi_e'\left(l_e\right)$ or $\left[1-T_e'\left(\pi\right)\right]\frac{WL_w}{\xi}\frac{1}{l_e}$, and we have

$$\frac{\phi_e''\left(l_e\right)l_e}{\phi_e'\left(l_e\right)}\frac{dl_e}{l_e} = \left[\frac{dL_w}{L_w}-\frac{dl_e}{l_e}\right]-\left[\frac{\pi T_e''(\pi)}{1-T_e'(\pi)}\frac{dL_w}{L_w}\right]-\frac{d\tau}{1-T_e'(\pi)}.$$

That is,

$$\frac{1+\varepsilon_e}{\varepsilon_e}\frac{dl_e}{l_e}=\frac{dL_w}{L_w}\left[1-\frac{\pi T_e''(\pi)}{1-T_e'(\pi)}\right]-\frac{d\tau}{1-T_e'(\pi)}. \tag{A37}$$

On the other hand, based on $WL_w = P\left(Q_{ij},\theta_e\right)Q_{ij}\frac{\xi}{\mu(\theta_e)}$, we have

$$WdL_w = PQ_{ij}\frac{\xi}{\mu\left(\theta_e\right)^2}\left[\frac{dl_e}{l_e}+\xi\frac{dL_w}{L_w}\right].$$

Dividing both sides of the above equation by WL_w or $PQ_{ij}\frac{\xi}{\mu(\theta_e)}$ gives

$$\frac{dL_w}{L_w}=\frac{1}{\mu\left(\theta_e\right)}\left[\frac{dl_e}{l_e}+\xi\frac{dL_w}{L_w}\right]. \tag{A38}$$

Combining (A37) and (A38) gives

$$\frac{dL_w}{L_w}=\frac{-\frac{d\tau}{1-T_e'(\pi)}}{\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu\left(\theta_e\right)-\xi\right]-\left[1-\frac{\pi T_e''(\pi)}{1-T_e'(\pi)}\right]}. \tag{A39}$$

Lastly, based on $\pi = PQ_{ij}-WL_w$ and $WL_w = PQ_{ij}\frac{\xi}{\mu(\theta_e)}$, we have $d\pi = \frac{P}{\mu(\theta_e)}\frac{dQ}{dL_w}dL_w + \frac{P}{\mu(\theta)}\frac{dQ}{dl_e}dl_e - WdL_w$. Thus,

$$\begin{aligned}\frac{d\pi}{\pi} &= \frac{PQ_{ij}}{\pi}\frac{\xi}{\mu\left(\theta_e\right)}\frac{dL_w}{L_w}+\frac{PQ_{ij}}{\pi}\frac{1}{\mu\left(\theta_e\right)}\frac{dl_e}{l_e}-\frac{WL_w}{\pi}\frac{dL_w}{L_w} \\ &= \left[\frac{\frac{\xi}{\mu(\theta_e)}}{1-\frac{\xi}{\mu(\theta_e)}}-\frac{\frac{\xi}{\mu(\theta_e)}}{1-\frac{\xi}{\mu(\theta_e)}}\right]\frac{dL_w}{L_w}+\frac{\frac{1}{\mu(\theta_e)}}{1-\frac{\xi}{\mu(\theta_e)}}\frac{dl_e}{l_e} \\ &= \frac{\frac{1}{\mu(\theta_e)}}{1-\frac{\xi}{\mu(\theta_e)}}\frac{dl_e}{l_e} \\ &= \frac{dL_w}{L_w}\end{aligned} \tag{A40}$$

where the last equation is derived by (A38). Moreover, we can also obtain the above equation through $WL_w = \left(\pi + WL_w\right)\frac{\xi}{\mu(\theta_e)}$, which is a combination of $\pi = PQ_{ij}-WL_w$ and $WL_w = PQ_{ij}\frac{\xi}{\mu(\theta_e)}$.

Combining (A39) and (A40) gives a non-linear elasticity of profit with respect to the net tax income rate:

$$\widetilde{\varepsilon}^{\pi}_{1-\tau_e}\left(\theta_e\right)\equiv\frac{\frac{d\pi(\theta_e)}{\pi(\theta_e)}}{-\frac{d\tau}{1-T_e'(\pi(\theta_e))}}=\frac{1}{\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu\left(\theta_e\right)-\xi\right]-\left[1-\frac{\pi(\theta_e)T_e''(\pi(\theta_e))}{1-T_e'(\pi(\theta_e))}\right]}. \tag{A41}$$

We thus create a linear elasticity of profit with respect to the net tax income rate as $\varepsilon^{\pi}_{1-\tau_e}\left(\theta_e\right)$ (e.g., see (40)), which is the elasticity of profit when $T_e\left(\cdot\right)$ is linear. Such a linear elasticity of profit can be observed from the data if the profit tax in the real economy is linear.

A.6.2 Price Elasticity

To make the expression more compact, we denote $P\left(Q_{ij},\theta_e\right)$ as the short form of the inverse demand function and $P\left(\theta_e\right)$ as the price. Solving the final good producer's problem, we immediately find the following in the equilibrium for any $\theta_e \in \Theta_e$:

$$P(\theta_e) = \chi(\theta_e) N^{-\frac{1}{\sigma}} A^{\frac{\sigma-1}{\sigma}} Q_{ij}\left(\theta_e\right)^{-\frac{1}{\sigma}} Q^{\frac{1}{\sigma}}, \tag{A42}$$

and the inverse demand function

$$P(Q_{ij},\theta_e) = \chi(\theta_e) N^{-\frac{1}{\sigma}} A^{\frac{\sigma-1}{\sigma}} Q_{ij}^{-\frac{1}{\eta(\theta_e)}} I^{-\left[\frac{1}{\eta(\theta_e)}-\frac{1}{\sigma}\right]\frac{\eta(\theta_e)}{\eta(\theta_e)-1}} \left[\begin{array}{c} \left(I-1\right) Q_{ij}\left(\theta_e\right)^{\frac{\eta(\theta_e)-1}{\eta(\theta_e)}} \\ +Q_{ij}^{\frac{\eta(\theta_e)-1}{\eta(\theta_e)}} \end{array}\right]^{\left[\frac{1}{\eta(\theta_e)}-\frac{1}{\sigma}\right]\frac{\eta(\theta_e)}{\eta(\theta_e)-1}} Q^{\frac{1}{\sigma}}. \tag{A43}$$

For later use, we define the own price elasticity, own inverse-demand elasticity, and cross inverse-demand elasticity as

$$\begin{aligned} \varepsilon^P_{Q_{ij}}\left(\theta_e\right) &\equiv \frac{\partial \ln P\left(\theta_e\right)}{\partial \ln Q_{ij}\left(\theta_e\right)} = -\frac{1}{\sigma}, \qquad \text{(A44)} \\ \varepsilon_{Q_{ij}}(\theta_e) &\equiv \frac{\partial \ln P\left(Q_{ij},\theta_e\right)}{\partial \ln Q_{ij}}|_{Q_{ij}=Q_{ij}(\theta_e)} = -\left[\frac{1}{\eta\left(\theta_e\right)}\frac{I-1}{I} + \frac{1}{\sigma}\frac{1}{I}\right], \\ \varepsilon_{Q_{-ij}}(\theta_e) &\equiv \frac{\partial \ln P\left(Q_{ij},\theta_e\right)}{\partial \ln Q_{ij}\left(\theta_e\right)}|_{Q_{ij}=Q_{ij}(\theta_e)} = \left[\frac{1}{\eta\left(\theta_e\right)} - \frac{1}{\sigma}\right]\frac{I-1}{I}, \forall \theta_e \in \Theta_e. \end{aligned}$$

Notice that $\mu\left(\theta_e\right) = \frac{1}{1+\varepsilon_{Q_{ij}}\left(\theta_e\right)}$, we have

$$\varepsilon_{Q_{-ij}}\left(\theta_e\right) = -\frac{1}{\mu\left(\theta_e\right)} + \frac{\sigma-1}{\sigma}, \ \forall \theta_e \in \Theta_e. \tag{A45}$$

Under our production technology, we have

$$\varepsilon_{Q_{ij}}(\theta_e) = \varepsilon^P_{Q_{ij}}\left(\theta_e\right) - \varepsilon_{Q_{-ij}}(\theta_e),$$

and

$$\begin{aligned} & \frac{\partial \ln P(Q_{ij},\theta_e)}{\partial \theta_e}|_{Q_{ij}=Q_{ij}(\theta_e)} \\ = & \frac{d\ln P(Q_{ij}(\theta_e),\theta_e)}{d\theta_e} - \varepsilon_{Q_{ij}}(\theta_e)\frac{d\ln Q_{ij}(\theta_e)}{d\theta_e} \\ = & \frac{\chi'(\theta_e)}{\chi(\theta_e)} - \frac{1}{\sigma}\frac{Q'_{ij}(\theta_e)}{Q_{ij}(\theta_e)} + \left[\frac{I-1}{I}\frac{1}{\eta(\theta_e)} + \frac{1}{I}\frac{1}{\sigma}\right]\frac{Q'_{ij}(\theta_e)}{Q_{ij}(\theta_e)} \\ = & \frac{\chi'(\theta_e)}{\chi(\theta_e)} + \left[\frac{1}{\eta(\theta_e)} - \frac{1}{\sigma}\right]\frac{I-1}{I}\frac{Q'_{ij}(\theta_e)}{Q_{ij}(\theta_e)} \\ = & \frac{\chi'(\theta_e)}{\chi(\theta_e)} + \varepsilon_{Q_{-ij}}(\theta_e)\frac{d\ln Q_{ij}(\theta_e)}{d\theta_e}, \forall \theta_e \in \Theta_e. \end{aligned} \tag{A46}$$

Specially, when $I=1$, we have $\frac{\partial \ln P\left(Q_{ij},\theta_e\right)}{\partial \theta_e} = \frac{\chi'(\theta_e)}{\chi(\theta_e)}$.

A.6.3 Wage Elasticity

We have defined $\varpi(\theta_e) = \frac{P(\theta_e)}{\mu(\theta_e)}\frac{\partial Q_{ij}(\theta_e)}{\partial L_w(\theta_e)}$. In addition, we define

$$\varepsilon^{\varpi}_{L_w}(\theta_e) = \frac{\partial \ln \varpi(\theta_e)}{\partial \ln L_w(\theta_e)} - \varepsilon^{\varpi}_{L_w}(\theta_e,\theta_e), \text{ and } \varepsilon^{\varpi}_{l_e}(\theta_e) = \frac{\partial \ln \varpi(\theta_e)}{\partial \ln l_e(\theta_e)} - \varepsilon^{\varpi}_{l_e}(\theta_e,\theta_e) \tag{A47}$$

as the own elasticity of wage with respect to labor inputs and effort, respectively, where

$$\varepsilon^{\varpi}_{L_w}(\theta'_e,\theta_e) = \begin{cases} \frac{\partial \ln \varpi(\theta'_e)}{\partial \ln L_w(\theta_e)}, & \theta'_e \neq \theta_e, \\ \lim_{\theta'_e\to\theta_e}\frac{\partial \ln \varpi(\theta'_e)}{\partial \ln L_w(\theta_e)}, & \theta'_e = \theta_e \end{cases} \textit{and} \tag{A48}$$

$$\varepsilon^{\varpi}_{l_e}(\theta'_e,\theta_e) = \begin{cases} \frac{\partial \ln \varpi(\theta'_e)}{\partial \ln L_e(\theta_e)}, & \theta'_e \neq \theta_e, \\ \lim_{\theta'_e\to\theta_e}\frac{\partial \ln \varpi(\theta'_e)}{\partial \ln l_e(\theta_e)}, & \theta'_e = \theta_e; \end{cases} \tag{A49}$$

are the cross elasticity of wage with respect to labor and capital inputs, respectively, $(\theta_e,\theta'_e) \in \Theta_e^2$.

Observe that under the assumptions on the technology, $\varepsilon^{\varpi}_{L_w}(\theta'_e,\theta_e)$ and $\varepsilon^{\varpi}_{l_e}(\theta'_e,\theta_e)$ are not dependent on θ'_e. Specifically, we have

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

Then by definition, we have

$$\varepsilon^{\varpi}_{L_w}(\theta_e) = \xi\left(1-\frac{1}{\sigma}\right) - 1 < 0, \quad \text{and} \quad \varepsilon^{\varpi}_{l_e}(\theta) = 1 - \frac{1}{\sigma} > 0. \tag{A50}$$

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

B Solution

B.1 Proof of Lemma 1

(i) According to the definition of $V_e(\theta_e'|\theta_e)$, we have

$$\frac{\partial V_e(\theta_e'|\theta_e)}{\partial \theta_e'} = c_e'\left(\theta_e'\right) - \phi_e'\left(l_e\left(\theta_e'|\theta_e\right)\right)\frac{\partial l_e\left(\theta_e'|\theta_e\right)}{\partial \theta_e'} \tag{B2}$$

According to the first-order incentive condition, we have $\lim_{\theta_e \to \theta_e'} \frac{\partial V_e(\theta_e'|\theta_e)}{\partial \theta_e'} = 0$. That is,

$$0 = \left[c_e'\left(\theta_e'\right) - \phi_e'\left(l_e\left(\theta_e'|\theta_e\right)\right)\frac{\partial l_e\left(\theta_e'|\theta_e\right)}{\partial \theta_e'}\right]|_{\theta_e=\theta_e'}, \tag{B3}$$

Adding (B2) into (B3), we have

$$\frac{\partial V_e(\theta_e'|\theta_e)}{\partial \theta_e'} = \left[\phi_e'\left(l_e\left(\theta_e'|\theta_e\right)\right)\frac{\partial l_e\left(\theta_e'|\theta_e\right)}{\partial \theta_e'}\right]|_{\theta_e=\theta_e'} - \phi_e'\left(l_e\left(\theta_e'|\theta_e\right)\right)\frac{\partial l_e\left(\theta_e'|\theta_e\right)}{\partial \theta_e'}$$

Using the mean value theorem, the sign of the right-hand side is given by

$$\frac{d\left[\phi_e'\left(l_e\left(\theta_e'|\theta_e^*\right)\right)\frac{\partial l_e(\theta_e'|\theta_e^*)}{\partial \theta_e'}\right]}{d\theta^*}\left(\theta_e' - \theta_e\right)$$

for some θ_e^* that lies between θ_e' and θ_e. If one has $\frac{d\left[\phi_e'\left(l_e\left(\theta_e'|\theta_e^*\right)\right)\frac{\partial l_e(\theta_e'|\theta_e^*)}{\partial \theta_e'}\right]}{d\theta^*} < 0$ for any $(\theta_e^*, \theta_e') \in \Theta^2$, the function $V_e(\theta_e'|\theta_e)$ will increase with θ_e' until $\theta_e' = \theta_e$ and then decreases with θ_e'. In conclusion, there is a unique local maximum point that is also the global maximizer of $V_e(\theta_e'|\theta_e)$. Thus under Assumption 1, the first-order incentive condition is not only necessary but also sufficient for the agent's problem.

Notice that

$$\begin{aligned}
&\frac{d\left[\phi_e'\left(l_e\left(\theta_e'|\theta_e^*\right)\right)\frac{\partial l_e(\theta_e'|\theta_e^*)}{\partial \theta_e'}\right]}{d\theta^*} \\
= \;& \phi_e''\left(l_e\left(\theta_e'|\theta_e^*\right)\right)\frac{\partial l_e\left(\theta_e'|\theta_e^*\right)}{\partial \theta_e^*}\frac{\partial l_e\left(\theta_e'|\theta_e^*\right)}{\partial \theta_e'} + \phi_e'\left(l_e\left(\theta_e'|\theta_e^*\right)\right)\frac{\partial^2 l_e\left(\theta_e'|\theta_e^*\right)}{\partial \theta_e^* \partial \theta_e'} \\
= \;& \phi_e'\left(l_e\left(\theta_e'|\theta_e^*\right)\right) l_e\left(\theta_e'|\theta_e^*\right)\left[\begin{array}{c} \frac{\phi_e''(l_e(\theta_e'|\theta_e^*))l_e(\theta_e'|\theta_e^*)}{\phi_e'(l_e(\theta_e'|\theta_e^*))}\frac{\partial \ln l_e(\theta_e'|\theta_e^*)}{\partial \theta_e^*}\frac{\partial \ln l_e(\theta_e'|\theta_e^*)}{\partial \theta_e'} \\ + \frac{\partial^2 \ln l_e(\theta_e'|\theta_e^*)}{\partial \theta_e^* \partial \theta_e'}, \end{array}\right]
\end{aligned}$$

so we have

$$\operatorname{sign}\left(\frac{d\left[\phi_e'\left(l_e\left(\theta_e'|\theta_e^*\right)\right)\frac{\partial l_e(\theta_e'|\theta_e^*)}{\partial \theta_e'}\right]}{d\theta^*}\right) = \operatorname{sign}\left(\begin{array}{c} \frac{\phi_e''(l_e(\theta_e'|\theta_e^*))l_e(\theta_e'|\theta_e^*)}{\phi_e'(l_e(\theta_e'|\theta_e^*))}\frac{\partial \ln l_e(\theta_e'|\theta_e^*)}{\partial \theta_e^*}\frac{\partial \ln l_e(\theta_e'|\theta_e^*)}{\partial \theta_e'} \\ + \frac{\partial^2 \ln l_e(\theta_e'|\theta_e^*)}{\partial \theta_e^* \partial \theta_e'} \end{array}\right)$$

Since $\frac{\phi_e''(l_e(\theta_e'|\theta_e^*))l_e(\theta_e'|\theta_e^*)}{\phi_e'(l_e(\theta_e'|\theta_e^*))}$ is positive, it follows that $\frac{\partial \ln l_e(\theta_e'|\theta_e^*)}{\partial \theta_e^*}\frac{\partial \ln l_e(\theta_e'|\theta_e^*)}{\partial \theta_e'} < 0$ and $\frac{\partial^2 \ln l_e(\theta_e'|\theta_e^*)}{\partial \theta_e^* \partial \theta_e'} = 0$ is a sufficient condition for $\frac{d\left[\phi_e'(l_e(\theta_e'|\theta_e^*))\frac{\partial l_e(\theta_e'|\theta_e^*)}{\partial \theta_e'}\right]}{d\theta^*} < 0$.

$l_e\left(\theta_e'|\theta_e\right)$ is determined by

$$P\left(Q_{ij}\left(x_e\left(\theta_e\right)l_e\left(\theta_e'|\theta_e\right), L_w\left(\theta_e'|\theta_e\right)\right), \theta_e\right)Q_{ij}\left(x_e\left(\theta_e\right)l_e\left(\theta_e'|\theta_e\right), L_w\left(\theta_e'|\theta_e\right)\right)(1-t_s) - WL_w\left(\theta_e'|\theta_e\right) = \pi\left(\theta_e'\right), \tag{B4}$$

where $L_w\left(\theta_e'|\theta_e\right)$ is the optimal labor input given that θ_e entrepreneur reports θ_e'. The inverse demand function $P\left(Q_{ij}, \theta_e\right)$ is given by (A43). In the following proof of part (i), we refer to $P\left(Q_{ij}, \theta_e\right)$ and Q_{ij} as the short forms of $P\left(Q_{ij}\left(x_e\left(\theta_e\right)l_e\left(\theta_e'|\theta_e\right), L_w\left(\theta_e'|\theta_e\right)\right), \theta_e\right)$ and $Q_{ij}\left(x_e\left(\theta_e\right)l_e\left(\theta_e'|\theta_e\right), L_w\left(\theta_e'|\theta_e\right)\right)$, respectively. Based on (B4), we have

$$\begin{aligned} \frac{\partial l_e\left(\theta_e'|\theta_e\right)}{\partial \theta_e} &= -\frac{\frac{\partial\left[P\left(Q_{ij},\theta_e\right)Q_{ij}\right]}{\partial Q_{ij}}\frac{\partial Q_{ij}}{\partial \theta_e} + \frac{\partial P\left(Q_{ij},\theta_e\right)}{\partial \theta_e}Q_{ij}}{\frac{\partial\left[P\left(Q_{ij},\theta_e\right)Q_{ij}\right]}{\partial Q_{ij}}\frac{\partial Q_{ij}}{\partial l_e}} \\ &= -\frac{\frac{\partial Q_{ij}}{\partial \theta_e} + \frac{\frac{\partial P\left(Q_{ij},\theta_e\right)}{\partial \theta_e}Q_{ij}}{P\left(Q_{ij},\theta_e\right)\left[1+\varepsilon_{Q_{ij}}(\theta_e)\right]}}{\frac{\partial Q_{ij}}{\partial l_e}} \\ &= -\frac{x_e'\left(\theta_e\right)}{x_e\left(\theta_e\right)}l_e\left(\theta_e'|\theta_e\right) - \frac{\partial \ln P\left(Q_{ij},\theta_e\right)}{\partial \theta_e}\frac{l_e\left(\theta_e'|\theta_e\right)}{1+\varepsilon_{Q_{ij}}\left(\theta_e\right)} < 0 \end{aligned} \tag{B5}$$

and

$$\begin{aligned} \frac{\partial l_e\left(\theta_e'|\theta_e\right)}{\partial \theta_e'} &= -\frac{-\pi'\left(\theta_e'\right)}{\frac{\partial\left[P\left(Q_{ij},\theta_e\right)Q_{ij}\right]}{\partial Q_{ij}}\frac{\partial Q_{ij}}{\partial l_e(\theta_e'|\theta_e)}} \\ &= -\frac{-\pi'\left(\theta_e'\right)}{\frac{\partial \ln\left[P\left(Q_{ij},\theta_e\right)Q_{ij}\right]}{\partial \ln Q_{ij}}\frac{\partial Q_{ij}}{\partial l_e(\theta_e'|\theta_e)}\frac{l_e(\theta_e'|\theta_e)}{Q_{ij}}\frac{P\left(Q_{ij},\theta_e\right)}{l_e(\theta_e'|\theta_e)}} \\ &= -\frac{-\pi'\left(\theta_e'\right)}{\left[1+\varepsilon_{Q_{-ij}}(\theta_e)\right]\frac{P\left(Q_{ij},\theta_e\right)}{l_e(\theta_e'|\theta_e)}} > 0. \end{aligned} \tag{B6}$$

In addition, we have

$$\frac{\partial \ln l_e\left(\theta_e'|\theta_e\right)}{\partial \theta_e} = -\frac{x_e'\left(\theta_e\right)}{x_e\left(\theta_e\right)} - \frac{\partial \ln P\left(Q_{ij}\left(x_e\left(\theta_e\right)l_e\left(\theta_e'|\theta_e\right), L_w\left(\theta_e'|\theta_e\right)\right), \theta_e\right)}{\partial \theta_e}\frac{1}{1+\varepsilon_{Q_{ij}}\left(\theta_e\right)} < 0 \tag{B7}$$

and

$$\frac{\partial^2 \ln l_e\left(\theta_e'|\theta_e\right)}{\partial \theta_e \partial \theta_e'} = -\frac{\partial^2 \ln P\left(Q_{ij},\theta_e\right)}{\partial \theta_e \partial Q_{ij}}\frac{\partial Q_{ij}}{\partial \theta_e'}\frac{1}{1+\varepsilon^P_{Q_{ij}}\left(\theta_e\right)}. \tag{B8}$$

According to (A46),

$$\frac{\partial \ln P\left(Q_{ij},\theta_e\right)}{\partial \theta_e}=\frac{\widetilde{\chi}'(\theta_e)}{\widetilde{\chi}(\theta_e)}-\frac{1}{\sigma}\frac{f_e'(\theta_e)}{f_e(\theta_e)}+\left[\frac{\sigma-1}{\sigma}-\frac{1}{\mu\left(\theta_e\right)}\right]\frac{Q_{ij}'\left(\theta_e\right)}{Q_{ij}\left(\theta_e\right)}, \tag{B9}$$

$\frac{\partial \ln P\left(Q_{ij},\theta_e\right)}{\partial \theta_e}$ is independent of Q_{ij} (note that the $\frac{Q_{ij}'(\theta_e)}{Q_{ij}(\theta_e)}$ on the right side of the above equation is treated as given by the agents when they report their types). Thus, we have $\frac{\partial^2 \ln P\left(Q_{ij},\theta\right)}{\partial\theta\partial Q_{ij}}=0$ and $\frac{\partial^2 \ln l_e(\theta'|\theta)}{\partial\theta\partial\theta'}=0$. In conclusion, we have $\frac{d\left[\phi_e'(l_e(\theta'|\theta^*))\frac{\partial l_e(\theta'|\theta^*)}{\partial\theta'}\right]}{d\theta^*}<0$.

(ii) Now we prove part (ii) of Lemma 1 (i.e., given (B10), (B11) is satisfied if and only if (35) is satisfied). According to the definition of $V_e(\theta)$, we have

$$V_e(\theta_e)=c_e\left(\theta_e\right)-\phi_e\left(l_e(\theta_e)\right).\forall\theta_e\in\Theta_e \tag{B10}$$

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 θ_e entrepreneur needs to finish the θ_e' task. 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

$$0=\left[c_e'\left(\theta_e'\right)-\phi_e'\left(l_e\left(\theta_e'|\theta_e\right)\right)\frac{\partial l_e\left(\theta'|\theta\right)}{\partial\theta'}\right]|_{\theta'=\theta},\forall\theta_e\in\Theta_e. \tag{B11}$$

First, note that by

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

we have

$$V_e'(\theta_e)=\frac{\partial V_e(\theta_e^*\left(\theta_e\right)|\theta_e)}{\partial\theta_e^*\left(\theta_e\right)}\frac{d\theta_e^*\left(\theta_e\right)}{d\theta_e}+\frac{\partial V_e(\theta_e^*\left(\theta_e\right)|\theta_e)}{\partial\theta_e} \tag{B12}$$

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

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

$$\frac{\partial V_e(\theta_e^*\left(\theta_e\right)|\theta_e)}{\partial\theta_e}=-\phi_e'\left(l_e\left(\theta_e^*\left(\theta_e\right)|\theta\right)\right)\frac{\partial l_e\left(\theta_e^*\left(\theta_e\right)|\theta_e\right)}{\partial\theta_e}, \tag{B13}$$

where by (B5), we have

$$\begin{aligned}\frac{\partial l_e\left(\theta_e^*\left(\theta_e\right)|\theta_e\right)}{\partial\theta_e} &= -\frac{x_e'\left(\theta_e\right)}{x_e\left(\theta_e\right)}l_e\left(\theta_e^*\left(\theta_e\right)|\theta_e\right)\\ &\quad -\frac{\partial \ln P\left(Q_{ij}\left(x_e\left(\theta_e\right)l_e\left(\theta_e^*\left(\theta_e\right)|\theta_e\right),L_w\left(\theta_e^*\left(\theta_e\right)|\theta_e\right)\right),\theta_e\right)}{\partial\theta_e}\frac{l_e\left(\theta_e^*\left(\theta_e\right)|\theta_e\right)}{1+\varepsilon_{Q_{-ij}}(\theta_e)}.\end{aligned} \tag{B14}$$

Last, a combination of (B12), (B13), and (B14) suggests that

$$V_e'(\theta_e) = \phi_e'\left(l_e\left(\theta_e^*\left(\theta_e\right)|\theta_e\right)\right) l_e\left(\theta_e^*\left(\theta_e\right)|\theta_e\right) \left[\begin{array}{c} \frac{x_e'(\theta_e)}{x_e(\theta_e)} + \\ \mu\left(\theta_e\right) \frac{\partial \ln P\left(Q_{ij}(x_e(\theta_e) l_e(\theta_e^*(\theta_e)|\theta_e), L_w(\theta_e^*(\theta_e)|\theta_e)), \theta_e\right)}{\partial \theta_e} \end{array} \right] \tag{B15}$$

if and only if $\frac{\partial V_e(\theta_e^*(\theta_e)|\theta_e)}{\partial \theta_e^*(\theta_e)} \frac{d\theta_e^*(\theta_e)}{d\theta_e} = 0$, which means that when the first-order incentive condition ($\frac{\partial V_e(\theta_e'|\theta_e)}{\partial \theta_e'}|_{\theta_e'=\theta_e} = 0$) is satisfied, we have (35); and if (35) holds (i.e., (B15) holds at $\theta_e^*\left(\theta_e\right) = \theta_e$), we must have $\frac{\partial V_e(\theta_e'|\theta_e)}{\partial \theta_e'}|_{\theta_e'=\theta_e} = 0$ (unless $\frac{d\theta_e^*(\theta_e)}{d\theta_e} = 0$, which is ruled out by Assumption 1).■

B.2 Proof of Lemma 2

We first show that a symmetric Cournot competitive tax equilibrium must satisfy parts 1 to 3. First, by the definition of SCCTE, (16) to (17) and (13) to (15) must be satisfied. Second, by the definition of SCCTE, agents maximize their utilities, which means (33) and (35) should be satisfied (e.g., see Lemma 1).

Next, suppose that we are given an allocation $\mathcal{A}$ and price $\mathcal{P}$ to satisfy the properties in parts 1 to 3. We now construct the tax system $\mathcal{T}$ (with $t_s = 0$), which together with the given allocation $\mathcal{A}$ and price $\mathcal{P}$ constructs an SCCTE. We first construct a policy system with the given allocation $\mathcal{A}$ and price $\mathcal{P}$. We then show that this constructed policy system together with $\mathcal{A}$ and $\mathcal{P}$ constructs an SCCTE.

First, we construct the policy system. By the definition of tax wedges and $t_s = 0$, the marginal tax rates are constructed as follows:

$$T_w'\left(y\left(\theta_w\right)\right) = 1 - \frac{\phi_w'\left(l_w\left(\theta_w\right)\right)}{\frac{P(\theta_e)}{\mu(\theta_e)} \frac{\partial Q_{ij}(\theta_e)}{\partial L_w(\theta_e)} x_w\left(\theta_w\right)}$$

and

$$T_e'\left(\pi\left(\theta_e\right)\right) = 1 - \frac{\phi_e'\left(l_e\left(\theta_e\right)\right)}{\frac{P(\theta_e)}{\mu(\theta_e)} \frac{\partial Q_{ij}(\theta_e)}{\partial l_e(\theta_e)}}.$$

We use agents' budget constraints to fix the labor income taxes. We first construct $T_w\left(\cdot\right)$. To do this, we substitute

$$T_w\left(y(\theta_w)\right) = T_w\left(y(\underline{\theta}_w)\right) + \int_{y(\underline{\theta}_w)}^{y(\theta_w)} T_w'\left(y\right) dy$$

into

$$y(\theta_w) - T_w\left(y(\theta_w)\right) - c_w(\theta_w) = 0$$

and show that there exists $T_w\left(y(\theta_w)\right)$ such that given allocation $\mathcal{A}$, price $\mathcal{P}$ and $\{T_w'\left(y(\theta_w)\right)\}_{\theta_w \in \Theta_w}$, the above equation is satisfied for any $\theta_w \in \Theta_w$.

To be consistent with the $\underline{\theta}_w$-type agent's budget constraint, $T_w\left(y(\underline{\theta}_w)\right)$ must satisfy

$$y(\underline{\theta}_w) - T_w\left(y(\underline{\theta}_w)\right) - c_w(\underline{\theta}_w) = 0.$$

We should show this $T_w\left(y(\underline{\theta})\right)$ is also consistent with other agents' budget constraints. This is equivalent to say that

$$y'(\theta_w)\left[1-T_w'\left(y(\theta_w)\right)\right]-c_w'(\theta_w)=0.$$

Substituting $1-T_w'\left(y(\theta_w)\right)$ with the FOC (18), the above equation is equivalent to

$$c_w'(\theta_w)-\frac{\phi_w'\left(l_w(\theta_w)\right)}{Wx_w\left(\theta_w\right)}y'(\theta_w)=0.$$

The above equations are true since we have

$$\begin{aligned} V_w'(\theta_w) &= \frac{\phi_w'\left(l_w(\theta_w)\right)l_w(\theta_w)x_w'\left(\theta_w\right)}{x_w\left(\theta_w\right)} \\ &= c_w'(\theta_w)-\frac{y'(\theta_w)}{Wx_w\left(\theta_w\right)}\phi_w'\left(l_w(\theta_w)\right)+\frac{\phi_w'\left(l_w(\theta_w)\right)l_w(\theta_w)x_w'\left(\theta_w\right)}{x_w\left(\theta_w\right)}. \end{aligned} \tag{B16}$$

The first equation of (B16) is the incentive condition, and the second equation is derived through the definition of $V_w(\theta_w)$. In conclusion, given the allocation, we can construct a unique labor income tax that is consistent with the allocation in the equilibrium.

The construction of $T_e(\cdot)$ is similar to the construction of $T_w(\cdot)$. Note that $T_w\left(y(\underline{\theta}_w)\right)$ can be different from $T_e\left(\pi(\underline{\theta}_e)\right)$. We substitute

$$T_e\left(\pi(\theta_e)\right)=T_e\left(\pi(\underline{\theta}_e)\right)+\int_{\pi(\underline{\theta}_e)}^{\pi(\theta_e)}T_e'\left(\pi\right)d\pi$$

into

$$\pi(\theta_e)-T_e\left(\pi(\theta_e)\right)-c_e(\theta_e)=0$$

and show that there exists $T_e\left(\pi(\theta_e)\right)$ such that given allocation $\mathcal{A}$, price $\mathcal{P}$ and $\{T_e'\left(\pi(\theta_e)\right)\}_{\theta_e\in\Theta_e}$, the above equation is satisfied for any $\theta_e\in\Theta_e$:

To be consistent with the $\underline{\theta}_e$-type agent's budget constraint, $T_e\left(\pi(\underline{\theta}_e)\right)$ must satisfy

$$\pi(\underline{\theta}_e)-T_e\left(\pi(\underline{\theta}_e)\right)-c_e(\underline{\theta}_e)=0.$$

We should show this $T_e\left(\pi(\underline{\theta}_e)\right)$ is also consistent with other agents' budget constraints. This is equivalent to saying that

$$\pi'(\theta_e)\left[1-T_e'\left(\pi(\theta_e)\right)\right]-c_e'(\theta_e)=0.$$

Substituting $1-T_e'\left(\pi(\theta_e)\right)$ with the FOC (19), the above equation is equivalent to

$$c_e'(\theta_e)-\frac{\phi_e'\left(l_e\left(\theta_e\right)\right)}{\frac{P(\theta_e)}{\mu(\theta_e)}\frac{\partial Q_{ij}(\theta_e)}{\partial l_e(\theta_e)}}\pi'(\theta_e)=0,$$

which is further equivalent to

$$c_e'(\theta_e) - \mu(\theta_e) \frac{\phi_e'(l_e(\theta_e))\, l_e(\theta_e)}{P(\theta_e)\, Q_{ij}(\theta_e)} \pi'(\theta_e) = 0. \tag{B17}$$

The above equations are true since we have

$$\begin{aligned} V_e'(\theta_e) &= \phi_e'(l_e(\theta_e))\, l_e(\theta_e) \left[\mu(\theta_e) \frac{\partial \ln P\left(Q_{ij}(\theta_e), \theta_e\right)}{\partial \theta_e} + \frac{\varkappa_e'(\theta_e)}{\varkappa_e(\theta_e)} \right] \\ &= c_e'(\theta_e) - \phi_e'(l_e(\theta_e))\, l_e(\theta_e) \frac{l_e'(\theta_e)}{l_e(\theta_e)}, \end{aligned} \tag{B18}$$

and

$$\pi'(\theta_e) = Q_{ij}(\theta_e) P\left(Q_{ij}(\theta_e), \theta_e\right) \left[\begin{array}{c} \frac{\partial \ln P\left(Q_{ij}(\theta_e), \theta_e\right)}{\partial \theta_e} + \\ \left[1 + \frac{\partial \ln P\left(Q_{ij}(\theta_e), \theta_e\right)}{\partial \ln Q_{ij}(\theta_e)}\right] \left[\frac{x_e'(\theta_e)}{x_e(\theta_e)} + \frac{l_e'(\theta_e)}{l_e(\theta_e)}\right] \end{array} \right]. \tag{B19}$$

Substituting $\frac{l_e'(\theta_e)}{l_e(\theta_e)}$ in (B18) by (B19) and using $1 + \frac{\partial \ln P\left(Q_{ij}(\theta_e), \theta_e\right)}{\partial \ln Q_{ij}(\theta_e)} = \frac{1}{\mu(\theta_e)}$ delivers (B17) immediately. The first equation of (B18) is the incentive condition, and the second equation is derived through the definition of $V_e(\theta_e)$. (B19) is derived from the definition of $\pi(\theta_e)$ (i.e., $\pi(\theta_e) = P\left(Q_{ij}(\theta_e), \theta_e\right) Q_{ij}(\theta_e) - WL_w(\theta_e)$) and the fact that the derivative of $\pi(\theta_e)$ with respect to $L_w(\theta_e)$ is zero.

In conclusion, given the allocation, we can construct a unique combination of labor income tax and profit tax that is consistent with the allocation in the equilibrium.

We now show that the allocation $\mathcal{A}$ and price $\mathcal{P}$ satisfying parts 1 to 3 and the constructed tax system $\mathcal{T}$ construct an SCCTE. First, the allocation satisfies the incentive conditions (33) and (35). Thus, according to the analysis in the subsections given before (see Lemma 1 for example), the allocation is consistent with agents' optimal choice. Second, the price $\mathcal{P}$ satisfies (16) and (17). Third, the market clear conditions (13) to (15) are satisfied. Lastly, agents' budget constraints (9) and (12) are embedded in the definitions of gross utilities and the construction of income taxes. In conclusion, the constructed tax system $\mathcal{T}$ together with the given allocation $\mathcal{A}$ and price $\mathcal{P}$ constructs an SCCTE.■

C Benchmark Results

C.1 Optimal Taxation

C.1.1 Lagrangian and First-order Conditions

We now take Lagrange multipliers to solve the planner's optimization problem.[44] The Lagrangian function for the planner's problem is

$$
\begin{aligned}
& \pounds\left(L_w, l_w, l_e, V_w, V_e, \delta, \Delta; \lambda, \psi_w, \psi_e\right) \\
= & \sum_{o\in\{w,e\}} N_o \int_{\theta_o} G\left(V_o(\theta_o)\right) \widetilde{f}_o\left(\theta_o\right) d\theta_o + \lambda\left[Q - \sum_{o\in\{w,e\}} N_o \int_{\theta_o} \left[V_o\left(\theta_o\right) + \phi_o\left(l_o\left(\theta_o\right)\right)\right] f_o\left(\theta_o\right) d\theta_o - R\right] \\
& + \lambda'\left[\int_{\theta_w} x_w\left(\theta_w\right) l_w\left(\theta_w\right) f_w\left(\theta_w\right) d\theta_w - N \int_{\theta_e} L_w\left(\theta_e\right) f_e\left(\theta_e\right) d\theta_e\right] \\
& + \int_{\theta_e} \varphi\left(\theta_e\right) \frac{d \ln \varpi\left(\theta_e, \theta_e l_e\left(\theta_e\right), L_w\left(\theta_e\right), Q\right)}{d\theta_e} d\theta_e \\
& + \int_{\theta_e} \kappa\left(\theta_e\right)\left[\delta\left(\theta_e\right) - \frac{d \ln Q_{ij}\left(\theta_e\right)}{d\theta_e}\right] d\theta_e \\
& + \int_{\theta_w} \psi_w\left(\theta_w\right)\left[l_w\left(\theta_w\right) \phi_w'\left(l_w\left(\theta_w\right)\right) \frac{x_w'\left(\theta_e\right)}{x_w\left(\theta_e\right)} - V_w'(\theta_w)\right] d\theta_w \\
& + \int_{\theta_e} \psi_e\left(\theta_e\right)\left[\phi_e'\left(l_e\left(\theta_e\right)\right) l_e\left(\theta_e\right)\left[\mu(\theta_e)\left(\frac{\chi'\left(\theta_e\right)}{\chi\left(\theta_e\right)} - \frac{1}{\sigma}\frac{f_e'(\theta_e)}{f_e(\theta_e)} + \varepsilon_{Q_{-ij}}\left(\theta\right)\delta\left(\theta_e\right)\right) + \frac{x_e'\left(\theta_e\right)}{x_e\left(\theta_e\right)}\right] - V_e'(\theta_e)\right] d\theta_e,
\end{aligned}
$$

where $\frac{\chi'(\theta_e)}{\chi(\theta_e)} - \frac{1}{\sigma}\frac{f_e'(\theta_e)}{f_e(\theta_e)} + \varepsilon_{Q_{-ij}}(\theta_e)\delta\left(\theta_e\right) = \frac{\partial \ln P\left(Q_{ij}(\theta_e),\theta_e\right)}{\partial\theta_e}$. Note that we have introduced $\delta\left(\theta_e\right) = \frac{d\ln Q_{ij}(\theta_e)}{d\theta_e}$ as a control value and that $\ln Q_{ij}\left(\theta_e\right)$ can be treated as a state variable. Constraint $\frac{d\ln\varpi(\theta_e,\theta_e l_e(\theta_e),L_w(\theta_e),Q)}{d\theta_e} = 0$ is used to guarantee that $\varpi\left(\theta_e\right) = \frac{P(\theta_e)}{\mu(\theta_e)}\frac{\partial Q_{ij}(\theta_e)}{\partial L_w(\theta_e)}$ is constant, which is a result of uniform sales taxes on the goods produced by firms.

Taking partial integrals yields the following:

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

and

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

and

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

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

$$
\kappa(\underline{\theta}_e) = \kappa(\overline{\theta}_e) = \varphi\left(\overline{\theta}_e\right) = \varphi\left(\underline{\theta}_e\right) = \psi_o(\underline{\theta}_o) = \psi_o(\overline{\theta}_o) = 0, o \in \{w, e\}. \tag{C2}
$$

[44] 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.

Thus,

$$\int_{\theta_e} \varphi'(\theta_e)\, d\theta_e = 0, \tag{C3}$$

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

$$\frac{\partial £}{\partial V_o(\theta_o)} = G'(V_o(\theta_o))N_o\widetilde{f}_o(\theta_o) + \psi_o'(\theta_o) - \lambda N_o f_o(\theta_o) = 0, o \in \{w, e\}, \tag{C4}$$

$$\frac{\partial £}{\partial \delta(\theta_e)} = \kappa(\theta_e) + \psi_e(\theta_e)\,\phi_e'(l_e(\theta_e))\, l_e(\theta_e)\,\mu(\theta_e)\varepsilon_{Q_{-ij}}(\theta_e) = 0, \tag{C5}$$

$$\frac{\partial £}{\partial l_w(\theta_w)} = -\lambda N_w \phi_w'(l_w(\theta_w))\, f_w(\theta_w) + \lambda' N_w x_w(\theta_w)\, f_w(\theta_w) + \psi_w(\theta_w)\,\frac{\phi_w'(l_w(\theta_w))}{x_w(\theta_w)}\frac{1+\varepsilon_w}{\varepsilon_w} = 0, \tag{C6}$$

$$\frac{\partial £}{\partial L_w(\theta_e)} = \left[\lambda P(\theta_e)\,\frac{\partial Q_{ij}(\theta_e)}{\partial L_w(\theta_e)} - \lambda'\right] N f_e(\theta_e) + \left[\begin{array}{c} \frac{\kappa'(\theta_e)}{L_w(\theta_e)}\frac{\partial \ln Q_{ij}(\theta_e)}{\partial \ln L_w(\theta_e)} \\ -\frac{\int_{\Theta_e}\varphi'(\theta_e')\frac{\partial \ln \varpi(\theta_e')}{\partial \ln L_w(\theta_e)}d\theta_e'}{L_w(\theta_e)} \end{array}\right] = 0, \tag{C7}$$

and

$$\begin{aligned} \frac{\partial £}{\partial l_e(\theta_e)} \quad = \quad & \psi_e(\theta_e)\,\phi_e'(l_e(\theta_e))\,\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu(\theta_e)\,\frac{\partial \ln P(Q(\theta_e),\theta_e)}{\partial \theta_e} + \frac{x_e'(\theta_e)}{x_e(\theta_e)}\right] \\ & +\lambda\left[P(\theta_e)\,\frac{\partial Q_{ij}(\theta_e)}{\partial l_e(\theta_e)} - \phi_e'(l_e(\theta_e))\right] N f_e(\theta_e) \\ & +\frac{\kappa'(\theta_e)}{l_e(\theta_e)}\frac{\partial \ln Q_{ij}(\theta_e)}{\partial \ln l_e(\theta_e)} - \frac{\int_\Theta \varphi'(\theta_e')\,\frac{\partial \ln \varpi(\theta_e')}{\partial \ln l_e(\theta_e)}d\theta_e'}{l_e(\theta_e)} = 0, \forall \theta_o \in \Theta_o. \end{aligned} \tag{C8}$$

C.1.2 Social Welfare Weight

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

$$\lambda = \int_{\theta_o} G'(V_o(\theta_o))\widetilde{f}_o(\theta_o)d\theta_o. \tag{C9}$$

Set

$$g_o(\theta_o) = \frac{G'(V_o(\theta_o))\widetilde{f}_o(\theta_o)}{\lambda f_o(\theta_o)} \tag{C10}$$

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

$$\bar{g}_o(\theta_o) = \frac{\int_{\theta_o}^{\overline{\theta}_o} g(x) f_o(x)\, dx}{1 - F_o(\theta_o)} \tag{C11}$$

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

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

$$\frac{\psi_o'(\theta_o)}{\lambda N_o f_o(\theta_o)} = 1 - g_o(\theta_o) \tag{C12}$$

Taking integration and using the boundary conditions gives

$$-\frac{\psi_o(\theta_o)}{\lambda N_o} = \int_{\theta_o}^{\overline{\theta}_o} [1 - g_o(x)] f_o(x) dx = [1 - \bar{g}_o(\theta_o)] [1 - F_o(\theta_o)] . \tag{C13}$$

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

$$\begin{aligned} \kappa(\theta_e) &= -\psi_e(\theta_e) \phi_e'(l_e(\theta_e)) l_e(\theta_e) \mu(\theta_e) \varepsilon_{Q_{-ij}}(\theta_e) \\ &= -\psi_e(\theta_e) P(\theta_e) Q_{ij}(\theta_e) [1 - \tau_e(\theta_e)] (1 - \tau_s) \varepsilon_{Q_{-ij}}(\theta_e), \end{aligned} \tag{C14}$$

where the second equation is derived by

$$\begin{aligned} \phi_e'(l_e(\theta_e)) l_e(\theta_e) &= \frac{\phi_e'(l_e(\theta_e)) l_e(\theta_e)}{\frac{\partial Q_{ij}(\theta_e)}{\partial L_e(\theta_e)} \frac{L_e(\theta_e)}{Q_{ij}(\theta_e)}} \\ &= \frac{\phi_e'(l_e(\theta_e))}{\frac{P(\theta_e)}{\mu(\theta_e)} \frac{\partial Q_{ij}(\theta_e)}{\partial L_e(\theta_e)} x_e(\theta_e) \frac{1}{Q_{ij}(\theta_e)} \frac{\mu(\theta_e)}{P(\theta_e)}} \\ &= \frac{P(\theta_e) Q_{ij}(\theta_e)}{\mu(\theta_e)} [1 - \tau_e(\theta_e)] (1 - \tau_s) . \end{aligned} \tag{C15}$$

In addition, we have

$$\begin{aligned} \kappa'(\theta_e) &= -\frac{d\left[\psi_e(\theta_e) \phi_e'(l_e(\theta_e)) l_e(\theta_e) \mu(\theta_e) \varepsilon_{Q_{-ij}}(\theta_e)\right]}{d\theta_e} \\ &= -\left[\begin{array}{c} \psi_e'(\theta_e) \phi_e'(l_e(\theta_e)) l_e(\theta_e) \mu(\theta_e) \varepsilon_{Q_{-ij}}(\theta_e) + \\ \psi_e(\theta_e) \phi_e'(l_e(\theta_e)) \frac{1+\varepsilon_e}{\varepsilon_e} l_e'(\theta_e) \mu(\theta_e) \varepsilon_{Q_{-ij}}(\theta_e) + \\ \psi_e(\theta_e) \phi_e'(l_e(\theta_e)) l_e(\theta_e) \frac{d \ln\left[\mu(\theta_e) \varepsilon_{Q_{-ij}}(\theta_e)\right]}{d\theta_e} \end{array} \right] \\ &= -\phi_e'(l_e(\theta_e)) l_e(\theta_e) \mu(\theta_e) \varepsilon_{Q_{-ij}}(\theta_e) \left[\begin{array}{c} \psi_e(\theta_e) \frac{1+\varepsilon_e}{\varepsilon_e} \frac{l_e'(\theta_e)}{l_e(\theta_e)} + \\ \psi_e'(\theta_e) + \psi_e(\theta_e) \frac{d \ln\left[\mu(\theta_e) \varepsilon_{Q_{-ij}}(\theta_e)\right]}{d\theta_e} \end{array} \right] \\ &= -P(\theta_e) Q_{ij}(\theta_e) [1 - \tau_e(\theta_e)] (1 - \tau_s) \varepsilon_{Q_{-ij}}(\theta_e) \left[\begin{array}{c} \psi_e(\theta_e) \frac{1+\varepsilon_e}{\varepsilon_e} \frac{l_e'(\theta_e)}{l_e(\theta_e)} + \\ \psi_e'(\theta_e) + \psi_e(\theta_e) \frac{d \ln\left[\mu(\theta_e) \varepsilon_{Q_{-ij}}(\theta_e)\right]}{d\theta_e} \end{array} \right] . \end{aligned} \tag{C16}$$

Substituting $\psi_e(\theta_e)$ and $\psi_e'(\theta_e)$ in (C16) and (C14) with (C12) and (C13), we have

$$\begin{aligned} \kappa(\theta_e) &= \lambda N_e [1 - \bar{g}_e(\theta_e)] [1 - F_e(\theta_e)] \phi_e'(l_e(\theta_e)) l_e(\theta_e) \mu(\theta_e) \varepsilon_{Q_{-ij}}(\theta_e) \\ &= \lambda N_e [1 - \bar{g}_e(\theta_e)] [1 - F_e(\theta_e)] P(\theta_e) Q_{ij}(\theta_e) [1 - \tau_e(\theta_e)] (1 - \tau_s) \varepsilon_{Q_{-ij}}(\theta_e), \end{aligned} \tag{C17}$$

and

$$\frac{\kappa'(\theta_e)}{\lambda N_e f_e(\theta_e)} = -P(\theta_e) Q_{ij}(\theta_e) [1-\tau_e(\theta_e)] (1-\tau_s) \varepsilon_{Q_{-ij}}(\theta_e) \left[\begin{array}{c} [1-g_e(\theta_e)] - \frac{[1-\bar{g}_e(\theta_e)][1-F_e(\theta_e)]}{f_e(\theta_e)} \\ \times \left[\begin{array}{c} \frac{1+\varepsilon_e}{\varepsilon_e} \frac{l_e'(\theta_e)}{l_e(\theta_e)} + \\ \frac{d \ln\left[\mu(\theta_e)\varepsilon_{Q_{-ij}}(\theta_e)\right]}{d\theta_e} \end{array} \right] \end{array} \right]. \tag{C18}$$

C.2 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) According to $\frac{\partial \pounds}{\partial L_w(\theta_e)}$, one has:

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

where $\int_{\theta_e} \varphi'(\theta_e') \frac{\partial \ln \varpi(\theta_e')}{\partial \ln L_w(\theta_e)} dx' = \varphi'(\theta_e)\varepsilon_{L_w}^{\varpi}(\theta_e)$ 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

$$W\mu(\theta_e) = \frac{\lambda'}{\lambda} - \frac{\kappa'(\theta_e)\xi}{\lambda L_w(\theta_e) N f_e(\theta_e)} + \frac{\varphi'(\theta_e)\varepsilon_{L_w}^{\varpi}(\theta_e)}{\lambda L_w(\theta_e) N f_e(\theta_e)}. \tag{C19}$$

Dividing both sides by $\frac{\varepsilon_{L_w}^{\varpi}(\theta_e)}{L_w(\theta_e) N f_e(\theta_e)}$ and integrating across θ_e gives

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

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

$$\begin{aligned} 1 &= \frac{\frac{\lambda'}{\lambda} \int_{\theta_e} \frac{L_w(\theta_e) N f_e(\theta_e)}{\varepsilon_{L_w}^{\varpi}(\theta_e)} d\theta_e}{W \int_{\theta_e} \mu(\theta_e) \frac{L_w(\theta_e) N f_e(\theta_e)}{\varepsilon_{L_w}^{\varpi}(\theta_e)} d\theta_e} - \frac{\int_{\theta_e} \frac{\kappa'(\theta_e)}{\lambda \varepsilon_{L_w}^{\varpi}(\theta_e)} \xi d\theta_e}{W \int_{\theta_e} \mu(\theta_e) \frac{L(\theta_e) N f_e(\theta_e)}{\varepsilon_{L_w}^{\varpi}(\theta_e)} d\theta_e} \\ &= \frac{\frac{\lambda'}{\lambda} \int_{\theta_e} \frac{L_w(\theta_e) N f_e(\theta_e)}{\varepsilon_{L_w}^{\varpi}(\theta_e)} d\theta_e}{W \int_{\theta_e} \mu(\theta_e) \frac{L_w(\theta_e) N f_e(\theta_e)}{\varepsilon_{L_w}^{\varpi}(\theta_e)} d\theta_e} + \int_{\theta_e} \frac{\kappa(\theta_e)}{\lambda} \frac{d \frac{\xi}{\varepsilon_{L_w}^{\varpi}(\theta_e)} / d\theta_e}{W \int_{\Theta_e} \mu(\theta_e) \frac{L_w(\theta_e) N f_e(\theta_e)}{\varepsilon_{L_w}^{\varpi}(\theta_e)} d\theta_e} d\theta_e, \end{aligned} \tag{C20}$$

where the second equation is derived by $\kappa(\underline{\theta}_e) = \kappa(\overline{\theta}_e) = 0$ and integration by parts.

Define

$$\varepsilon_{1-\tau}^{Q_{ij}}(\theta_e) \equiv \frac{\frac{\xi}{\varepsilon_{L_w}^{\varpi}(\theta_e)}}{\int_{\theta_e} \mu(\theta_e) \frac{W L(\theta_e) N f_e(\theta_e)}{\varepsilon_{L_w}^{\varpi}(\theta_e)} d\theta_e} \tag{C21}$$

Note that under our production function, labor inputs are perfectly substitutable. Thus, $\varepsilon_{L_w}^{\varpi}(\theta_e)$ is independent of θ_e and $\frac{d\varepsilon_{1-\tau}^{Q_{ij}}(\theta_e)}{d\theta_e} = 0$. Combining the above definitions and (C20) gives

$$\begin{aligned} 1 &= \frac{\lambda'}{\lambda W \mu} + \int_{\theta_e} \frac{\kappa'(\theta_e)}{\lambda} \frac{d\varepsilon_{1-\tau}^{Q_{ij}}(\theta_e)}{d\theta_e} d\theta_e \\ &= \frac{\lambda'}{\lambda W \mu}, \end{aligned} \tag{C22}$$

where the second equation is derived by $\frac{d\varepsilon_{1-\tau}^{Q_{ij}}(\theta_e)}{d\theta_e} = 0$.

According to (C7), we have

$$\frac{1}{\frac{\phi_w'(l_w(\theta_w))}{x_w(\theta_w)}} = \frac{\lambda}{\lambda'} \left[1 - \frac{x_w'(\theta_w)}{x_w(\theta_w)} \frac{\psi_w(\theta_w)}{\lambda N_w f_w(\theta_w)} \frac{1+\varepsilon_w}{\varepsilon_w} \right]. \tag{C23}$$

Substituting $\frac{\phi_w'(l_w(\theta_w))}{x_w(\theta_w)}$ with $[1-\tau_w(\theta_w)]\,W$ gives

$$\frac{1}{1-\tau_w(\theta_w)} = \frac{W\lambda}{\lambda'} \left[1 - \frac{x_w'(\theta_w)}{x_w(\theta_w)} \frac{\psi_w(\theta_w)}{\lambda N_w f_w(\theta_w)} \frac{1+\varepsilon_w}{\varepsilon_w} \right] \tag{C24}$$

Second, combining (C24), (C13), and (C22) gives

$$\frac{1}{1-\tau_w(\theta_w)} = \frac{1}{\mu} \left[1 + [1-\bar{g}_w(\theta_w)] \frac{1-F_w(\theta_w)}{f_w(\theta_w)} \frac{x_w'(\theta_w)}{x_w(\theta_w)} \frac{1+\varepsilon_w}{\varepsilon_w} \right]. \tag{C25}$$

(ii) We derive an optimal profit tax formula in part (a) of the following proof. Then we simplify the expression in parts (b) and (c).

(a) Divide both sides of (C8) by $\lambda N_e f_e(\theta_e) P(\theta_e) \frac{\partial Q_{ij}(\theta_e)}{\partial L_e(\theta_e)}$, and we have

$$\begin{aligned} & 1 - \frac{\phi_e'(l_e(\theta_e))}{P(\theta_e) \frac{\partial Q_{ij}(\theta_e)}{\partial l_e(\theta_e)}} \\ = & -\frac{\psi_e(\theta_e)}{\lambda N_e f_e(\theta_e)} \frac{\phi_e'(l_e(\theta_e))}{P(\theta_e) \frac{\partial Q_{ij}(\theta_e)}{\partial l_e(\theta_e)}} \frac{1+\varepsilon_e}{\varepsilon_e} \left[\mu(\theta_e) \frac{\partial \ln P(Q(\theta_e), \theta_e)}{\partial \theta_e} + \frac{x_e'(\theta_e)}{x_e(\theta_e)} \right] \\ & - \frac{\kappa'(\theta_e)}{\lambda l_e(\theta_e) P(\theta_e) \frac{\partial Q_{ij}(\theta_e)}{\partial l_e(\theta_e)} N f_e(\theta_e)} + \frac{\varphi'(\theta_e) \varepsilon_{l_e}^{\varpi}(\theta_e)}{\lambda l_e(\theta_e) P(\theta_e) \frac{\partial Q_{ij}(\theta_e)}{\partial l_e(\theta_e)} N f_e(\theta_e)}, \end{aligned} \tag{C26}$$

where we use $\frac{\partial \ln Q_{ij}(\theta_e)}{\partial \ln l_e(\theta_e)} = 1$ and $\int_\Theta \varphi'(\theta_e') \frac{\partial \ln \varpi(\theta_e')}{\partial \ln l_e(\theta_e)} d\theta_e = \varphi'(\theta_e)\, \varepsilon_{l_e}^{\varpi}(\theta_e)$ to simplify the expression.

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

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

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

$$\begin{aligned}\frac{\widetilde{\tau}_e(\theta_e)}{1-\widetilde{\tau}_e(\theta_e)} &= -\frac{\psi_e(\theta_e)}{\lambda N_e f_e(\theta_e)}\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu(\theta_e)\frac{\partial \ln P(Q(\theta_e),\theta_e)}{\partial\theta_e}+\frac{x_e'(\theta_e)}{x_e(\theta_e)}\right] \\ &\quad -\frac{1}{1-\widetilde{\tau}_e(\theta_e)}\frac{1}{P(\theta_e)Q_{ij}(\theta_e)}\left[\frac{\kappa'(\theta_e)}{\lambda N_e f_e(\theta_e)}-\frac{\varphi'(\theta_e)}{\lambda N_e f_e(\theta_e)}\varepsilon^{\varpi}_{l_e}(\theta_e)\right]\end{aligned} \tag{C27}$$

or

$$\begin{aligned}\frac{1}{1-\widetilde{\tau}_e(\theta_e)} &= 1-\frac{\psi_e(\theta_e)}{\lambda N_e f_e(\theta_e)}\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu(\theta_e)\frac{\partial \ln P(Q(\theta_e),\theta_e)}{\partial\theta_e}+\frac{x_e'(\theta_e)}{x_e(\theta_e)}\right] \\ &\quad -\frac{1}{1-\widetilde{\tau}_e(\theta_e)}\frac{1}{P(\theta_e)Q_{ij}(\theta_e)}\left[\frac{\kappa'(\theta_e)}{\lambda N_e f_e(\theta_e)}-\frac{\varphi'(\theta_e)}{\lambda N_e f_e(\theta_e)}\varepsilon^{\varpi}_{l_e}(\theta_e)\right].\end{aligned} \tag{C28}$$

Combining (C27) and (C13) gives

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

Using (C19) to substitute $\frac{\varphi'(\theta_e)\varepsilon^{\varpi}_{l_e}(\theta_e)}{\lambda N_e f_e(\theta_e)}$,[45] we have

$$\begin{aligned}\frac{\widetilde{\tau}_e(\theta_e)}{1-\widetilde{\tau}_e(\theta_e)} &= [1-\bar{g}_e(\theta_e)]\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu(\theta_e)\frac{\partial \ln P(Q(\theta_e),\theta_e)}{\partial\theta_e}+\frac{x_e'(\theta_e)}{x_e(\theta_e)}\right] \\ &\quad -\frac{1}{1-\widetilde{\tau}_e(\theta_e)}\frac{1}{P(\theta_e)Q_{ij}(\theta_e)}\frac{\kappa'(\theta_e)}{\lambda N_e f_e(\theta_e)}\left[1-\xi\frac{\varepsilon^{\varpi}_{l_e}(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)}\right] \\ &\quad -\frac{1}{1-\widetilde{\tau}_e(\theta_e)}\frac{L_w(\theta_e)}{P(\theta_e)Q_{ij}(\theta_e)}\frac{\lambda'}{\lambda}\left[1-\frac{\lambda}{\lambda'}\frac{W\mu(\theta_e)}{1-\tau_s}\right]\frac{\varepsilon^{\varpi}_{l_e}(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)}.\end{aligned} \tag{C29}$$

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

[45]Equation (C19) suggests that $\frac{\varphi'(\theta_e)\varepsilon^{\varpi}_{l_e}(\theta_e)}{\lambda N_e f_e(\theta_e)} = \left[\left[\frac{W\mu(\theta)}{1-\tau_s}-\frac{\lambda'}{\lambda}\right]L_w(\theta_e)+\frac{\kappa'(\theta_e)\xi}{\lambda N_e f_e(\theta_e)}\right]\frac{\varepsilon^{\varpi}_{l_e}(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)}$.

ing $\kappa'(\theta_e)$ with (C18), we have the following equation:[46]

$$-\frac{1}{1-\widetilde{\tau}_e(\theta_e)}\frac{1}{P(\theta_e)Q_{ij}(\theta_e)}\frac{\kappa'(\theta_e)}{\lambda N_e f_e(\theta_e)} \tag{C30}$$

$$= \frac{1-\tau_e(\theta_e)}{1-\widetilde{\tau}_e(\theta_e)}\varepsilon_{Q_{-ij}}(\theta_e)\left[\begin{array}{c}[1-g_e(\theta_e)]-\frac{[1-\bar{g}_e(\theta_e)][1-F_e(\theta_e)]}{f_e(\theta_e)} \\ \times\left[\begin{array}{c}\frac{1+\varepsilon_e}{\varepsilon_e}\frac{l_e'(\theta_e)}{l_e(\theta_e)}+ \\ \frac{d\ln\left[\mu(\theta_e)\varepsilon_{Q_{-ij}}(\theta_e)\right]}{d\theta_e}\end{array}\right]\end{array}\right].$$

Second, notice that $\frac{L_w(\theta_e)W}{P(\theta_e)Q_{ij}(\theta_e)}=\frac{\xi}{\mu(\theta_e)}$ and $\frac{\lambda'}{\lambda W}=\mu$ (e.g., see (C22)). The last term of (C29) equals

$$-\frac{1}{1-\widetilde{\tau}_e(\theta_e)}\frac{L_w(\theta_e)}{P(\theta_e)Q_{ij}(\theta_e)}\frac{\lambda'}{\lambda}\left[1-\frac{\lambda}{\lambda'}\frac{W\mu(\theta_e)}{1-\tau_s}\right]\frac{\varepsilon^{\omega}_{l_e}(\theta_e)}{\varepsilon^{\omega}_{L_w}(\theta_e)} \tag{C31}$$

$$= -\frac{1-\tau_s}{1-\widetilde{\tau}_e(\theta_e)}\frac{\xi}{\mu(\theta_e)}\mu\left[1-\frac{\mu(\theta_e)}{\mu}\right]\frac{\varepsilon^{\omega}_{l_e}(\theta_e)}{\varepsilon^{\omega}_{L_w}(\theta_e)}$$

$$= -\frac{\xi}{1-\widetilde{\tau}_e(\theta_e)}\left[\frac{\mu}{\mu(\theta_e)}-1\right]\frac{\varepsilon^{\omega}_{l_e}(\theta_e)}{\varepsilon^{\omega}_{L_w}(\theta_e)}.$$

Combining equations (C30) to (C31) gives

$$\frac{\widetilde{\tau}_e(\theta_e)}{1-\widetilde{\tau}_e(\theta_e)} \tag{C32}$$

$$= [1-\bar{g}_e(\theta_e)]\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu(\theta_e)\frac{\partial\ln P\left(Q_{ij}(\theta_e),\theta_e\right)}{\partial\theta_e}+\frac{x_e'(\theta_e)}{x_e(\theta_e)}\right]$$

$$+\frac{1-\tau_e(\theta_e)}{1-\widetilde{\tau}_e(\theta_e)}\varepsilon_{Q_{-ij}}(\theta_e)\left[\begin{array}{c}[1-g_e(\theta_e)]- \\ \frac{[1-\bar{g}_e(\theta_e)][1-F_e(\theta_e)]}{f_e(\theta_e)}\left[\begin{array}{c}\frac{1+\varepsilon_e}{\varepsilon_e}\frac{l_e'(\theta_e)}{l_e(\theta_e)} \\ +\frac{d\ln\left[\mu(\theta_e)\varepsilon_{Q_{-ij}}(\theta_e)\right]}{d\theta_e}\end{array}\right]\end{array}\right]\left[1-\xi\frac{\varepsilon^{\omega}_{l_e}(\theta_e)}{\varepsilon^{\omega}_{L_w}(\theta_e)}\right]$$

$$-\frac{\xi}{1-\widetilde{\tau}_e(\theta_e)}\left[\frac{\mu}{\mu(\theta_e)}-1\right]\frac{\varepsilon^{\omega}_{l_e}(\theta_e)}{\varepsilon^{\omega}_{L_w}(\theta_e)}.$$

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

In addition, substituting $1-\widetilde{\tau}_e(\theta_e)$ by $\frac{1-\tau_e(\theta_e)}{\mu(\theta_e)}$, we have

$$\begin{aligned}
&\frac{1}{1-\tau_e(\theta_e)} \qquad \text{(C33)}\\
= \;& \left[1+[1-\bar{g}_e(\theta_e)]\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu(\theta_e)\frac{\partial \ln P\left(Q_{ij}(\theta_e),\theta_e\right)}{\partial\theta_e}+\frac{x_e'(\theta_e)}{x_e(\theta_e)}\right]\right]\frac{1}{\mu(\theta_e)}\\
&+\varepsilon_{Q_{-ij}}(\theta_e)\left[\begin{array}{c}[1-g_e(\theta_e)]-\\ \frac{[1-\bar{g}_e(\theta_e)][1-F_e(\theta_e)]}{f_e(\theta_e)}\left[\begin{array}{c}\frac{1+\varepsilon_e}{\varepsilon_e}\frac{l_e'(\theta_e)}{l_e(\theta_e)}\\ +\frac{d\ln\left[\mu(\theta_e)\varepsilon_{Q_{-ij}}(\theta_e)\right]}{d\theta_e}\end{array}\right]\end{array}\right]\left[1-\xi\frac{\varepsilon^{\varpi}_{l_e}(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)}\right]\\
&+\frac{1}{1-\tau_e(\theta_e)}\left[1-\frac{\mu}{\mu(\theta_e)}\right]\xi\frac{\varepsilon^{\varpi}_{l_e}(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)}.
\end{aligned}$$

Using

$$RE(\theta_e)\equiv\left[\frac{\mu}{\mu(\theta_e)}-1\right]$$

and

$$\widetilde{IRE}(\theta_e)\equiv\varepsilon_{Q_{-ij}}(\theta_e)\left[\begin{array}{c}[1-g_e(\theta_e)]-\frac{[1-\bar{g}_e(\theta_e)][1-F_e(\theta_e)]}{f_e(\theta_e)}\\ \times\left[\frac{1+\varepsilon_e}{\varepsilon_e}\frac{l_e'(\theta_e)}{l_e(\theta_e)}+\frac{d\ln\left[\mu(\theta_e)\varepsilon_{Q_{-ij}}(\theta_e)\right]}{d\theta_e}\right]\end{array}\right],$$

we have

$$\begin{aligned}
\frac{1}{1-\tau_e(\theta_e)} \;=\;& \frac{1+[1-\bar{g}_e(\theta_e)]\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu(\theta_e)\frac{\partial\ln P(Q(\theta_e),\theta_e)}{\partial\theta_e}+\frac{x_e'(\theta_e)}{x_e(\theta_e)}\right]}{\mu(\theta_e)} \qquad \text{(C34)}\\
&+\widetilde{IRE}(\theta_e)\overbrace{\left[1-\xi\frac{\varepsilon^{\varpi}_{l_e}(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)}\right]}^{\text{Elasticity of } Q_{ij} \text{ w.r.t } l_e}\\
&+\frac{1}{1-\tau_e(\theta_e)}\xi RE(\theta_e)\overbrace{\left[-\frac{\varepsilon^{\varpi}_{l_e}(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)}\right]}^{\text{Elasticity of } L_w \text{ w.r.t } l_e},
\end{aligned}$$

which is equivalent to

$$\frac{1}{1-\tau_e(\theta_e)}=\frac{\frac{1+[1-\bar{g}_e(\theta_e)]\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu(\theta_e)\frac{\partial\ln P(Q(\theta_e),\theta_e)}{\partial\theta_e}+\frac{x_e'(\theta_e)}{x_e(\theta_e)}\right]}{\mu(\theta_e)}+\widetilde{IRE}(\theta_e)\left[1-\xi\frac{\varepsilon^{\varpi}_{l_e}(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)}\right]}{1+RE(\theta_e)\,\xi\frac{\varepsilon^{\varpi}_{l_e}(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)}}. \qquad \text{(C35)}$$

(b) To derive optimal profit tax formular in termes of parameters, we first derive $\frac{d\ln L_w(\theta_e)}{d\theta_e}$, $\frac{d\ln l_e(\theta_e)}{d\theta_e}$ and $\frac{d\ln\pi(\theta_e)}{d\theta_e}$ in terms of θ_e and profit tax rate. Using (A1) and $Q_{ij}(\theta_e)=x_e(\theta_e)\,l_e(\theta_e)\,L_{w,ij}(\theta_e)^{\xi}$ to substitute $P(\theta_e)$ and $Q_{ij}(\theta_e)$ in (A5) and rearrange the equation, we have the following in the equilibrium when

$t_s = 0$:

$$L_w(\theta_e) = \left[\frac{X(\theta_e)}{\mu(\theta_e)} \frac{\xi A^{\frac{\sigma-1}{\sigma}} Q^{\frac{1}{\sigma}}}{W N_e^{\frac{1}{\sigma}}} l_e(\theta_e)^{\frac{\sigma-1}{\sigma}} \right]^{\frac{1}{1-\xi\frac{\sigma-1}{\sigma}}}, \tag{C36}$$

and

$$\frac{d\ln L_w(\theta_e)}{d\theta_e} = \frac{1}{1-\xi\frac{\sigma-1}{\sigma}} \frac{d\ln X(\theta_e)/\mu(\theta_e)}{d\theta_e} + \frac{\frac{\sigma-1}{\sigma}}{1-\xi\frac{\sigma-1}{\sigma}} \frac{l_e'(\theta_e)}{l_e(\theta_e)}, \forall\theta \in \Theta. \tag{C37}$$

The entrepreneurial effort $l_e(\theta_e)$ satisfies the first order condition (e.g., see (A6))

$$\frac{P(\theta_e) Q_{ij}(\theta_e)}{\mu(\theta_e)} [1-\tau_e(\theta_e)] = l_e(\theta_e)^{1+\frac{1}{\varepsilon_e}},$$

where $\tau_e(\theta_e) = T'(\pi(\theta_e))$. Using (A1) and the expression of $Q_{ij}(\theta_e)$ to substitute $P_{ij}(\theta_e)$ and $Q_{ij}(\theta_e)$ again, we have

$$\left[\frac{X(\theta_e)}{\mu(\theta_e)} \frac{\xi A^{\frac{\sigma-1}{\sigma}} Q^{\frac{1}{\sigma}}}{W N_e^{\frac{1}{\sigma}}} \right]^{\frac{1}{1-\xi\frac{\sigma-1}{\sigma}}} [1-\tau_e(\theta_e)] W = l_e(\theta_e)^{\frac{\frac{\varepsilon_e+1}{\varepsilon_e} - \frac{\sigma-1}{\sigma}\left(1+\frac{\varepsilon_e+1}{\varepsilon_e}\xi\right)}{1-\xi\frac{\sigma-1}{\sigma}}}. \tag{C38}$$

Taking the derivation of the items on both sides of the above formula to obtain the following equation:

$$\frac{d\ln l_e(\theta_e)}{d\theta_e} = \frac{\frac{d\ln X(\theta_e)/\mu(\theta_e)}{d\theta_e} + \left[1-\xi\frac{\sigma-1}{\sigma}\right] \frac{d\ln[1-\tau_e(\theta_e)]}{d\theta_e}}{\frac{\varepsilon_e+1}{\varepsilon_e} - \frac{\sigma-1}{\sigma}\left(1+\frac{\varepsilon_e+1}{\varepsilon_e}\xi\right)}, \forall\theta \in \Theta. \tag{C39}$$

Combination of firm's first order condition (A5), and the definition of $\pi(\theta_e)$, $\pi(\theta_e) = P_{ij}(\theta_e) Q_{ij}(\theta_e) - W L_{w,ij}(\theta_e)$, gives

$$\pi(\theta_e) = \frac{\mu(\theta_e) - \xi}{\mu(\theta_e)} P(\theta_e) Q_{ij}(\theta_e), \forall\theta \in \Theta. \tag{C40}$$

Substituting $P_{ij}(\theta_e) Q_{ij}(\theta_e)$ in (C40) by (A5), and then $L_w(\theta_e)$ by (C36), we have

$$\begin{aligned} \pi(\theta_e) &= [\mu(\theta_e) - \xi] \frac{X(\theta_e)}{\mu(\theta_e)} l_e(\theta_e)^{\frac{\sigma-1}{\sigma}} L_w(\theta_e)^{\xi\frac{\sigma-1}{\sigma}} \frac{A^{\frac{\sigma-1}{\sigma}} Q^{\frac{1}{\sigma}}}{W N_e^{\frac{1}{\sigma}}} W \\ &= [\mu(\theta_e) - \xi] \left[\frac{X(\theta_e)}{\mu(\theta_e)} \frac{\xi A^{\frac{\sigma-1}{\sigma}} Q^{\frac{1}{\sigma}}}{W N_e^{\frac{1}{\sigma}}} \right]^{\frac{1}{1-\xi\frac{\sigma-1}{\sigma}}} l_e(\theta_e)^{\frac{1-\frac{1}{\sigma}}{1-\xi\frac{\sigma-1}{\sigma}}} W, \forall\theta \in \Theta, \end{aligned} \tag{C41}$$

Taking the derivation of the items on both sides of (C41) and substitute $\frac{d\ln l_e(\theta_e)}{d\theta_e}$ by the above equation, we have

$$\frac{d\ln \pi(\theta_e)}{d\theta_e} = \frac{\frac{1+\varepsilon_e}{\varepsilon_e} \frac{d\ln[X(\theta_e)/\mu(\theta_e)]}{d\theta_e} + \frac{\sigma-1}{\sigma} \frac{d\ln[1-\tau_e(\theta_e)]}{d\theta_e}}{\frac{1+\varepsilon_e}{\varepsilon_e}\left(1-\xi\frac{\sigma-1}{\sigma}\right) - \frac{\sigma-1}{\sigma}} + \frac{d\ln[\mu(\theta_e) - \xi]}{d\theta_e}, \forall\theta \in \Theta. \tag{C42}$$

Therefore, we define

$$H(\theta_e) \equiv \frac{1-F_e(\theta_e)}{f_e(\theta_e)} \left[\frac{\frac{1+\varepsilon_e}{\varepsilon_e} \frac{d\ln[X(\theta_e)/\mu(\theta_e)]}{d\theta_e}}{\frac{1+\varepsilon_e}{\varepsilon_e}\left(1-\xi\frac{\sigma-1}{\sigma}\right) - \frac{\sigma-1}{\sigma}} + \frac{d\ln[\mu(\theta_e) - \xi]}{d\theta_e} \right], \forall\theta \in \Theta,$$

which is $\frac{1-F_e(\theta_e)}{f_e(\theta_e)} \frac{d \ln \pi(\theta_e)}{d\theta_e}$ (the hazard ratio of profit at $\pi(\theta_e)$) when $\frac{d\tau_e(\theta_e)}{d\theta_e} = 0$. Note that once $\pi(\theta_e)$ is established by the data, $H(\theta_e)$ can be derived by the profit distribution when the original profit tax is linear.

(c) We now derive a more explicit expression of $\mu(\theta_e) \frac{\partial \ln P\left(Q_{ij}(\theta_e),\theta_e\right)}{\partial\theta_e} + \frac{x_e'(\theta_e)}{x_e(\theta_e)}$. Remind that (e.g., see (A45) and (A46)) $\varepsilon_{Q_{-ij}}(\theta_e) = -\frac{1}{\mu(\theta_e)} + \frac{\sigma-1}{\sigma}$ and

$$\frac{\partial \ln P\left(Q_{ij},\theta_e\right)}{\partial\theta_e} = \frac{\chi'(\theta_e)}{\chi(\theta_e)} + \varepsilon_{Q_{-ij}}(\theta_e) \frac{d \ln Q_{ij}(\theta_e)}{d\theta_e}, \forall\theta \in \Theta. \tag{C43}$$

where

$$\begin{aligned}
\frac{d \ln Q_{ij}(\theta_e)}{d\theta_e} &= \frac{x_e'(\theta_e)}{x_e(\theta_e)} + \frac{d \ln l_e(\theta_e)}{d\theta_e} + \xi \frac{d \ln L_{w,ij}(\theta_e)}{d\theta_e} \\
&= \frac{x_e'(\theta_e)}{x_e(\theta_e)} + \frac{1}{1-\xi\frac{\sigma-1}{\sigma}} \frac{l_e'(\theta_e)}{l_e(\theta_e)} + \\
&\quad \frac{\xi}{1-\xi\frac{\sigma-1}{\sigma}} \frac{d}{d\theta_e} \ln \frac{X(\theta_e)}{\mu(\theta_e)}.
\end{aligned}$$

The second equation of the above equations is derived by (C37). Substituting $\frac{l_e'(\theta_e)}{l_e(\theta_e)}$ by (C39), we have

$$\begin{aligned}
&\frac{d \ln Q_{ij}(\theta_e)}{d\theta_e} \\
= &\frac{x_e'(\theta_e)}{x_e(\theta_e)} + \frac{\xi}{1-\xi\frac{\sigma-1}{\sigma}} \frac{d}{d\theta_e} \ln \frac{X(\theta_e)}{\mu(\theta_e)} + \\
&\frac{1}{1-\xi\frac{\sigma-1}{\sigma}} \frac{\frac{d}{d\theta_e} \ln \frac{X(\theta_e)}{\mu(\theta_e)} + \left[1-\xi\frac{\sigma-1}{\sigma}\right] \frac{d \ln[1-\tau_e(\theta_e)]}{d\theta_e}}{\frac{\varepsilon_e+1}{\varepsilon_e} - \frac{\sigma-1}{\sigma}\left(1+\frac{\varepsilon_e+1}{\varepsilon_e}\xi\right)} \\
= &\frac{x_e'(\theta_e)}{x_e(\theta_e)} + \frac{\left(1+\xi\frac{\varepsilon_e+1}{\varepsilon_e}\right) \frac{d}{d\theta_e} \ln \frac{X(\theta_e)}{\mu(\theta_e)} + \frac{d \ln[1-\tau_e(\theta_e)]}{d\theta_e}}{\frac{\varepsilon_e+1}{\varepsilon_e} - \frac{\sigma-1}{\sigma}\left(1+\frac{\varepsilon_e+1}{\varepsilon_e}\xi\right)}
\end{aligned}$$

Substituting $\frac{d \ln Q_{ij}(\theta_e)}{d\theta_e}$ in (A46) by the above equation gives

$$\begin{aligned}
&\mu(\theta_e) \frac{\partial \ln P\left(Q_{ij}(\theta_e),\theta_e\right)}{\partial\theta_e} + \frac{x_e'(\theta_e)}{x_e(\theta_e)} \\
= &\mu(\theta_e) \left[\frac{\chi'(\theta_e)}{\chi(\theta_e)} + \varepsilon_{Q_{-ij}}(\theta_e) \frac{d \ln Q_{ij}(\theta_e)}{d\theta_e}\right] + \frac{x_e'(\theta_e)}{x_e(\theta_e)} \\
= &\mu(\theta_e) \frac{\chi'(\theta_e)}{\chi(\theta_e)} + \left[\mu(\theta_e) \varepsilon_{Q_{-ij}}(\theta_e) + 1\right] \frac{x_e'(\theta_e)}{x_e(\theta_e)} + \\
&\mu(\theta_e) \varepsilon_{Q_{-ij}}(\theta_e) \frac{\left(1+\xi\frac{\varepsilon_e+1}{\varepsilon_e}\right) \frac{d}{d\theta_e} \ln \frac{X(\theta_e)}{\mu(\theta_e)} + \frac{d \ln[1-\tau_e(\theta_e)]}{d\theta_e}}{\frac{\varepsilon_e+1}{\varepsilon_e}\left(1-\xi\frac{\sigma-1}{\sigma}\right) - \frac{\sigma-1}{\sigma}},
\end{aligned}$$

where

$$\begin{aligned}
& \mu(\theta_e)\frac{\chi'(\theta_e)}{\chi(\theta_e)}+\left[\mu(\theta_e)\varepsilon_{Q_{-ij}}(\theta_e)+1\right]\frac{x_e'(\theta_e)}{x_e(\theta_e)} \\
= \;& \mu(\theta_e)\frac{\chi'(\theta_e)}{\chi(\theta_e)}+\mu(\theta_e)\frac{\sigma-1}{\sigma}\frac{x_e'(\theta_e)}{x_e(\theta_e)} \\
= \;& \mu(\theta_e)\frac{d}{d\theta_e}\left[\ln\frac{X(\theta_e)}{\mu(\theta_e)}\right]+\mu(\theta_e)\frac{d\ln\mu(\theta_e)}{d\theta_e} \\
= \;& \left[\mu(\theta_e)\varepsilon_{Q_{-ij}}(\theta_e)+1\right]\frac{\sigma}{\sigma-1}\frac{d}{d\theta_e}\left[\ln\frac{X(\theta_e)}{\mu(\theta_e)}\right]+\mu(\theta_e)\frac{d\ln\mu(\theta_e)}{d\theta_e}.
\end{aligned}$$

The first and third equations of the above equations are derived by $\varepsilon_{Q_{-ij}}(\theta_e)=-\frac{1}{\mu(\theta_e)}+\frac{\sigma-1}{\sigma}$, and the second equation is derived by $X(\theta_e)=\chi(\theta_e)\,x_e(\theta_e)^{\frac{\sigma-1}{\sigma}}$. Therefore, we have

$$\begin{aligned}
& \frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu(\theta_e)\frac{\partial\ln P\left(Q_{ij}(\theta_e),\theta_e\right)}{\partial\theta_e}+\frac{x_e'(\theta_e)}{x_e(\theta_e)}\right] \\
= \;& \frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{1+\varepsilon_e}{\varepsilon_e}\left\{\begin{array}{c}\left[\mu(\theta_e)\varepsilon_{Q_{-ij}}(\theta_e)+1\right]\frac{\sigma}{\sigma-1}\frac{d}{d\theta_e}\left[\ln\frac{X(\theta_e)}{\mu(\theta_e)}\right]+\mu(\theta_e)\frac{d\ln\mu(\theta_e)}{d\theta_e} \\ +\mu(\theta_e)\varepsilon_{Q_{-ij}}(\theta_e)\frac{\left(1+\xi\frac{\varepsilon_e+1}{\varepsilon_e}\right)\frac{d}{d\theta_e}\left[\ln\frac{X(\theta_e)}{\mu(\theta_e)}\right]+\frac{d\ln[1-\tau_e(\theta_e)]}{d\theta_e}}{\frac{\varepsilon_e+1}{\varepsilon_e}\left(1-\xi\frac{\sigma-1}{\sigma}\right)-\frac{\sigma-1}{\sigma}}\end{array}\right\} \\
= \;& \frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{d\ln[\mu(\theta_e)-\xi]}{d\theta_e}\frac{1+\varepsilon_e}{\varepsilon_e}[\mu(\theta_e)-\xi]+ \\
& \frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{\frac{1+\varepsilon_e}{\varepsilon_e}\frac{d}{d\theta_e}\ln\frac{X(\theta_e)}{\mu(\theta_e)}}{\frac{1+\varepsilon_e}{\varepsilon_e}\left(1-\xi\frac{\sigma-1}{\sigma}\right)-\frac{\sigma-1}{\sigma}}\left[\begin{array}{c}\left[\mu(\theta_e)\varepsilon_{Q_{-ij}}(\theta_e)+1\right]\frac{\sigma}{\sigma-1}\frac{\varepsilon_e+1}{\varepsilon_e} \\ -\left(1+\xi\frac{\varepsilon_e+1}{\varepsilon_e}\right)\end{array}\right] \\
& +\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{1+\varepsilon_e}{\varepsilon_e}\frac{\mu(\theta_e)\varepsilon_{Q_{-ij}}(\theta_e)\frac{d\ln[1-\tau_e(\theta_e)]}{d\theta_e}}{\frac{1+\varepsilon_e}{\varepsilon_e}\left(1-\xi\frac{\sigma-1}{\sigma}\right)-\frac{\sigma-1}{\sigma}},
\end{aligned}$$

where the second 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 $\frac{d}{d\theta_e}\ln\frac{X(\theta_e)}{\mu(\theta_e)}$. Using (45), which implies

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

we have

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

where according to $\varepsilon_{Q_{-ij}}(\theta_e) = -\frac{1}{\mu(\theta_e)} + \frac{\sigma-1}{\sigma}$, $\left[\mu(\theta_e)\varepsilon_{Q_{-ij}}(\theta_e) + 1\right]\frac{\sigma}{\sigma-1} = \mu(\theta_e)$.

Therefore,

$$\begin{aligned} & \frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu(\theta_e)\frac{\partial \ln P\left(Q_{ij}(\theta_e),\theta_e\right)}{\partial\theta_e} + \frac{x_e'(\theta_e)}{x_e(\theta_e)}\right] \\ = \; & \frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{d\ln[\mu(\theta_e)-\xi]}{d\theta_e} + H(\theta_e)\left[\frac{\varepsilon_e+1}{\varepsilon_e}[\mu(\theta_e)-\xi]-1\right] \\ & + \frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{1+\varepsilon_e}{\varepsilon_e}\frac{\mu(\theta_e)\varepsilon_{Q_{-ij}}(\theta_e)\frac{d\ln[1-\tau_e(\theta_e)]}{d\theta_e}}{\frac{1+\varepsilon_e}{\varepsilon_e}\left(1-\xi\frac{\sigma-1}{\sigma}\right)-\frac{\sigma-1}{\sigma}}. \end{aligned} \tag{C44}$$

Substituting $\frac{l_e'(\theta_e)}{l_e(\theta_e)}$ in $\widetilde{IRE}(\theta_e)$ by (C39), we have, for any $\theta_e \in \Theta_e$,

$$\widetilde{IRE}(\theta_e) = [1-g_e(\theta_e)]\varepsilon_{Q_{-ij}}(\theta_e) - \varepsilon_{Q_{-ij}}(\theta_e)[1-\bar{g}_e(\theta_e)]\left\{\begin{array}{c} H(\theta_e) + \frac{\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{d\ln[\mu(\theta_e)-\xi]}{d\theta_e}}{\left[1-\xi\frac{\varepsilon^{\varpi}_{l_e}(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)}\right]\mu(\theta_e)\varepsilon_{Q_{-ij}}(\theta_e)} \\ + \frac{\frac{1+\varepsilon_e}{\varepsilon_e}\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{d\ln[1-\tau_e(\theta_e)]}{d\theta_e}}{\left[1-\xi\frac{\varepsilon^{\varpi}_{l_e}(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)}\right]\left[\frac{\varepsilon_e+1}{\varepsilon_e}-\frac{\sigma-1}{\sigma}\left(1+\frac{\varepsilon_e+1}{\varepsilon_e}\xi\right)\right]} \end{array}\right\}. \tag{C45}$$

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 (C35) by (C44) and (C45), respectively, we have, for any $\theta_e \in \Theta_e$,

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

Notice that $1-\xi\frac{\varepsilon^{\varpi}_{l_e}(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)} = \frac{1}{1-\xi\frac{\sigma-1}{\sigma}}$ (e.g., see (39)), one can see that the sum of terms multiplied by $\frac{d\ln[1-\tau_e(\theta_e)]}{d\theta_e}$ 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 $IRE(\theta_e)$ (e.g., see (44)), we have (42).

On the other hand, combining (C36) and (C38), we have

$$L_w(\theta_e) = \left[\left(\frac{X(\theta_e)}{\mu(\theta_e)}\frac{\xi A^{\frac{\sigma-1}{\sigma}}Q^{\frac{1}{\sigma}}}{WN_e^{\frac{1}{\sigma}}}\right)^{\frac{\varepsilon_e+1}{\varepsilon_e}}[1-\tau_e(\theta_e)]^{\frac{\sigma-1}{\sigma}}W^{\frac{\sigma-1}{\sigma}}\right]^{\frac{1}{\frac{1+\varepsilon_e}{\varepsilon_e}\left(1-\xi\frac{\sigma-1}{\sigma}\right)-\frac{\sigma-1}{\sigma}}}.$$

Substituting $L_w(\theta_e)$ in (24) by the above equation, we have (46) and (47).■

C.3 Proof of Corollary 1

Substituting elasticities, $RE(\theta_e)$ and $IRE(\theta_e)$ in (42) by (39), (40), (43) and (44), we have

$$1-\tau_e(\theta_e) = \frac{\frac{\frac{\sigma}{\sigma-1}}{\frac{\sigma}{\sigma-1}-\xi} - \frac{\xi}{\frac{\sigma}{\sigma-1}-\xi}\frac{\mu}{\mu(\theta_e)}}{\frac{1+[1-\bar{g}_e(\theta_e)]H(\theta_e)\left[\frac{1}{\varepsilon_e}\mu(\theta_e)-\frac{1+\varepsilon_e}{\varepsilon_e}\xi+\frac{1}{\sigma-1}\right]+\left[\mu(\theta_e)-\frac{\sigma}{\sigma-1}\right][1-g_e(\theta_e)]}{\mu(\theta_e)}}$$

Notice that $1<\sigma<\eta(\theta_e)$, $\frac{\sigma}{\sigma-1}\geq\mu(\theta_e)=\frac{1}{1-\left[\frac{1}{\eta(\theta_e)}\frac{I-1}{I}+\frac{1}{\sigma}\frac{1}{I}\right]}>1$, $0\leq\xi<1$. We have

$$1-\frac{\xi}{\frac{\sigma}{\sigma-1}-\xi}RE(\theta_e) = \frac{1}{\frac{\sigma}{\sigma-1}-\xi}\left[\frac{\sigma}{\sigma-1}\frac{\mu(\theta_e)}{\mu}-\xi\right]>0.$$

Then, we have

$$1+[1-\bar{g}_e(\theta_e)]H(\theta_e)\left[\frac{1}{\varepsilon_e}\mu(\theta_e)-\frac{1+\varepsilon_e}{\varepsilon_e}\xi+\frac{1}{\sigma-1}\right]+\left[\mu(\theta_e)-\frac{\sigma}{\sigma-1}\right][1-g_e(\theta_e)]\geq 0$$

since $1-\tau_e(\theta_e)\geq 0$. According to the above inequality, we have $\gamma(\theta_e)\geq 0$ and $\overline{\gamma}(\theta_e)\geq 0$. When $\frac{1+\varepsilon_e}{\varepsilon_e}\left(\frac{\sigma}{\sigma-1}-\xi\right)=2$, we have

$$\begin{aligned}\mu &= \frac{\int_{\theta_e}\mu(\theta_e)[1-\tau_e(\theta_e)]\left(\frac{X(\theta_e)}{\mu(\theta_e)}\right)^{\frac{\varepsilon_e+1}{\varepsilon_e}\frac{\sigma}{\sigma-1}}f_e(\theta_e)\,d\theta_e}{\int_{\theta_e}[1-\tau_e(\theta_e)]\left(\frac{X(\theta_e)}{\mu(\theta_e)}\right)^{\frac{\varepsilon_e+1}{\varepsilon_e}\frac{\sigma}{\sigma-1}}f_e(\theta_e)\,d\theta_e}\\ &= \frac{\int_{\theta_e}\mu(\theta_e)\gamma(\theta_e)\left[\frac{\sigma}{\sigma-1}\mu(\theta_e)-\xi\mu\right]d\theta_e}{\int_{\theta_e}\gamma(\theta_e)\left[\frac{\sigma}{\sigma-1}\mu(\theta_e)-\xi\mu\right]d\theta_e},\end{aligned}$$

where the second equation is derived by the relationship between $\gamma(\theta_e)$ and $[1-\tau_e(\theta_e)]$. Accordingly, we have

$$\begin{aligned}&\frac{\sigma}{\sigma-1}\mu\int_{\theta_e}\gamma(\theta_e)\mu(\theta_e)\,d\theta_e-\xi\mu^2\int_{\theta_e}\gamma(\theta_e)d\theta_e\\ =\ &\frac{\sigma}{\sigma-1}\int_{\theta_e}\mu(\theta_e)^2\gamma(\theta_e)d\theta_e-\xi\mu\int_{\theta_e}\mu(\theta_e)\gamma(\theta_e)d\theta_e,\end{aligned}$$

or equivalently

$$0 = \xi\mu^2 - \left[\left(\frac{\sigma}{\sigma-1} + \xi\right)\int_{\theta_e}\overline{\gamma}(\theta_e)\mu\left(\theta_e\right)d\theta_e\right]\mu + \frac{\sigma}{\sigma-1}\int_{\theta_e}\mu\left(\theta_e\right)^2\overline{\gamma}(\theta_e)d\theta_e, \tag{C46}$$

which is a quadratic equation of μ.

We define

$$\Delta = \left(\frac{\sigma}{\sigma-1} - \xi\right)^2\left[\int_{\theta_e}\overline{\gamma}(\theta_e)\mu\left(\theta_e\right)d\theta_e\right]^2 - 4\frac{\sigma}{\sigma-1}\xi\left[\int_{\theta_e}\mu^2\left(\theta_e\right)\overline{\gamma}(\theta_e)d\theta_e - \left(\int_{\theta_e}\overline{\gamma}(\theta_e)\mu\left(\theta_e\right)d\theta_e\right)^2\right]$$

as the discriminant of (C46). Set

$$\mathbf{E}_{\overline{\gamma}}\left(\mu\right) = \int_{\theta_e}\overline{\gamma}(\theta_e)\mu\left(\theta_e\right)d\theta_e, \mathbf{Var}_{\overline{\gamma}}\left(\mu\right) = \left[\int_{\theta_e}\mu^2\left(\theta_e\right)\overline{\gamma}(\theta_e)d\theta_e - \left(\int_{\theta_e}\overline{\gamma}(\theta_e)\mu\left(\theta_e\right)d\theta_e\right)^2\right]$$

and

$$\widetilde{\mu}\left(\theta_e\right) = \left(\sigma-1\right)\left[\mu\left(\theta_e\right) - 1\right].$$

We have $\widetilde{\mu}\left(\theta_e\right) \in (0,1]$, because $\mu\left(\theta_e\right) \in (1, \frac{\sigma}{\sigma-1}]$. Set

$$\begin{aligned}
\mathbf{E}_{\overline{\gamma}}\left(\widetilde{\mu}\right) &= \int_{\theta_e}\overline{\gamma}(\theta_e)\widetilde{\mu}\left(\theta_e\right)d\theta_e, \mathbf{E}_{\overline{\gamma}}\left(\widetilde{\mu}^2\right) = \int_{\theta_e}\overline{\gamma}(\theta_e)\widetilde{\mu}\left(\theta_e\right)^2 d\theta_e \\
\mathbf{Var}_{\overline{\gamma}}\left(\widetilde{\mu}\right) &= \left[\int_{\theta_e}\widetilde{\mu}\left(\theta_e\right)^2\overline{\gamma}(\theta_e)d\theta_e - \left(\int_{\theta_e}\overline{\gamma}(\theta_e)\widetilde{\mu}\left(\theta_e\right)d\theta_e\right)^2\right].
\end{aligned}$$

Then, we have

$$\Delta = \left(\frac{\sigma}{\sigma-1} - \xi\right)^2\left[1 + \frac{1}{\sigma-1}\mathbf{E}_{\overline{\gamma}}\left(\widetilde{\mu}\right)\right]^2 - 4\frac{\sigma}{\sigma-1}\frac{\xi}{\left(\sigma-1\right)^2}\mathbf{Var}_{\overline{\gamma}}\left(\mu\right).$$

One necessary condition for the exist of solution to (C46) is $\Delta \geq 0$. Notice that $\mathbf{E}_{\overline{\gamma}}\left(\widetilde{\mu}^2\right) \leq \mathbf{E}_{\overline{\gamma}}\left(\widetilde{\mu}\right)$. We have

$$\begin{aligned}
\Delta &= \left(\frac{\sigma}{\sigma-1} - \xi\right)^2 + \frac{2}{\sigma-1}\left(\frac{\sigma}{\sigma-1} - \xi\right)^2\mathbf{E}_{\overline{\gamma}}\left(\widetilde{\mu}\right) + \frac{\left(\frac{\sigma}{\sigma-1} + \xi\right)^2}{\left(\sigma-1\right)^2}\left[\mathbf{E}_{\overline{\gamma}}\left(\widetilde{\mu}\right)\right]^2 \\
&\quad - \frac{2}{\sigma-1}\left(\frac{\sigma}{\sigma-1}\right)^2\frac{2\xi}{\sigma}\mathbf{E}_{\overline{\gamma}}\left(\widetilde{\mu}^2\right) \\
&\geq \Delta_H\left(\mathbf{E}_{\overline{\gamma}}\left(\widetilde{\mu}\right)\right),
\end{aligned} \tag{C47}$$

where we set

$$\Delta_H\left(\mathbf{E}_{\overline{\gamma}}\left(\widetilde{\mu}\right)\right) = \left(\frac{\sigma}{\sigma-1} - \xi\right)^2 - \frac{2}{\sigma-1}\left(\frac{\sigma}{\sigma-1}\right)^2\left[\frac{2\xi}{\sigma} - \left(1 - \frac{\sigma-1}{\sigma}\xi\right)^2\right]\mathbf{E}_{\overline{\gamma}}\left(\widetilde{\mu}\right) + \frac{\left(\frac{\sigma}{\sigma-1} + \xi\right)^2}{\left(\sigma-1\right)^2}\left[\mathbf{E}_{\overline{\gamma}}\left(\widetilde{\mu}\right)\right]^2$$

as a quadratic function of $\mathbf{E}_{\overline{\gamma}}\left(\widetilde{\mu}\right)$.

The minimium value of Δ_H is derived at $\mu^* \equiv \frac{(\sigma-1)\left[\frac{2\sigma-1}{\sigma^2}\xi^2-(1-\xi)^2\right]}{\left(1+\frac{\sigma-1}{\sigma}\xi\right)^2} < 1$. However, $\mathbf{E}_{\overline{\gamma}}\left(\widetilde{\mu}\right) \in (0,1]$, and thus μ^* may not belong to the domain of $\mathbf{E}_{\overline{\gamma}}\left(\widetilde{\mu}\right)$.

If $\mu^* \leq 0$, to prove that $\Delta \geq 0$, we only need to prove that $\Delta_H \geq 0$ when $\mathbf{E}_{\overline{\gamma}}\left(\widetilde{\mu}\right) = 1$ and $\mathbf{E}_{\overline{\gamma}}\left(\widetilde{\mu}\right) = 0$. This is true. According to (C47), when $\mathbf{E}_{\overline{\gamma}}\left(\widetilde{\mu}\right) = 1$, we have

$$\begin{aligned}\Delta_H &= \left(\frac{\sigma}{\sigma-1}-\xi\right)^2 - \frac{2}{\sigma-1}\left(\frac{\sigma}{\sigma-1}\right)^2\left[\frac{2\xi}{\sigma}-\left(1-\frac{\sigma-1}{\sigma}\xi\right)^2\right] + \frac{\left(\frac{\sigma}{\sigma-1}+\xi\right)^2}{(\sigma-1)^2} \\ &= \left(\frac{\sigma}{\sigma-1}\right)^2\left(\frac{\sigma}{\sigma-1}-\xi\right)^2 - \left(\frac{\sigma}{\sigma-1}\right)^2\frac{2}{\sigma-1}\frac{2\xi}{\sigma} + \left(\frac{\sigma}{\sigma-1}\right)^2\frac{2}{\sigma-1}\frac{2\xi}{\sigma} \\ &= \left(\frac{\sigma}{\sigma-1}\right)^2\left(\frac{\sigma}{\sigma-1}-\xi\right)^2 > 0\end{aligned}$$

When $\mathbf{E}_{\overline{\gamma}}\left(\widetilde{\mu}\right) = 0$, we have

$$\Delta_H = \left(\frac{\sigma}{\sigma-1}-\xi\right)^2 > 0.$$

If $\mu^* > 0$ (note that μ^* must be lower than one), to prove $\Delta \geq 0$, we only need to prove that $\Delta_H\left(\mu^*\right) \geq 0$. To see this, first note that when $\mu^* \in (0,1)$, we have

$$\frac{2\sigma-1}{\sigma^2} > \left(\frac{1}{\xi}-1\right)^2$$

and

$$\frac{2}{\sigma-1}\left(\frac{\sigma}{\sigma-1}\right)^2\left[\frac{2\xi}{\sigma}-\left(1-\frac{\sigma-1}{\sigma}\xi\right)^2\right] > 0.$$

We set

$$\widetilde{\Delta} = \left\{\frac{2}{\sigma-1}\left(\frac{\sigma}{\sigma-1}\right)^2\left[\frac{2\xi}{\sigma}-\left(1-\frac{\sigma-1}{\sigma}\xi\right)^2\right]\right\}^2 - 4\left(\frac{\sigma}{\sigma-1}-\xi\right)^2\frac{\left(\frac{\sigma}{\sigma-1}+\xi\right)^2}{(\sigma-1)^2}$$

as the discriminant of $\Delta_H\left(\mathbf{E}_{\overline{\gamma}}\left(\widetilde{\mu}\right)\right) > 0$. $\widetilde{\Delta} < 0$ is a sufficient condition for $\Delta > 0$. To prove $\widetilde{\Delta} < 0$, we only need to show that

$$\frac{1}{\sigma-1}\left(\frac{\sigma}{\sigma-1}\right)^2\left[\frac{2\xi}{\sigma}-\left(1-\frac{\sigma-1}{\sigma}\xi\right)^2\right] < \frac{\left(\frac{\sigma}{\sigma-1}-\xi\right)\left(\frac{\sigma}{\sigma-1}+\xi\right)}{(\sigma-1)},$$

which is equivalent to

$$2\cdot\left(\frac{\sigma}{\sigma-1}\right)^2(1-\xi) > 0.$$

Notice that the above inequality must hold. Thus, we must have $\Delta > 0$. In conclusion, there are two solutions to the quadratic equation (C46). However, $\Delta > 0$ does not necessarily mean that the solutions are all in the domain of μ (i.e., $(1, \frac{\sigma}{\sigma-1}]$).

In the following analysis, we prove that there exist unique solution in the domain of μ. The two potential solutions are μ_1 and μ_2:

$$\begin{aligned}
\mu_1 &= \frac{\left(\frac{\sigma}{\sigma-1}+\xi\right)\int_{\theta_e}\overline{\gamma}(\theta_e)\mu\left(\theta_e\right)d\theta_e}{2\xi}-\frac{\sqrt{\Delta}}{2\xi}\geq\int_{\theta_e}\overline{\gamma}(\theta_e)\mu\left(\theta_e\right)d\theta_e \\
\mu_2 &= \frac{\left(\frac{\sigma}{\sigma-1}+\xi\right)\int_{\theta_e}\overline{\gamma}(\theta_e)\mu\left(\theta_e\right)d\theta_e}{2\xi}+\frac{\sqrt{\Delta}}{2\xi}\leq\frac{\sigma}{\sigma-1}\frac{1}{\xi}\int_{\theta_e}\overline{\gamma}(\theta_e)\mu\left(\theta_e\right)d\theta_e.
\end{aligned}$$

In the following analysis, we prove that $\mu_2>\frac{\sigma}{\sigma-1}$ and $\mu_1\in(1,\frac{\sigma}{\sigma-1}]$. To prove this, we only need to show that

$$\sqrt{\Delta}>\left|2\xi\frac{\sigma}{\sigma-1}-\left(\frac{\sigma}{\sigma-1}+\xi\right)\int_{\theta_e}\overline{\gamma}(\theta_e)\mu\left(\theta_e\right)d\theta_e\right|.$$

In particular, $\mu_2>\frac{\sigma}{\sigma-1}$ is equivalent to

$$\mu_2=\frac{\left(\frac{\sigma}{\sigma-1}+\xi\right)\int_{\theta_e}\overline{\gamma}(\theta_e)\mu\left(\theta_e\right)d\theta_e}{2\xi}+\frac{\sqrt{\Delta}}{2\xi}>\frac{\sigma}{\sigma-1},$$

i.e.,

$$\sqrt{\Delta}>2\xi\frac{\sigma}{\sigma-1}-\left(\frac{\sigma}{\sigma-1}+\xi\right)\int_{\theta_e}\overline{\gamma}(\theta_e)\mu\left(\theta_e\right)d\theta_e.$$

Substituting $\mu\left(\theta_e\right)$ in the above inequality by $\mu\left(\theta_e\right)=1+\frac{1}{\sigma-1}\widetilde{\mu}\left(\theta_e\right)$, we have

$$\begin{aligned}
\sqrt{\Delta} &> 2\xi\frac{\sigma}{\sigma-1}-\left(\frac{\sigma}{\sigma-1}+\xi\right)\left[1+\frac{1}{\sigma-1}\mathbf{E}_{\overline{\gamma}}\left(\widetilde{\mu}\right)\right] \\
&= \left(\frac{\sigma+1}{\sigma-1}\xi-\frac{\sigma}{\sigma-1}\right)-\left(\frac{\sigma}{\sigma-1}+\xi\right)\frac{1}{\sigma-1}\mathbf{E}_{\overline{\gamma}}\left(\widetilde{\mu}\right).
\end{aligned}$$

When $\xi\leq\frac{\sigma}{\sigma+1}$, the above inequality must holds. When $\xi>\frac{\sigma}{\sigma+1}$, to prove the above inequality, we only need to show

$$\Delta-\left[\left(\frac{\sigma+1}{\sigma-1}\xi-\frac{\sigma}{\sigma-1}\right)-\left(\frac{\sigma}{\sigma-1}+\xi\right)\frac{1}{\sigma-1}\mathbf{E}_{\overline{\gamma}}\left(\widetilde{\mu}\right)\right]^2>0.$$

To see this, notice that

$$\begin{aligned}
&\Delta-\left[\left(\frac{\sigma+1}{\sigma-1}\xi-\frac{\sigma}{\sigma-1}\right)-\left(\frac{\sigma}{\sigma-1}+\xi\right)\frac{1}{\sigma-1}E_{\overline{\gamma}}\left(\widetilde{\mu}\right)\right]^2 \\
=&\left(\frac{\sigma}{\sigma-1}-\xi\right)^2+\frac{2}{\sigma-1}\left(\frac{\sigma}{\sigma-1}-\xi\right)^2\mathbf{E}_{\overline{\gamma}}\left(\widetilde{\mu}\right)+\frac{\left(\frac{\sigma}{\sigma-1}+\xi\right)^2}{(\sigma-1)^2}\left[\mathbf{E}_{\overline{\gamma}}\left(\widetilde{\mu}\right)\right]^2 \\
&-\frac{2}{\sigma-1}\left(\frac{\sigma}{\sigma-1}\right)^2\frac{2\xi}{\sigma}\mathbf{E}_{\overline{\gamma}}\left(\widetilde{\mu}^2\right)-\left[\left(\frac{\sigma+1}{\sigma-1}\xi-\frac{\sigma}{\sigma-1}\right)-\left(\frac{\sigma}{\sigma-1}+\xi\right)\frac{1}{\sigma-1}\mathbf{E}_{\overline{\gamma}}\left(\widetilde{\mu}\right)\right]^2 \\
>&\ \Delta_M,
\end{aligned}$$

where

$$\Delta_M=\Delta_H\left(\mathbf{E}_{\overline{\gamma}}\left(\widetilde{\mu}\right)\right)-\left[\left(\frac{\sigma+1}{\sigma-1}\xi-\frac{\sigma}{\sigma-1}\right)-\left(\frac{\sigma}{\sigma-1}+\xi\right)\frac{1}{\sigma-1}\mathbf{E}_{\overline{\gamma}}\left(\widetilde{\mu}\right)\right]^2.$$

Rearranging the right side the the above equation, we have

$$\Delta_M = \frac{2}{\sigma-1}\frac{2\sigma}{\sigma-1}\xi\left(1-\xi\right)\left[1-\mathbf{E}_{\overline{\gamma}}\left(\widetilde{\mu}\right)\right] \geq 0.$$

Since

$$\Delta - \left[\left(\frac{\sigma+1}{\sigma-1}\xi - \frac{\sigma}{\sigma-1}\right) - \left(\frac{\sigma}{\sigma-1}+\xi\right)\frac{1}{\sigma-1}\mathbf{E}_{\overline{\gamma}}\left(\widetilde{\mu}\right)\right]^2 > \Delta_M > 0,$$

we have

$$\mu_2 = \frac{\sigma}{\sigma-1}\frac{1}{\xi}\int_{\theta_e}\overline{\gamma}(\theta_e)\mu\left(\theta_e\right)d\theta_e > \frac{\sigma}{\sigma-1}.$$

We now prove that

$$\frac{\sigma}{\sigma-1} \geq \mu_1 = \frac{\left(\frac{\sigma}{\sigma-1}+\xi\right)\int_{\theta_e}\overline{\gamma}(\theta_e)\mu\left(\theta_e\right)d\theta_e}{2\xi} - \frac{\sqrt{\Delta}}{2\xi} > 1$$

First, $\mu_1 \geq \int_{\theta_e}\overline{\gamma}(\theta_e)\mu\left(\theta_e\right)d\theta_e$, where $\mu\left(\theta_e\right) \in (1, \frac{\sigma}{\sigma-1}]$ and $\overline{\gamma}(\theta_e)$ is a densify with support Θ_e. Thus, $\mu_1 > 1$. Second, notice that

$$\Delta - \left[\left(\frac{\sigma+1}{\sigma-1}\xi - \frac{\sigma}{\sigma-1}\right) - \left(\frac{\sigma}{\sigma-1}+\xi\right)\frac{1}{\sigma-1}\mathbf{E}_{\overline{\gamma}}\left(\widetilde{\mu}\right)\right]^2 \geq 0,$$

(e.g., see (C47)) we have

$$\begin{aligned}\sqrt{\Delta} &\geq \left(\frac{\sigma}{\sigma-1}+\xi\right)\left[1+\frac{1}{\sigma-1}\mathbf{E}_{\overline{\gamma}}\left(\widetilde{\mu}\right)\right]\left(\frac{\sigma}{\sigma-1}+\xi\right) - \frac{2\xi\sigma}{\sigma-1} \\ &= \left(\frac{\sigma}{\sigma-1}+\xi\right)\frac{1}{\sigma-1}\mathbf{E}_{\overline{\gamma}}\left(\widetilde{\mu}\right) - \left(\frac{\sigma+1}{\sigma-1}\xi - \frac{\sigma}{\sigma-1}\right)\end{aligned}$$

The above inequality implies

$$\sqrt{\Delta} \geq \left(\frac{\sigma}{\sigma-1}+\xi\right)\int_{\theta_e}\overline{\gamma}(\theta_e)\mu\left(\theta_e\right)d\theta_e - \frac{2\xi\sigma}{\sigma-1},$$

which gives

$$\mu_1 = \frac{\left(\frac{\sigma}{\sigma-1}+\xi\right)\int_{\theta_e}\overline{\gamma}(\theta_e)\mu\left(\theta_e\right)d\theta_e}{2\xi} - \frac{\sqrt{\Delta}}{2\xi} \leq \frac{\sigma}{\sigma-1}.$$

In conclusion, we have

$$\mu = \frac{\left(\frac{\sigma}{\sigma-1}+\xi\right)\int_{\theta_e}\overline{\gamma}(\theta_e)\mu\left(\theta_e\right)d\theta_e}{2\xi} - \frac{\sqrt{\Delta}}{2\xi}.$$

■

C.4 Proof of Proposition 5

(i) Rewriting the general tax formula for the entrepreneurs in equation (42) for the case with uniform markups, we have part one of Proposition 5. By the definitions of $\varepsilon^{\pi}_{1-\tau_e}\left(\theta_e\right)$ and $\varepsilon_{Q_{-ij}}\left(\theta_e\right)$ (i.e., (40) and

(39)), we have

$$\frac{1}{1-\tau_e(\theta_e)} = \frac{1+[1-\bar{g}_e(\theta_e)]\,H(\theta_e)\left[\frac{1+\varepsilon_e}{\varepsilon_e}[\mu-\xi]-1\right]}{\mu} + \frac{\frac{\sigma}{\sigma-1}}{\frac{\sigma}{\sigma-1}-\xi}\left[\frac{\sigma-1}{\sigma}-\frac{1}{\mu}\right]\left\{[1-g_e(\theta_e)]-[1-\bar{g}_e(\theta_e)]\,H(\theta_e)\right\}.$$

(ii) Supposing $g_o(\cdot)$ (such that $\bar{g}_o(\cdot)$) is exogenous, we have

$$\begin{aligned}
\frac{d\left[\frac{1}{1-\tau_e(\theta_e)}\right]}{d\mu} &= \frac{1}{\mu^2}\left[\begin{array}{c}\mu\,[1-\bar{g}_e(\theta_e)]\,H(\theta_e)\frac{\varepsilon_e+1}{\varepsilon_e}-[1-\bar{g}_e(\theta_e)]\,H(\theta_e)\left[(\mu-\xi)\frac{\varepsilon_e+1}{\varepsilon_e}-1\right]\\ +\frac{\frac{\sigma}{\sigma-1}}{\frac{\sigma}{\sigma-1}-\xi}\left\{[1-g_e(\theta_e)]-[1-\bar{g}_e(\theta_e)]\,H(\theta_e)\right\}-1\end{array}\right]\\
&= \frac{1}{\mu^2}\left[\begin{array}{c}[1-\bar{g}_e(\theta_e)]\,H(\theta_e)\left[\xi\frac{\varepsilon_e+1}{\varepsilon_e}+1\right]\\ +\frac{\frac{\sigma}{\sigma-1}}{\frac{\sigma}{\sigma-1}-\xi}\left\{[1-g_e(\theta_e)]-[1-\bar{g}_e(\theta_e)]\,H(\theta_e)\right\}-1\end{array}\right]\\
&= \frac{1}{\mu^2}\left[\begin{array}{c}[1-\bar{g}_e(\theta_e)]\,H(\theta_e)\xi\left[\frac{\varepsilon_e+1}{\varepsilon_e}-\frac{1}{\frac{\sigma}{\sigma-1}-\xi}\right]\\ +[1-g_e(\theta_e)]\frac{\frac{\sigma}{\sigma-1}}{\frac{\sigma}{\sigma-1}-\xi}-1\end{array}\right]
\end{aligned}$$

and

$$\begin{aligned}
\frac{d\left[\frac{1-\tau_w(\theta_w)}{1-\tau_e(\theta_e)}\right]}{d\mu} &= \frac{1-\tau_w(\theta_w)}{\mu}\frac{d\left[\frac{\mu}{1-\tau_e(\theta_e)}\right]}{d\mu}\\
&= \frac{1-\tau_w(\theta_w)}{\mu}\left[\begin{array}{c}[1-\bar{g}_e(\theta_e)]\,H(\theta_e)\frac{\varepsilon_e+1}{\varepsilon_e}+\\ \frac{1}{\frac{\sigma}{\sigma-1}-\xi}\left\{[1-g_e(\theta_e)]-[1-\bar{g}_e(\theta_e)]\,H(\theta_e)\right\}\end{array}\right].
\end{aligned}$$

According to the above equations, $\tau_e(\theta_e)$ increases in μ iff

$$[1-\bar{g}_e(\theta_e)]\,H(\theta_e)\xi\left[\frac{\varepsilon_e+1}{\varepsilon_e}-\frac{1}{\frac{\sigma}{\sigma-1}-\xi}\right]+[1-g_e(\theta_e)]\frac{\frac{\sigma}{\sigma-1}}{\frac{\sigma}{\sigma-1}-\xi}-1>0,$$

i.e.,

$$\begin{aligned}
g_e(\theta_e) &< 1-\frac{1-[1-\bar{g}_e(\theta_e)]\,H(\theta_e)\xi\left[\frac{\varepsilon_e+1}{\varepsilon_e}-\frac{1}{\frac{\sigma}{\sigma-1}-\xi}\right]}{\frac{\frac{\sigma}{\sigma-1}}{\frac{\sigma}{\sigma-1}-\xi}} \qquad \text{(C48)}\\
&= \frac{\xi(\sigma-1)}{\sigma}\left\{1+[1-\bar{g}_e(\theta_e)]\,H(\theta_e)\left[\frac{\varepsilon_e+1}{\varepsilon_e}\left(\frac{\sigma}{\sigma-1}-\xi\right)-1\right]\right\}.
\end{aligned}$$

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

$$[1-\bar{g}_e(\theta_e)]\,H(\theta_e)\frac{\varepsilon_e+1}{\varepsilon_e}+\frac{1}{\frac{\sigma}{\sigma-1}-\xi}\left\{[1-g_e(\theta_e)]-[1-\bar{g}_e(\theta_e)]\,H(\theta_e)\right\}>0,$$

i.e.,

$$g_e(\theta_e) < 1 + [1 - \bar{g}_e(\theta_e)]\, H(\theta_e) \left[\frac{\varepsilon_e + 1}{\varepsilon_e} \left(\frac{\sigma}{\sigma - 1} - \xi \right) - 1 \right], \tag{C49}$$

or equivalently

$$g_e(\theta_e) - 1 + [1 - \bar{g}_e(\theta_e)]\, H(\theta_e) < [1 - \bar{g}_e(\theta_e)]\, H(\theta_e) \frac{\varepsilon_e + 1}{\varepsilon_e} \left(\frac{\sigma}{\sigma - 1} - \xi \right),$$

which must be true when $[1 - g_e(\theta_e)] - [1 - \bar{g}_e(\theta_e)]\, H(\theta_e) > 0$ and $H(\theta_e) > 0$.

(iii) When $g_e(\theta_e) = \bar{g}_e(\theta_e)$ and $H(\theta_e) > 0$, inequality (C48) is equivalent to

$$g_e(\theta_e) < \frac{\frac{\xi(\sigma-1)}{\sigma} \left\{ 1 + H(\theta_e) \left[\frac{\varepsilon_e+1}{\varepsilon_e} \left(\frac{\sigma}{\sigma-1} - \xi \right) - 1 \right] \right\}}{1 + \frac{\xi(\sigma-1)}{\sigma} H(\theta_e) \left[\frac{\varepsilon_e+1}{\varepsilon_e} \left(\frac{\sigma}{\sigma-1} - \xi \right) - 1 \right]},$$

where $\frac{1 + H(\theta_e) \left[\frac{\varepsilon_e+1}{\varepsilon_e} \left(\frac{\sigma}{\sigma-1} - \xi \right) - 1 \right]}{\frac{\sigma}{\xi(\sigma-1)} + H(\theta_e) \left[\frac{\varepsilon_e+1}{\varepsilon_e} \left(\frac{\sigma}{\sigma-1} - \xi \right) - 1 \right]}$ increases in $H(\theta_e) \left[\frac{\varepsilon_e+1}{\varepsilon_e} \left(\frac{\sigma}{\sigma-1} - \xi \right) - 1 \right]$, because $\frac{\sigma}{\xi(\sigma-1)} > 1$. Besides, under condition (29), we have $\frac{\varepsilon_e+1}{\varepsilon_e} \left(\frac{\sigma}{\sigma-1} - \xi \right) - 1 > 0$ and that $H(\theta_e) \left[\frac{\varepsilon_e+1}{\varepsilon_e} \left(\frac{\sigma}{\sigma-1} - \xi \right) - 1 \right]$ increases in $H(\theta_e)$. Thus, when $H(\theta_e) > 0$, the above inequality holds if

$$g_e(\theta_e) \leq \frac{\xi\, (\sigma - 1)}{\sigma}.$$

Analogously, when $g_e(\theta_e) = \bar{g}_e(\theta_e) < 1$ and $H(\theta_e) > 0$, inequality (C49) must holds.■

C.5 Proof of Corollary 2

(i) When firm-level markup is constant, we have $H\left(\theta_e\right) = \frac{1 - F_e(\theta_e)}{f_e(\theta_e)} \frac{d \ln X(\theta_e)}{d\theta_e} \frac{\frac{1+\varepsilon_e}{\varepsilon_e}}{\frac{1+\varepsilon_e}{\varepsilon_e} \left(1 - \xi \frac{\sigma-1}{\sigma} \right) - \frac{\sigma-1}{\sigma}}$, which is constant if $\frac{1 - F_e(\theta_e)}{f_e(\theta_e)} \frac{d \ln X(\theta_e)}{d\theta_e}$ is constant. Thus $H\left(\theta_e\right)$ is constant for $\theta_e \geq \theta_e^*$. On the other hand, notice that if for any $\theta_e \geq \theta_e^\star$, $g_e(\theta_e) = \widehat{g}_e$ is constant, for any $\theta_e \geq \theta_e^\star$, $g_e(\theta_e)$ is also constant and equal to $\widehat{g}_e$. With the above two findings, part one of Corollary 2 follow directly from Theorem 1.

(ii) To prove part two of Corollary 2, first note that $\frac{1}{1-\widehat{\tau}_e}$ and thus $\widehat{\tau}_e$ increases with μ. Notice that for any $\theta_e \in \Theta_e$, $\mu\left(\theta_e\right)$ non-decreases with the decrease of I, μ as a weighted average of $\mu\left(\theta_e\right)$ must non-decreases with the decrease of I. Thus a condition guarantees that, given μ, $\frac{1}{1-\widehat{\tau}_e}$ increases with $\widehat{\mu}$ is a sufficient condition for $\widehat{\tau}_e$ to be increasing with the decrease of I. We now show that (C50) is such a condition.

To see this, we first treat $\widehat{g}_e$ and μ as given and take the derivative of $\frac{1}{1-\widehat{\tau}_e}$ with respect to $\widehat{\mu}$:

$$\begin{aligned} \frac{d\left(\frac{1}{1-\widehat{\tau}_e}\right)}{d\widehat{\mu}} \quad = \quad & \frac{-\frac{1}{\widehat{\mu}^2} \left[1 + \left(1 - \widehat{g}_e\right) \widehat{H} \frac{1}{\widehat{\varepsilon}^{\pi}_{1-\tau_e}} \right] + \frac{1}{\widehat{\mu}} \left(1 - \widehat{g}_e\right) \widehat{H} \frac{1+\varepsilon_e}{\varepsilon_e} + \frac{\frac{\sigma}{\sigma-1} \frac{1}{\widehat{\mu}^2}}{\frac{\sigma}{\sigma-1} - \xi} \left(1 - \widehat{g}_e\right) \left(1 - \widehat{H}\right)}{1 - \frac{\xi}{\frac{\sigma}{\sigma-1} - \xi} \frac{\mu - \widehat{\mu}}{\widehat{\mu}}} \\ & - \frac{\frac{\xi}{\frac{\sigma}{\sigma-1} - \xi} \frac{\mu}{\widehat{\mu}^2} \left[\frac{1}{\widehat{\mu}} \left[1 + \left(1 - \widehat{g}_e\right) \widehat{H} \frac{1}{\widehat{\varepsilon}^{\pi}_{1-\tau_e}} \right] + \frac{1 - \frac{\sigma}{\sigma-1} \frac{1}{\widehat{\mu}}}{\frac{\sigma}{\sigma-1} - \xi} \left(1 - \widehat{g}_e\right) \left(1 - \widehat{H}\right) \right]}{\left[1 - \frac{\xi}{\frac{\sigma}{\sigma-1} - \xi} \frac{\mu - \widehat{\mu}}{\widehat{\mu}} \right]^2} \end{aligned}$$

Rearranging the right side of the above equation, we find that $\frac{d}{d\widehat{\mu}}\left(\frac{1}{1-\widehat{\tau}_e}\right) > 0$ if and only if

$$\left(1-\widehat{g}_e\right)\widehat{H}\left[\begin{array}{c}\left(\frac{\sigma}{\sigma-1}-\xi\right)\left[\left(\widehat{\mu}-\xi\mu\frac{\sigma-1}{\sigma}\right)\frac{1+\varepsilon_e}{\varepsilon_e}-\frac{1}{\widehat{\varepsilon}^{\pi}_{1-\tau_e}}\right]\\+\left(\frac{\sigma}{\sigma-1}-\xi\mu\frac{\sigma-1}{\sigma}\right)\end{array}\right] > \frac{\sigma}{\sigma-1}-\xi.$$

Note that the term in the bracket is positive, because for any $\theta_e \in \Theta_e$, $\mu\left(\theta_e\right) \leq \frac{\sigma}{\sigma-1}$. Thus, we have

$$1-\widehat{g}_e > \frac{\frac{\sigma}{\sigma-1}-\xi}{\widehat{H}\left[\begin{array}{c}\left(\frac{\sigma}{\sigma-1}-\xi\right)\left[\left(\widehat{\mu}-\xi\mu\frac{\sigma-1}{\sigma}\right)\frac{1+\varepsilon_e}{\varepsilon_e}-\frac{1}{\widehat{\varepsilon}^{\pi}_{1-\tau_e}}\right]\\+\left(\frac{\sigma}{\sigma-1}-\xi\mu\frac{\sigma-1}{\sigma}\right)\end{array}\right]}.$$

Dividing both the numerator and denominator of the right side of the above inequality by $\frac{\sigma}{\sigma-1}-\xi$ and substituting $\frac{1}{\widehat{\varepsilon}^{\pi}_{1-\tau_e}}$ by $\frac{1+\varepsilon_e}{\varepsilon_e}\left(\widehat{\mu}-\xi\right)-1$, we have

$$\widehat{g}_e < 1-\frac{1}{\left[\xi\left(1-\mu\frac{\sigma-1}{\sigma}\right)\frac{1+\varepsilon_e}{\varepsilon_e}+1+\frac{\frac{\sigma}{\sigma-1}-\xi\mu\frac{\sigma-1}{\sigma}}{\frac{\sigma}{\sigma-1}-\xi}\right]\widehat{H}}. \tag{C50}$$

(C50). Notice that $\mu \leq \frac{\sigma}{\sigma-1}$, the term in the bracket of the right side of (53) is not less than 2. Thus, condition (53) satisfies if $\widehat{g}_e < 1-\frac{1}{2\widehat{H}}$.

Last, since for any $I > 1$, (54) is a sufficient condition for $\widehat{\tau}_e$ to be increasing with the marginal decrease of I (just suppose that I is continuous), and $\widehat{\tau}_e$ is twice continuously differentiable with respect to I, according to the mean value theorem, (54) is also a sufficient condition for $\widehat{\tau}_e$ to be increasing with the decrease of I from $n+1$ to n, where $n \in \mathbb{N}_+$.■

D Discussion and Robustness

D.1 Proof of Theorem 2

We prove Theorem 2 following the Lagrangian problem presented in Appendix C.1.1. Note that the expression of markup (20) is now generalized to be $\mu(\theta_e) = \frac{P(\theta_e)\frac{\partial Q_{ij}(\theta_e)}{\partial L_w(\theta_e)}[1-\tau_s(\theta_e)]}{W}$. By the definition of $\varpi(\theta_e)$ (e.g., see (38)), we have $\varpi(\theta_e) = P\left(Q_{ij}(\theta_e), \theta_e\right)\left[1+\varepsilon_{Q_{ij}}(\theta_e)\right]\frac{\partial Q_{ij}(\theta_e)}{\partial L_w(\theta_e)}$. Notice that $1-\tau_s(\theta_e) = \frac{W}{\varpi(\theta_e)}$, as in the benchmark model, we hvae $\mu(\theta_e) = \frac{1}{1+\varepsilon_{Q_{ij}}(\theta_e)}$.

(i) When the uniform restriction on $\varpi(\theta_e, \theta_e l_e(\theta_e), L_w(\theta_e), Q)$ is loosened and $\varphi(\theta_e) = 0$. Under this case, according to the expression of $\frac{\partial \pounds}{\partial L_w(\theta_e)}$ (e.g., (C7)), we have

$$\begin{aligned} P(\theta_e)\frac{\partial Q_{ij}(\theta_e)}{\partial L_w(\theta_e)} &= \frac{\lambda'}{\lambda} - \frac{\kappa'(\theta_e)}{\lambda L_w(\theta_e) N f_e(\theta_e)}\frac{\partial \ln Q_{ij}(\theta_e)}{\partial \ln L_w(\theta_e)} \\ &= \frac{\lambda'}{\lambda} - \frac{\kappa'(\theta_e)\xi}{\lambda L_w(\theta_e) N f_e(\theta_e)}. \end{aligned} \tag{D2}$$

Dividing both sides by $\frac{\varepsilon^{\varpi}_{L_w}(\theta_e)}{L_w(\theta_e) N f_e(\theta_e)}$ and integrating across θ_e gives

$$\int_{\theta_e} P(\theta_e) Q_{ij}(\theta_e)\xi\frac{N f_e(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)}d\theta_e = \frac{\lambda'}{\lambda}\int_{\theta_e}\frac{L_w(\theta_e) N f_e(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)}d\theta_e - \int_{\theta_e}\frac{\kappa'(\theta_e)}{\lambda\varepsilon^{\varpi}_{L_w}(\theta_e)}\xi d\theta_e,$$

Using $\kappa(\underline{\theta}_e) = \kappa(\overline{\theta}_e) = 0$ and integration by parts, we have

$$\begin{aligned} \frac{\lambda'}{\lambda W} &= \frac{\xi\int_{\theta_e} P(\theta_e) Q_{ij}(\theta_e) N f_e(\theta_e) d\theta_e}{W\int_{\theta_e} L_w(\theta_e) N f_e(\theta_e) d\theta_e} \\ &= \frac{\xi\int_{\theta_e}\frac{P(\theta_e)Q_{ij}(\theta_e)}{W L_w(\theta_e)} W L_w(\theta_e) N f_e(\theta_e) d\theta_e}{\int_{\theta_e} W L_w(\theta_e) N f_e(\theta_e) d\theta_e} \\ &= \frac{\xi\int_{\theta_e}\frac{\mu(\theta_e)}{\xi[1-\tau^E_s(\theta_e)]} W L_w(\theta_e) N f_e(\theta_e) d\theta_e}{\int_{\theta_e} W L_w(\theta_e) N f_e(\theta_e) d\theta_e} \\ &= \widetilde{\mu}. \end{aligned}$$

where the third equation is derived by the generalized definition of markup (i.e., $\mu(\theta_e) = \frac{P(\theta_e)\frac{\partial Q_{ij}(\theta_e)}{\partial L_w(\theta_e)}[1-\tau_s(\theta_e)]}{W}$).

By the definition of tax wedges (e.g., see (37)), we can substitute $P(\theta_e)\frac{\partial Q_{ij}(\theta_e)}{\partial L_w(\theta_e)}$ in (D2) with $\frac{W\mu(\theta_e)}{1-\tau_s^E(\theta_e)}$:

$$\begin{aligned}
\frac{1}{1-\tau_s^E(\theta_e)} &= \frac{\lambda'}{\lambda W\mu(\theta_e)} - \frac{\kappa'(\theta_e)\xi}{\lambda L_w(\theta_e) N f_e(\theta_e) W\mu(\theta_e)} \\
&= \frac{\widetilde{\mu}}{\mu(\theta_e)} - \frac{\kappa'(\theta_e)\xi}{\lambda L_w(\theta_e) N f_e(\theta_e) W\mu(\theta_e)} \\
&= \frac{\widetilde{\mu}}{\mu(\theta_e)} + \frac{P(\theta_e) Q_{ij}(\theta_e)\left[1-\tau_s^E(\theta_e)\right]\xi\left[1-\tau_e^E(\theta_e)\right]\varepsilon_{Q_{-ij}}(\theta_e)\left[\begin{array}{c}\psi_e(\theta_e)\frac{1+\varepsilon_e}{\varepsilon_e}\frac{l_e'(\theta_e)}{l_e(\theta_e)}+ \\ \psi_e'(\theta_e)+\psi_e(\theta_e)\frac{d\ln\left[\mu(\theta_e)\varepsilon_{Q_{-ij}}(\theta_e)\right]}{d\theta_e}\end{array}\right]}{L_w(\theta_e) W\mu(\theta_e)\lambda N f_e(\theta_e)} \\
&= \frac{\widetilde{\mu}}{\mu(\theta_e)} + \left[1-\tau_e^E(\theta_e)\right]\varepsilon_{Q_{-ij}}(\theta_e)\left[\begin{array}{c}\left[1-g_e(\theta_e)\right]- \\ \frac{[1-\bar{g}_e(\theta_e)][1-F_e(\theta_e)]}{f_e(\theta_e)}\left[\frac{1+\varepsilon_e}{\varepsilon_e}\frac{l_e'(\theta_e)}{l_e(\theta_e)}+\frac{d\ln\left[\mu(\theta_e)\varepsilon_{Q_{-ij}}(\theta_e)\right]}{d\theta_e}\right]\end{array}\right],
\end{aligned} \tag{D3}$$

where the third equation is derived by the modified (C16):

$$\begin{aligned}
\kappa'(\theta_e) &= -\frac{d\left[\psi_e(\theta_e)\phi_e'(l_e(\theta_e)) l_e(\theta_e)\widehat{\mu}(\theta_e)\varepsilon_{Q_{-ij}}(\theta_e)\right]}{d\theta_e} \\
&= -\left[\begin{array}{c}\psi_e'(\theta_e)\phi_e'(l_e(\theta_e)) l_e(\theta_e)\widehat{\mu}(\theta_e)\varepsilon_{Q_{-ij}}(\theta_e)+ \\ \psi_e(\theta_e)\phi_e'(l_e(\theta_e))\frac{1+\varepsilon_e}{\varepsilon_e} l_e'(\theta_e)\widehat{\mu}(\theta_e)\varepsilon_{Q_{-ij}}(\theta_e)+ \\ \psi_e(\theta_e)\phi_e'(l_e(\theta_e)) l_e(\theta_e)\frac{d\ln\left[\widehat{\mu}(\theta_e)\varepsilon_{Q_{-ij}}(\theta_e)\right]}{d\theta_e}\end{array}\right] \\
&= -\phi_e'(l_e(\theta_e)) l_e(\theta_e)\widehat{\mu}(\theta_e)\varepsilon_{Q_{-ij}}(\theta_e)\left[\begin{array}{c}\psi_e(\theta_e)\frac{1+\varepsilon_e}{\varepsilon_e}\frac{l_e'(\theta_e)}{l_e(\theta_e)}+ \\ \psi_e'(\theta_e)+\psi_e(\theta_e)\frac{d\ln\left[\widehat{\mu}(\theta_e)\varepsilon_{Q_{-ij}}(\theta_e)\right]}{d\theta_e}\end{array}\right]
\end{aligned} \tag{D4}$$

and the tax wedge, $1-\tau_e^E(\theta_e) = \frac{\phi_e'(l_e(\theta_e))}{\frac{P(\theta_e)}{\mu(\theta_e)}\frac{\partial Q_{ij}(\theta_e)}{\partial l_e(\theta_e)}\left[1-\tau_s^E(\theta_e)\right]}$ (e.g., see (37)). These two equations implies

$$\kappa'(\theta_e) = -P(\theta_e) Q_{ij}(\theta_e)\left[1-\tau_e^E(\theta_e)\right]\left[1-\tau_s^E(\theta_e)\right]\varepsilon_{Q_{-ij}}(\theta_e)\left[\begin{array}{c}\psi_e(\theta_e)\frac{1+\varepsilon_e}{\varepsilon_e}\frac{l_e'(\theta_e)}{l_e(\theta_e)}+ \\ \psi_e'(\theta_e)+\psi_e(\theta_e)\frac{d\ln\left[\mu(\theta_e)\varepsilon_{Q_{-ij}}(\theta_e)\right]}{d\theta_e}\end{array}\right].$$

The last equation of (D3) is derived by (C12), (C13) and $1-\tau_s^E(\theta_e) = \frac{W\mu(\theta_e)}{P(\theta_e)\frac{\partial Q_{ij}(\theta_e)}{\partial L_w(\theta_e)}} = \frac{WL_w(\theta_e)\mu(\theta_e)}{\xi P(\theta_e) Q_{ij}(\theta_e)}$.

(ii) According to the expression of $\frac{\partial \pounds}{\partial l_w(\theta_w)}$ (e.g., (C6)), we have

$$\frac{1}{\frac{\phi_w'(l_w(\theta_w))}{x_w(\theta_w)}} = \frac{\lambda}{\lambda'}\left[1-\frac{x_w'(\theta_w)}{x_w(\theta_w)}\frac{\psi_w(\theta_w)}{\lambda N_w f_w(\theta_w)}\frac{1+\varepsilon_w}{\varepsilon_w}\right]. \tag{D5}$$

Substituting $\frac{\phi_w'(l_w(\theta_w))}{x_w(\theta_w)}$ by $\left[1-\tau_w^E(\theta_w)\right]W$ gives

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

(iii) According to the expression of $\frac{\partial \pounds}{\partial l_e(\theta_e)}$ (e.g., (C8)), we have

$$\frac{\widetilde{\tau}_e(\theta_e)}{1-\widetilde{\tau}_e(\theta_e)} = [1-\bar{g}_e(\theta_e)]\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu(\theta_e)\frac{\partial \ln P(Q(\theta_e),\theta_e)}{\partial\theta_e}+\frac{x_e'(\theta_e)}{x_e(\theta_e)}\right] - \frac{1}{1-\widetilde{\tau}_e(\theta_e)}\frac{WL_w(\theta_e)}{P(\theta_e)Q_{ij}(\theta_e)}\frac{1}{\xi}\frac{\kappa'(\theta_e)\xi}{\lambda N_e f_e(\theta_e)WL_w(\theta_e)}.$$

Reminber that $1-\widetilde{\tau}_e(\theta_e)=\frac{\phi_e'(l_e(\theta_e))}{P(\theta_e)\frac{\partial Q_{ij}(\theta_e)}{\partial l_e(\theta_e)}}=\frac{\left[1-\tau_e^E(\theta_e)\right]\left[1-\tau_s^E(\theta_e)\right]}{\mu(\theta_e)}$.

Using $\frac{\mu(\theta_e)}{1-\tau_s^E(\theta_e)}=\widetilde{\mu}-\frac{\kappa'(\theta_e)\xi}{\lambda WL_w(\theta_e)Nf_e(\theta_e)}$ to substitute $\frac{\kappa'(\theta_e)\xi}{\lambda WL_w(\theta_e)Nf_e(\theta_e)}$ and $\frac{\xi\left[1-\tau_s^E(\theta_e)\right]}{\mu(\theta_e)}$ to substitute $\frac{WL_w(\theta_e)}{P(\theta_e)Q_{ij}(\theta_e)}$, we have

$$\frac{1}{1-\widetilde{\tau}_e(\theta_e)} = 1+[1-\bar{g}_e(\theta_e)]\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu(\theta_e)\frac{\partial \ln P(Q(\theta_e),\theta_e)}{\partial\theta_e}+\frac{x_e'(\theta_e)}{x_e(\theta_e)}\right] + \frac{1}{1-\widetilde{\tau}_e(\theta_e)}\left[1-\widetilde{\mu}\frac{1-\tau_s^E(\theta_e)}{\mu(\theta_e)}\right]. \tag{D6}$$

Lastly, substituting $1-\widetilde{\tau}_e(\theta_e)$ with $\frac{\left[1-\tau_e^E(\theta_e)\right]\left[1-\tau_s^E(\theta_e)\right]}{\mu(\theta_e)}$, we have (60).■

D.2 Supplement to Theorem 2: A More Explicit Expression

According to the definitions of tax wedges, we have

$$\frac{\xi\left[1-\tau_s^E(\theta_e)\right]}{\mu(\theta_e)}P(\theta_e)Q(\theta_e)=WL_w(\theta_e) \tag{D7}$$

and

$$\frac{P(\theta_e)Q_{ij}(\theta_e)\left[1-\tau_s^E(\theta_e)\right]}{\mu(\theta_e)}\left[1-\tau_e^E(\theta_e)\right]=l_e(\theta_e)^{1+\frac{1}{\varepsilon_e}}. \tag{D8}$$

Substituting $P(\theta_e)$ and $Q(\theta_e)$ in (D7) by (A1) and $Q_{ij}(\theta_e)=x_e(\theta_e)l_e(\theta_e)L_w(\theta_e)^{\xi}$, we have

$$\left(\left[1-\tau_s^E(\theta_e)\right]\frac{\chi(\theta_e)x_e(\theta_e)^{1-\frac{1}{\sigma}}}{\mu(\theta_e)}\right)\frac{\xi A^{\frac{\sigma-1}{\sigma}}Q^{\frac{1}{\sigma}}}{N^{\frac{1}{\sigma}}W}l_e(\theta_e)^{1-\frac{1}{\sigma}}L_w(\theta_e)^{\xi\frac{\sigma-1}{\sigma}}=L_w(\theta_e)$$

and therefore

$$L_w(\theta_e)=\left(\frac{X(\theta_e)\left[1-\tau_s^E(\theta_e)\right]}{\mu(\theta_e)}\right)^{\frac{1}{1-\xi\frac{\sigma-1}{\sigma}}}l_e(\theta_e)^{\frac{1-\frac{1}{\sigma}}{1-\xi\frac{\sigma-1}{\sigma}}}\left(\frac{\xi A^{\frac{\sigma-1}{\sigma}}Q^{\frac{1}{\sigma}}}{WN^{\frac{1}{\sigma}}}\right)^{\frac{1}{1-\xi\frac{\sigma-1}{\sigma}}}$$

On the other hand, combining the two first order conditions delivers

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

Combining the above two equations, we can solve for $L_w(\theta_e)$ and $l_e(\theta_e)$ and derive

$$\frac{d\ln L_w(\theta_e)}{d\theta_e} = \frac{\frac{d}{d\theta_e}\left[\ln\frac{X(\theta_e)\left[1-\tau_s^E(\theta_e)\right]}{\mu(\theta_e)}\right]}{1-\xi\frac{\sigma-1}{\sigma}} + \frac{\frac{\sigma-1}{\sigma}}{1-\xi\frac{\sigma-1}{\sigma}}\frac{l_e'(\theta_e)}{l_e(\theta_e)}$$

and

$$\frac{d\ln l_e(\theta_e)}{d\theta_e} = \frac{\frac{d}{d\theta_e}\left[\ln\frac{X(\theta_e)\left[1-\tau_s^E(\theta_e)\right]}{\mu(\theta_e)}\right] + \left[1-\xi\frac{\sigma-1}{\sigma}\right]\frac{d\ln\left[1-\tau_e^E(\theta_e)\right]}{d\theta_e}}{\frac{\varepsilon_e+1}{\varepsilon_e} - \frac{\sigma-1}{\sigma}\left(1+\frac{\varepsilon_e+1}{\varepsilon_e}\xi\right)} \tag{D9}$$

Comparing the above two equations to (C37) and (C39), one can see that they are the same except that now $X(\theta_e)$ is replaced by $X(\theta_e)\left[1-\tau_s^E(\theta_e)\right]$. In addition, we have

$$\begin{aligned}
& \mu(\theta_e)\frac{\partial\ln P(Q(\theta_e),\theta_e)}{\partial\theta_e} + \frac{x_e'(\theta_e)}{x_e(\theta_e)} \\
= \;& \mu(\theta_e)\frac{\chi'(\theta_e)}{\chi(\theta_e)} + \frac{x_e'(\theta_e)}{x_e(\theta_e)} + \mu(\theta_e)\varepsilon_{Q_{-ij}}(\theta_e)\frac{d\ln Q_{ij}(\theta_e)}{d\theta_e} \\
= \;& \mu(\theta_e)\frac{\chi'(\theta_e)}{\chi(\theta_e)} + \left[1+\mu(\theta_e)\varepsilon_{Q_{-ij}}(\theta_e)\right]\frac{x_e'(\theta_e)}{x_e(\theta_e)} + \frac{\mu(\theta_e)\varepsilon_{Q_{-ij}}(\theta_e)}{1-\xi\frac{\sigma-1}{\sigma}}\frac{d\ln l_e(\theta_e)}{d\theta_e} \\
& + \frac{\xi\mu(\theta_e)\varepsilon_{Q_{-ij}}(\theta_e)}{1-\xi\frac{\sigma-1}{\sigma}}\frac{d\left[\ln\frac{X(\theta_e)}{\mu(\theta_e)} + \ln\left[1-\tau_s^E(\theta_e)\right]\right]}{d\theta_e},
\end{aligned}$$

where

$$\begin{aligned}
& \mu(\theta_e)\frac{\chi'(\theta_e)}{\chi(\theta_e)} + \left[1+\mu(\theta_e)\varepsilon_{Q_{-ij}}(\theta_e)\right]\frac{x_e'(\theta_e)}{x_e(\theta_e)} \\
= \;& \mu(\theta_e)\left[\frac{\chi'(\theta_e)}{\chi(\theta_e)} + \frac{\sigma-1}{\sigma}\frac{x_e'(\theta_e)}{x_e(\theta_e)}\right] \\
= \;& \mu(\theta_e)\frac{d\ln\frac{X(\theta_e)}{\mu(\theta_e)}}{d\theta_e} + \frac{d\mu(\theta_e)}{d\theta_e}.
\end{aligned}$$

Therefore, we have

$$
\begin{aligned}
& \mu\left(\theta_e\right) \frac{\partial \ln P\left(Q\left(\theta_e\right), \theta_e\right)}{\partial \theta_e}+\frac{x_e^{\prime}\left(\theta_e\right)}{x_e\left(\theta_e\right)} \\
= & \frac{\mu\left(\theta_e\right) \varepsilon_{Q_{-ij}}(\theta_e)}{1-\xi \frac{\sigma-1}{\sigma}} \frac{d \ln l_e\left(\theta_e\right)}{d \theta_e}+\frac{\mu\left(\theta_e\right)-\xi}{1-\xi \frac{\sigma-1}{\sigma}} \frac{d \ln \left[X\left(\theta_e\right) / \mu\left(\theta_e\right)\right]}{d \theta_e} \\
& +\frac{\xi \mu\left(\theta_e\right) \varepsilon_{Q_{-ij}}(\theta_e)}{1-\xi \frac{\sigma-1}{\sigma}} \frac{d \ln \left[1-\tau_s^E\left(\theta_e\right)\right]}{d \theta_e}+\frac{d \mu\left(\theta_e\right)}{d \theta_e} \\
= & \left[\left(\mu\left(\theta_e\right)-\xi\right) \frac{\varepsilon_e+1}{\varepsilon_e}-1\right] \frac{\frac{d \ln \left[X\left(\theta_e\right) / \mu\left(\theta_e\right)\right]}{d \theta_e}}{\frac{\varepsilon_e+1}{\varepsilon_e}-\frac{\sigma-1}{\sigma}\left(1+\frac{\varepsilon_e+1}{\varepsilon_e} \xi\right)} \\
& +\mu\left(\theta_e\right) \varepsilon_{Q_{-ij}}(\theta_e) \frac{\left(1+\xi \frac{\varepsilon_e+1}{\varepsilon_e}\right) \frac{d \ln \left[1-\tau_s^E\left(\theta_e\right)\right]}{d \theta_e}+\frac{d \ln \left[1-\tau_e^E\left(\theta_e\right)\right]}{d \theta_e}}{\frac{\varepsilon_e+1}{\varepsilon_e}-\frac{\sigma-1}{\sigma}\left(1+\frac{\varepsilon_e+1}{\varepsilon_e} \xi\right)}+\frac{d \mu\left(\theta_e\right)}{d \theta_e},
\end{aligned}
$$

where the second equation is derived by (D9). In addition, we have

$$
\begin{aligned}
& \frac{1-F_e(\theta_e)}{f_e(\theta_e)} \frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu\left(\theta_e\right) \frac{\partial \ln P\left(Q\left(\theta_e\right), \theta_e\right)}{\partial \theta_e}+\frac{x_e^{\prime}\left(\theta_e\right)}{x_e\left(\theta_e\right)}\right] \\
= & \left[\left(\mu\left(\theta_e\right)-\xi\right) \frac{\varepsilon_e+1}{\varepsilon_e}-1\right] H(\theta_e)+\frac{1-F_e(\theta_e)}{f_e\left(\theta_e\right)} \frac{d \ln \left[\mu\left(\theta_e\right)-\xi\right]}{d \theta_e} \\
& +\mu\left(\theta_e\right) \varepsilon_{Q_{-ij}}(\theta_e) \frac{1-F_e(\theta_e)}{f_e(\theta_e)} \frac{1+\varepsilon_e}{\varepsilon_e} \frac{\left(1+\xi \frac{\varepsilon_e+1}{\varepsilon_e}\right) \frac{d \ln \left[1-\tau_s^E\left(\theta_e\right)\right]}{d \theta_e}+\frac{d \ln \left[1-\tau_e^E\left(\theta_e\right)\right]}{d \theta_e}}{\frac{\varepsilon_e+1}{\varepsilon_e}-\frac{\sigma-1}{\sigma}\left(1+\frac{\varepsilon_e+1}{\varepsilon_e} \xi\right)}
\end{aligned}
$$

Substituting $\frac{1-F_e(\theta_e)}{f_e(\theta_e)} \frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu\left(\theta_e\right) \frac{\partial \ln P\left(Q\left(\theta_e\right), \theta_e\right)}{\partial \theta_e}+\frac{x_e^{\prime}\left(\theta_e\right)}{x_e\left(\theta_e\right)}\right]$ and $\frac{d \ln l_e(\theta_e)}{d \theta_e}$ in (60) and (61) by the above equation and (D9), respectively, and rearrange the formulas, we have

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

and

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

Setting

$$
\begin{aligned}
H_0\left(\theta_e\right) &= \left\{ \begin{array}{c} \left[1-\bar{g}_e(\theta_e)\right]\left\{\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{\frac{\sigma-1}{\sigma}\xi}{\mu(\theta_e)}\frac{d\ln[\mu(\theta_e)-\xi]}{d\theta_e}-\left[\frac{\sigma-1}{\sigma}-\left(1-\frac{\xi}{\mu(\theta_e)}\right)\frac{\varepsilon_e+1}{\varepsilon_e}\right]H(\theta_e)\right\} \\ +\varepsilon_{Q_{-ij}}(\theta_e)\left[1-g_e(\theta_e)\right]-\frac{1}{\mu(\theta_e)} \end{array} \right\} \\
H_1\left(\theta_e\right) &= \frac{1+\left[1-\bar{g}_e(\theta_e)\right]\left\{ \begin{array}{c} \left[\left[\mu\left(\theta_e\right)-\xi\right]\frac{\varepsilon_e+1}{\varepsilon_e}-1\right]H(\theta_e) \\ +\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{d\ln[\mu(\theta_e)-\xi]}{d\theta_e} \\ -\mu\left(\theta_e\right)\varepsilon_{Q_{-ij}}(\theta_e)\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{1+\varepsilon_e}{\varepsilon_e}\frac{\left(1+\xi\frac{\varepsilon_e+1}{\varepsilon_e}\right)\frac{d\ln H_0(\theta_e)}{d\theta_e}}{\frac{\varepsilon_e+1}{\varepsilon_e}-\frac{\sigma-1}{\sigma}\left(1+\frac{\varepsilon_e+1}{\varepsilon_e}\xi\right)} \end{array} \right\}}{\widetilde{\mu}} \\
H_2\left(\theta_e\right) &= \frac{\left[1-\bar{g}_e(\theta_e)\right]\mu\left(\theta_e\right)\varepsilon_{Q_{-ij}}(\theta_e)\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{\frac{1+\varepsilon_e}{\varepsilon_e}\xi}{1-\frac{\sigma-1}{\sigma}\left(\frac{\varepsilon_e}{1+\varepsilon_e}+\xi\right)}}{\widetilde{\mu}},
\end{aligned}
$$

we have

$$
\frac{1}{1-\tau_s^E\left(\theta_e\right)} = H_0\left(\theta_e\right)\left[1-\tau_e^E\left(\theta_e\right)\right]
$$

and

$$
\frac{1}{1-\tau_e^E\left(\theta_e\right)} = H_1\left(\theta_e\right) - H_2\left(\theta_e\right)\frac{d\ln\left[1-\tau_e^E\left(\theta_e\right)\right]}{d\theta_e}.
$$

Solving the above differential equation, we have

$$
1-\tau_e^E\left(\theta_e\right) = \left[1-\tau_e^E\left(\overline{\theta}_e\right)\right]e^{-\int_{\theta_e}^{\overline{\theta}_e}\frac{H_1(s)}{H_2(s)}ds} + \int_{\theta_e}^{\overline{\theta}_e} e^{-\int_{\theta_e}^{s}\frac{H_1(u)}{H_2(u)}du}\frac{1}{H_2\left(s\right)}ds,
$$

Last, according to (60), when $\overline{\theta}_e$ is finite, such that $1-F_e\left(\overline{\theta}_e\right)=1$, $\frac{\tau_e^E\left(\overline{\theta}_e\right)}{1-\tau_e^E\left(\overline{\theta}_e\right)} = \frac{1-\widetilde{\mu}}{\widetilde{\mu}}$.■

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

CHARI, V. V., M. GOLOSOV, AND A. TSYVINSKI (2012): "Prizes and patents: Using market signals to provide incentives for innovations," *Journal of Economic Theory*, 147, 781–801.

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

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

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

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

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

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

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

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.

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.

GRUBER, J. AND E. SAEZ (2002): "The elasticity of taxable income: evidence and implications," *Journal of public Economics*, 84, 1–32.

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

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

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.

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

KLEVEN, H. J. AND E. A. SCHULTZ (2014): "Estimating taxable income responses using Danish tax reforms," *American Economic Journal: Economic Policy*, 6, 271–301.

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.

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.

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.

SACHS, D., A. TSYVINSKI, AND N. WERQUIN (2020): "Nonlinear tax incidence and optimal taxation in general equilibrium," *Econometrica*, 88, 469–493.

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

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

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

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

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

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

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

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

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

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

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

SCHOTTMÜLLER, C. (2015): "Adverse selection without single crossing: Monotone solutions," *Journal of Economic Theory*, 158, 127–164.

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.