Optimal Taxation and Market Power

Jan Eeckhout
Chunyang Fu
Wenjian Li
Xi Weng

April 2021

Barcelona GSE Working Paper Series

Working Paper nº 1250

Optimal Taxation and Market Power*

Jan Eeckhout, UPF Barcelona[†]

Chunyang Fu, UCASS[‡]

Wenjian Li, ZUEL[§]

Xi Weng, Peking University[¶]

April 8, 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 labor income of workers and on the profits of entrepreneurs to induce a constrained efficient allocation. Our results show that optimal taxation in the presence of market power can substantially increase welfare, but it also highlights the severe constraints that the Planner faces to correct the negative externality from market power, using the income tax as a Pigouvian taxes. Pigouvian taxes compete with Mirrleesian incentive concerns, which generally leads to opposing forces. Overall, we find that due to incentive concerns, 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 numerous colleagues and seminar participants for many useful comments and insightful discussions. 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, and Spanish Ministry of Economy and Competitiveness, through the Severo Ochoa Programme for Centres of Excellence in R&D (CEX2019-000915-S).

[†] jan.eeckhout@upf.edu – ICREA-GSE-CREi

[‡] fuchunyang@pku.edu.cn

[§] wenjian_li@pku.edu.cn

[¶] wengxi125@gsm.pku.edu.cn

1 Introduction

Market power has an impact on both inequality and on 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. We therefore 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 the 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 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 heterogeneity 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 the equilibrium allocation and inequality. Second, we derive the optimal taxation policy to implement the planner's second best allocation. The optimal tax scheme now combines a Pigouvian correction of the externality due to market power with the design of the Mirrleesian incentive problem.

[1] Antitrust policy faces many challenges, not least because the determinants of market power are has 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 in the market such Mergers and Acquisitions. See amongst others Sutton (1991) and Sutton (2001) and De Loecker, Eeckhout, and Mongey (2019).

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 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 taxes, the optimal taxation policy in the presence of market power 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 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 the 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 skill premium within the pool of entrepreneurs, which requires a 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 decrease 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

[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 consumption 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. We therefore assume sales taxes are zero.

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.

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 skill premium 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 is as follows. In our simulated economy, the marginal tax rate is positive on average for both workers and entrepreneurs, and it increases in market power for entrepreneurs while it decreases for workers. The lump-sum tax for entrepreneurs is substantially higher than for workers, which means there is a lump-sum transfer from the entrepreneur to the workers. For entrepreneurs, the average tax rate and the total tax burden is positive and increases in market power, for workers it is negative and decreases. Across types within an occupation, higher skilled agents (both for entrepreneurs and workers) face lower marginal tax rates (as is the case for superstars, Scheuer and Werning (2017)), but the tax burden is non-monotonic. Comparing the optimal taxation allocation with Laissez-faire, optimal taxation increases welfare (trivially), but output and the wage rate are lower, and markups increase. With taxes, inequality decreases substantially, especially for the entrepreneurs.

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 et al. (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] 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 can influence output prices. Therefore, the presence of market power in the output market is notably distinct from the existing literature on optimal income taxation with market power (e.g., see Jaravel and Olivi (2019) and Kushnir and Zubrickas (2019)).

Our paper is related to the literature on optimal taxation with endogenous prices or wages (e.g., see Stiglitz (1982); Naito (1999); Saez (2004); Scheuer (2014); Sachs et al. (2020); Cui et al. (2020)). This literature emphasize the general equilibrium effect of tax 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 optimal profit tax now is dependent on the market structure. In particular, a reduction in profit tax encourages entrepreneurial effort and output. It thereby decreases the price of the competitor's product in the same submarket and hence achieves redistribution indirectly. Interestingly, when there is no competitor in the submarket, i.e., entrepreneurs have monopoly power over their own products, this effect of taxation disappears. This is because the strategic interaction between agents is absent under monopoly, which, together with entrepreneur's price setting action, totally 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 et al. (2015); Ales and Sleet (2016); Scheuer and Werning (2017); Ales et al. (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 skill premiums 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 result of this literature is the "additivity property":[6] optimal taxa-

[4]The taxation is Mirrleesian in the sense that both labor income and profit taxes are allowed to be arbitrarily nonlinear 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]The additivity property can be treated as a special case of the "principle of targeting"proposed by Dixit (1985).

tion under the presence of an externality can be expressed additively by some Pigouvian taxes and the optimal taxes in a circumstance otherwise the same without externality. 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), but also the redistributive tax change with the externality induced by market power. For one thing, social welfare weights, which are crucial to tax design, change with the extent of market power. For another, even if the social welfare weights are exogenous, the redistributive tax changes because of the skill premium, especially of the entrepreneurs, changes with the markup.

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.

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 θ_o, distributed according to the cumulative density function $F_o(\theta_o)$ with density $f_o(\theta_o)$. The measure of entrepreneurs is $N_e = N$; the measure of workers is normalized to $N_w = 1$. There is a representative firms 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.[7] 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 an agent of type θ_o.

Market Structure. The labor and final good market is 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

[7] Our utility function is separable between consumption and labor, and we eliminates income effects. The assumption is crucial for 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).

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 $J(\theta_e) = \frac{Nf(\theta_e)}{I}$. 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): $\sigma < \eta(\theta_e)$.

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 intermediates 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.[8] Because in general, the equilibrium labor inputs depends 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 $0 < \xi \leq 1$.[9] 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

[8]The assumption of efficiency units drastically simplifies the solution of the model but it is not innocuous. The efficiency units assumption rules out sorting because firms are indifferent across worker types as long as they provide exactly the same efficiency units. See amongst others Sattinger (1975a), Sattinger (1993) and Eeckhout and Kircher (2018) how the assumption of efficiency implies 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.

[9]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 which determines the boundaries of the firm.

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)[10] to output goods Q:

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

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

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

where $\chi(\theta_e)$ is a distribution parameter. It reflects the input weights for intermediate goods produced by the different entrepreneurs. It thus determines the demand and price for the intermediate goods given the output. To abstract from the love-of-variety effect related to I, we require $\zeta = -\frac{1}{\sigma}$ and $\zeta(\theta_e) = -\frac{1}{\eta(\theta_e)}$ in the following analysis.

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 efficiency 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, $y(\theta_w) = W x_w(\theta_w) \, l_w(\theta_w)$.[11]

[10]For notational simplicity and without loss, we assume the elasticity of substitution between intermediate inputs θ_e is the same as the elasticity of substitution between inputs in different markets j as there is no market power at both levels of aggregation. 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.

[11]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. (2020).

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

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 the labor incomes and entrepreneurial profits. Besides, we assume that the government can levy a linear sales tax or a linear tax on labor inputs (such as salary tax), which is usually used in the real economy.

The planner therefore 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 nonlinear 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.[12]

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 say that the social resource constraint is satisfied.

[12] 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 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.

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. 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:

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

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

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

where $Q_{ij}(\theta_e)$ is the quantity supplied of the intermediate good as defined in equation (A23). 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):

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

and

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

and

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

where R is the exogenous government revenue.

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

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

and

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

and

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

and

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

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 SC-

CTE, 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:*

1. $y(\theta_w)$ *is differentiable, strictly positive, and strictly increasing in* θ_w*;*
2. $\pi(\theta_e)$ *is differentiable, strictly positive, and positive increasing in* θ_e*;*
3. $\mu_{ij}(\theta_e)\frac{\partial \ln P\left(Q_{ij}(\theta_e),\theta_e\right)}{\partial \theta_e}+\frac{x_e'(\theta_e)}{x_e(\theta_e)}$ *is strictly positive.*

An incentive compatible allocation under the Spence-Mirrlees condition requires labor income to be non-decreasing in wage.[13] For simplicity, we assume that $y(\theta_w)$ is strictly increasing in θ_w. 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 3 of Assumption 1 is used to guarantee that to obtain certain amount of profit, the amount of effort invested by a higher-skilled entrepreneur is lower than that of a lower-skilled entrepreneur.[14] This assumption is needed to identify individuals of heterogeneous skills when prices or wages of factors are endogenous (e.g., see Sachs et al. (2020) and Cui et al. (2020)).

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

$$\mu(\theta_e) \equiv \frac{P(\theta_e)}{MC(\theta_e)} = \frac{P(\theta_e)}{\frac{W}{\frac{\partial Q_{ij}(\theta_e)}{\partial L_w(\theta_e)}(1-t_s)}}. \tag{20}$$

[13]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.

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

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 $\mu(\theta_e)$.[15] As in Atkeson and Burstein (2008), under the benchmark technology with nested CES preferences, the inverse-demand elasticity can be written in weighted form[16]

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

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. The markup therefore depends on the weighted sum of the elasticity of substitution between intermediate 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.[17]

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

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

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

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

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

[16] See Appendix A.6.3 for details.

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

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

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

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

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

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,[18] *the aggregate labor share ν is decreasing in market power (decrease in I) when*

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

and

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

Proof. See Appendix A.3. ■

Part (i) of Proposition 1 already hints at the fact that taxes cannot "solve" the problem of market power. To achieve the first best, 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 (ii) 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.[19] In addition, the parameter restrictions have intuitive eco-

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

[19]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).

nomic interpretations. Condition (28) guarantees that the equilibrium wage is increasing in TFP (e.g., see (A20)), while condition (29) ensures the labor demand is decreasing in W (e.g., see (A10)).

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;

(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

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

and individual and aggregate entrepreneur utility increase if and only if

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

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 the 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 which follows, the planner adopts feasible direct truthful mechanisms $\{c_w(\theta_w), y(\theta_w)\}$ for workers, and similarly $\{c_e(\theta_e), \pi(\theta_e)\}$ 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 type, and assigns a reward contingent 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 profits, and $c_e(\theta'_e)$ of after-tax profits.

Key to the planner's problem is that it takes the agents' private information as given, and can only induce truth-telling if it satisfies the agents' incentive compatibility constraints.

Incentive Compatibility of the Worker. Workers are atomistic and take the offered mechanisms as given. They report 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 by $V_w(\theta'_w|\theta_w) = c_w(\theta'_w) - \phi_w\left(\frac{y(\theta'_w)}{x_w(\theta_w)W}\right)$ the utility of θ_w worker who reports θ'_w. Using envelop 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 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

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

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

Lemma 1 *The first order necessary incentive condition* $\frac{\partial V_e(\theta_e'|\theta_e)}{\partial \theta_e'}|_{\theta_e'=\theta_e} = 0$ *is (i) not only a necessary but also a sufficient condition to entrepreneur's problem; and (ii) given the definition of* $V_e(\theta)$ *and the inverse demand function in the SCCTE, the first order necessary incentive condition is equivalent to*

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

where

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

and $\varepsilon_{Q_{-ij}}(\theta_e) \equiv \left[\frac{1}{\eta(\theta_e)} - \frac{1}{\sigma}\right]\frac{I-1}{I} < 0$ *is the cross inverse-demand elasticity,*[20] $\theta_e \in \Theta_e$.

Proof. See Appendix B.1 and Appendix A.6.3. ■

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. For future reference, we call $\mu\left(\theta_e\right)\frac{\partial \ln P\left(Q_{ij}(\theta_e),\theta_e\right)}{\partial\theta_e} + \frac{x_e'(\theta_e)}{x_e(\theta_e)}$ the entrepreneurial skill premium, since it determines $V_e'(\theta_e)$ given entrepreneurial effort $l_e\left(\theta_e\right)$. Analogously, $\frac{x_w'(\theta_w)}{x_w(\theta_w)} = \frac{d\ln x_w(\theta_w)}{d\theta_w}$ is worker's skill premium, which is the percentage change of individual wage rate with respect to skill θ_w, and determines $V_w'(\theta_w)$ given labor supply $l_w\left(\theta_w\right)$. Now we explain the incentive condition in two different situations:

(i) When $I = 1$, entrepreneurial skill premium is $\mu\left(\theta_e\right)\frac{\chi'(\theta_e)}{\chi(\theta_e)} + \frac{x_e'(\theta_e)}{x_e(\theta_e)}$. It is increasing in the market power $\mu\left(\theta_e\right)$ 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 $x_e\left(\theta_e\right)$ and $\chi\left(\theta_e\right)$ affect the gross utility in different ways. More specifically, given inputs, raising $\chi\left(\theta_e\right)$ increases the price directly. With the increase of price, the effort needed to finish the task becomes lower, and since the price goes up with the reduction of effort, there is a multiplier effect of raising $\chi\left(\theta_e\right)$ which is in terms of markup. On the other hand, increasing $x_e\left(\theta_e\right)$ raises output and lowers the price. Therefore, in order to obtain the same growth of profit, $x_e\left(\theta_e\right)$ should be increased at a higher rate compared to $\chi\left(\theta_e\right)$.

(ii) 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, since $Q_{ij}\left(\theta_e\right)$ 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.

[20] The elasticity of inverse demand with respect to the output of competitor in the same submarket.

This is not necessarily true though, because $\varepsilon_{Q_{-ij}}(\theta_e)\frac{Q'_{ij}(\theta_e)}{Q_{ij}(\theta_e)}$ also changes when the markup increases. In Appendix A.2.1, we show that under condition (29), the entrepreneurial skill premium of θ_e is increasing with $\mu(\theta_e)$ when $\frac{d\ln a(\theta_e)}{d\theta_e} > 0$ with $a(\theta_e) \equiv x_e(\theta_e)^{\frac{\sigma-1}{\sigma}}\chi(\theta_e)$ a composite productivity of entrepreneur. Moreover, we show that the entrepreneurial skill premium generally raises with the introduce of markup inequality.

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 through 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 firm's own output Q_{ij}. There are two interesting findings with the incentive condition:

(i) 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 of prices have 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.

(ii) When $I > 1$, the indirect redistribution effect presents again. It is dependent on the the strategic interaction between competitors in a submarket, which is shown by the last term in the right side of equation (36). Specifically, an increase in the competitors' outputs decreases the price of good 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, 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 labor supply l_w; (iii) The tax wedge $\tau_e(\theta_e)$ between the marginal disutility and income of the entrepreneur's labor supply l_e. Specifically, we shall define the three types of tax wedges as

$$\tau_s(\theta_e) = 1 - \frac{W}{\frac{P(\theta_e)}{\mu(\theta_e)}\frac{\partial Q_{ij}(\theta_e)}{\partial L_w(\theta_e)}}, \quad \tau_w(\theta_w) = 1 - \frac{\phi'_w(l_w(\theta_w))}{W x_w(\theta_w)}, \quad \text{and} \quad \tau_e(\theta_e) = 1 - \frac{\phi'_e(l_e(\theta_e))}{\frac{P(\theta_e)}{\mu(\theta_e)}\frac{\partial Q_{ij}(\theta_e)}{\partial l_e(\theta_e)}\left[1-\tau_s(\theta_e)\right]}.$$

Due to policy constraint, the government can not levy firm-specific or nonlinear sales tax, which means $\tau_s(\theta_e)$ is restricted to be uniform. Then, these tax wedges can be implemented by the tax system introduced before (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 $\tau_e(\theta_e) = 1 - [1 - T'_e(\pi(\theta_e))]$.[21] 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 loose the policy constraint and provide the optimal tax wedges including $\tau_s(\theta_e)$.

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

Lemma 2 *Suppose that 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:*

1. *Incentive conditions (33) and (35) are satisfied;*
2. *Prices and wage satisfy (16) and (17);*
3. *Markets 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 1-3 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 the 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.[22] 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.[23]

---

[21] The FOCs imply $t_s = 1 - \frac{W}{\frac{P(\theta_e)}{\mu(\theta_e)} \frac{\partial Q_{ij}(\theta_e)}{\partial L_w(\theta_e)}}$, $T'_w(y(\theta_w)) = 1 - \frac{\phi'_w(l_w(\theta_w))}{W\theta_w}$ and $T'_e(\pi(\theta_e)) = 1 - \frac{\phi'_e(l_e(\theta_e))}{\frac{P(\theta_e)}{\mu(\theta_e)} \frac{\partial Q_{ij}(\theta_e)}{\partial l_e(\theta_e)} (1-t_s)}$.

[22] 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 the income taxes are restricted to be uniform (e.g., see Scheuer (2014)).

[23] To see this, suppose that $\{T_w(y), T_e(\pi), t_s\}$ is the optimal tax that implements the second best allocation, there exists another optimal tax system $\{T^\#_w(y), T^\#_e(\pi), t^\#_s\}$ which can implement the second best allocation with $t^\#_s = 0$. Then the tax system can be constructed such that $1 - T^\#_o(x) = [1 - T'_o(x)](1 - t_s), x \in \mathbb{R}_+$.

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.

It 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, we reformulate the Planner's problem where the Planner now chooses the control variables $\{V_w(\theta_w), l_w(\theta_w), V_e(\theta_e), L_w(\theta_e), l_e(\theta_e)\}_{\theta_o\in\Theta_o}$ to maximize the social welfare (6), subject to the incentive conditions (33) and (35); the feasibility conditions (13) – (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. We define

$$\varpi(\theta_e) \equiv \frac{P(\theta_e)}{\mu(\theta_e)} \frac{\partial Q_{ij}(\theta_e)}{\partial L_w(\theta_e)} \tag{37}$$

as the marginal revenue of labor inputs since $\frac{\partial\left[P\left(Q_{ij},\theta_e\right)Q_{ij}\right]}{\partial L_w(\theta_e)} = \frac{P(\theta_e)}{\mu(\theta_e)}\frac{\partial Q_{ij}(\theta_e)}{\partial L_w(\theta_e)}$. Then, the policy constraint can be written as $\frac{d\ln\varpi(\theta_e)}{d\theta_e} = 0$.

In this reformulated planner's problem, we can now introduce some short-hand notation for the social welfare weights and the elasticities that appear in the solution to the planner's problem. We denote by $g_o(\theta_o)$ and $\bar{g}_o(\theta_o)$ the marginal and weighted social welfare weights:

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

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 by $\varepsilon^{\varpi}_{L_w}(\theta_e)$ and $\varepsilon^{\varpi}_{l_e}(\theta)$ the own elasticities of wage with respect to labor inputs and effort, respectively. These wage elasticities under our technology are given by

$$\varepsilon^{\varpi}_{L_w}(\theta_e) = \xi\left(1-\frac{1}{\sigma}\right) - 1,\ \varepsilon^{\varpi}_{l_e}(\theta) = 1-\frac{1}{\sigma}. \tag{38}$$

See Appendix A.6.4 for details about these wage elasticities. We define a non-linear elasticity of profit with respect to net-tax income rate by (39):

$$\varepsilon^{\pi}_{1-\tau_e}(\pi(\theta_e)) \equiv \frac{1}{\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu(\theta_e)-\xi\right] - \left[1-\frac{\pi(\theta_e)T''_e(\pi(\theta_e))}{1-T'_e(\pi(\theta_e))}\right]}. \tag{39}$$

This elasticity captures θ_e firm's reaction to net-tax income rate when they taking others' actions as given (see Appendix A.6.2 for detail). Notably, the elasticity is generally decreasing in raising of markup, which is clear when the profit tax is linear (so that $T''_e = 0$).

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; (iv) the general case of oligopolistic competition with heterogeneous markups. Each of these special cases gradually show the different components of the optimal tax wedges.

Theorem 1 *For any $\theta_e \in \Theta_e$, when $\tau_s = 0$, the optimal tax wedges satisfy:*

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

$$\frac{1}{1-\tau_e(\theta_e)} = \frac{[1-\bar{g}_e(\theta_e)] \frac{1-F_e(\theta_e)}{f_e(\theta_e)} \frac{1+\varepsilon_e}{\varepsilon_e} \left[\mu(\theta_e) \frac{\partial \ln P(Q(\theta_e),\theta_e)}{\partial \theta_e} + \frac{x_e'(\theta_e)}{x_e(\theta_e)}\right] + 1 + \mu(\theta_e)\, IRE(\theta_e) \left[1 + \frac{\xi}{\frac{\sigma}{\sigma-1}-\xi}\right]}{\mu(\theta_e)\left[1 - RE(\theta_e) \frac{\xi}{\frac{\sigma}{\sigma-1}-\xi}\right]}, \tag{41}$$

where $\frac{1}{\frac{\sigma}{\sigma-1}-\xi} = -\frac{\varepsilon_{l_e}^{\varpi}(\theta_e)}{\varepsilon_{L_w}^{\varpi}(\theta_e)} < 0$. *The Reallocation Effect $RE(\theta_e)$ and Indirect Redistribution Effect $IRE(\theta_e)$ are defined as*

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

$$IRE(\theta_e) \equiv \varepsilon_{Q_{-ij}}(\theta_e) \left\{ [1-g_e(\theta_e)] - \frac{[1-\bar{g}_e(\theta_e)][1-F_e(\theta_e)]}{f_e(\theta_e)} \left[\frac{1+\varepsilon_e}{\varepsilon_e} \frac{l_e'(\theta_e)}{l_e(\theta_e)} + \frac{d \ln \left[\mu(\theta_e)\varepsilon_{Q_{-ij}}(\theta_e)\right]}{d\theta_e} \right] \right\}. \tag{43}$$

Proof. See Appendix C.2 ∎

The term $RE(\theta_e)$ captures the Reallocation Effect of taxes. When the markup of θ_e firm is higher than the modified average markup μ, $RE(\theta_e)$ of tax decreases $\tau_e(\theta_e)$, and vice versa. This is because the labor demand of high-markup firm is more inefficiently low as compared to low-markup firm. Thus interventions in product market should reallocate labor factors to the high-mark firms.

As an explaination to $IRE(\theta_e)$, suppose that tax rates and markups are constant within an interval

$(\underline{\theta}_e^*, \overline{\theta}_e^*) \in \Theta_e^2$. Then we have $\frac{d \ln\left[\mu(\theta_e)\varepsilon_{Q_{-ij}}(\theta_e)\right]}{d\theta_e} = 0$ and $\frac{1+\varepsilon_e}{\varepsilon_e}\frac{l_e'(\theta_e)}{l_e(\theta_e)} = \frac{\pi_e'(\theta_e)}{\pi_e(\theta_e)}$ (e.g., see Lemma 3). Thus[24]

$$IRE\left(\theta_e\right) \equiv \varepsilon_{Q_{-ij}}(\theta_e)\left[\left[g_e(\theta_e)-1\right]+\left[1-\bar{g}_e(\theta_e)\right]\frac{1-F_\pi(\pi\left(\theta_e\right))}{f_\pi\left(\pi\left(\theta_e\right)\right)\pi\left(\theta_e\right)}\right], \forall(\underline{\theta}_e^*, \overline{\theta}_e^*) \in \Theta_e^2. \quad (44)$$

The term $IRE\left(\theta_e\right)$ captures the Indirect Redistribution Effect of tax. Decreasing $\tau_e\left(\theta_e\right)$ increases the output of firms (i.e., $Q_{-ij}(\theta_e)$) which decreases the price of products in θ_e submarket ($P\left(\theta_e\right)$). The decrease of price brings two redistribution effects: A local redistribution effect captured by $\varepsilon_{Q_{-ij}}(\theta_e)\left[1-g_e(\theta_e)\right]$ and a cumulative redistribution effect captured by $\varepsilon_{Q_{-ij}}(\theta_e)\left[1-\bar{g}_e(\theta_e)\right]\frac{1-F_\pi(\pi(\theta_e))}{f_\pi(\pi(\theta_e))\pi(\theta_e)}$. Intuitively, decrease of price $P\left(\theta_e\right)$ reduces the after-tax income of θ_e entrepreneur, which promote equality and social welfare if and only if $g_e(\theta_e) < 1$. This decrease of 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 $\tau_e\left(\theta_e\right)$.

Introducing heterogeneous agents brings two additional elements to tax design: (i) a taxation motive for redistribution induced by within-occupation income differences; (ii) a taxation motive to increase efficiency, induced by the inter-firm allocation of factors.

(i) Homogeneous Agents

First, we analyze the optimal taxation formulae when workers and entrepreneurs are homogeneous.

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

$$\tau_w = \tau_e = 1 - \mu. \quad (45)$$

Proof. When worker and entrepreneur types are homogeneous $\bar{g}_o = g_o = 1$, and the optimal tax formulas (40) and (41) can be simplified to (45). ■

A few aspect of this result deserve mention. First, note that since all entrepreneurs are identical, the firm-level markup is $\mu(\theta_e)$ 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 the cases of monopolistic competition and oligopolistic competition. Third, the optimal tax rate is *negative* because the markup is larger than one.

The interpretation of the optimal tax formula is straightforward. When agents are identical, the incentive constraints are mute because they are trivially satisfied. As a result, the optimal tax wedge exactly offsets the distortion due to the markup. It may seem surprising, but this affects both workers and entrepreneurs equally because the planner can only affect output by affecting their incentive to

[24] (C45) suggests that $\frac{L_w'(\theta_e)}{L_w(\theta_e)} = \frac{1+\varepsilon_e}{\varepsilon_e}\frac{l_e'(\theta_e)}{l_e(\theta_e)}, \forall(\underline{\theta}_e^*, \overline{\theta}_e^*) \in \Theta_e^2$. Thus, hazard ratio of profit (i.e., $\frac{1-F_\pi(\pi(\theta_e))}{f_\pi(\pi(\theta_e))\pi(\theta_e)}$) in the above formula can be replaced by a hazard ration of labor inputs.

produce and provide effort. The planner would be able to achieve this second-best outcome also with the 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 of workers and entrepreneurs to output is the same, the income tax wedges τ_w and τ_e are identical.

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 income tax system is to correct the markup distortion and transfer between groups. As a result, we can think of the tax here playing the role of a Pigouvian tax that corrects an externality or distortion in the output market. 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. The tax burden of each occupation 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 markup power and the level of the markup μ plays a key role for 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. Higher markups will lead to a higher tax burden on entrepreneurs, whereas for the workers the effect is ambiguous. Higher markups leads to higher marginal taxes, but there is also more redistribution due to the increase in inequality between workers and entrepreneurs.

(ii) Monopolistic Competition and Uniform Markups

We now turn to an economy without strategic interaction within each market j, i.e., with a monopolistic producer in each market where $I = 1$. In addition, we assume that markups are uniform, i.e., the residual demand elasticity is constant $\eta(\theta_e) = \eta$. Entrepreneurs are heterogeneous in productivity θ_e, but their market 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 tax wedges satisfy:*

$$\frac{1}{1-\tau_w(\theta_w)} = \frac{1 + [1 - \bar{g}_w(\theta_w)] \frac{1+\varepsilon_w}{\varepsilon_w} \frac{1-F_w(\theta_w)}{f_w(\theta_w)} \frac{x_w'(\theta_w)}{x_w(\theta_w)}}{\mu}, \tag{46}$$

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

Proof. When $I = 1$, one has $\varepsilon_{Q_{-ij}}(\theta_e) = 0$ and $\mu(\theta_e) = \mu$, (40) and (41) are reduced to be (46) and (47), respectively. ■

Comparing the optimal tax rates under heterogenous and homogenous agents establishes that the heterogeneity between agents calls for a higher tax wedge. The markup μ is divided by an expression that immediately stems from the incentive constraint, both for the worker and the entrepreneur. The denominator is larger than one and thus the marginal tax is lower than in the case of homogeneous agents and identical markups. Because when the government has a preference for equality, it will raise tax rate to generate tax revenue and transfer to the individuals. However, the government should also take tax's distortion on effort into consideration. In the end the trade off between dead-weight loss and redistribution benefits is captured by the denominator.

While there is a lot of similarity in the expression of the tax wedge for the workers and the entrepreneurs, there is one marked difference between the two. That difference in the numerator of the right side of (47) goes to the heart of one of the main findings of this paper. For both workers and entrepreneurs, the extent to which the marginal tax rate is lower depends on the Pareto weight $\bar{g}_o(\theta_o)$, the elasticity of labor supply ε_o and the hazard ratio of the distribution of types $F_o(\theta_o)$. In addition, the denominator depends on the technology of production. First, it depends on the productivity $x_o(\theta_o)$ of the individual worker and entrepreneur through $\frac{x_o'(\theta_o)}{x_o(\theta_o)}$. As the productivity of an agent increases, the tax wedge decreases in order to provide appropriate incentives.

But the tax wedge between the worker and the entrepreneur differs through the term $\mu\frac{\chi'(\theta_e)}{\chi(\theta_e)}$ which is equal to $\mu\frac{\partial \ln P(Q(\theta_e),\theta_e)}{\partial\theta_e}$. This term is only present for entrepreneurs and not for workers, and captures the two key forces. First, it depends on the technology and how it different sectors differ in productivity $\chi(\theta_e)$, and second, how big the markup is.[25]

Relative Net-tax Income Rate between Groups. Two interesting findings are: (i) The relative net-tax income rates of profit to labor income will decrease with the markup (i.e., $\frac{1-\tau_e(\theta_e)}{1-\tau_w(\theta_w)}$ is decreasing in μ), which can be easily seen from

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

(ii) Optimal tax reform is dependent on the technology. Specifically, under uniform x_e and heterogenous χ, the policy implication of the above equation will be decreasing the net-tax income rate of profit as relative to the net-tax income rate of labor income when the markup raises. While under uniform χ and heterogenous x_e, the policy implication will be no need to changing the relative rate of net-tax income rates between profit and labor income.

In addition to the above results, we have the following resutls:[26] When $\frac{\chi'(\theta_e)}{\chi(\theta_e)}=0$, $\tau_e\left(\theta_e\right)>\tau_w\left(\theta_w\right)$ is a sufficient and necessary condition for $\frac{d[\tau_e(\theta_e)-\tau_w(\theta_w)]}{d\mu}>0$; when $\frac{\chi'(\theta_e)}{\chi(\theta_e)}>0$, $\tau_e\left(\theta_e\right)>\tau_w\left(\theta_w\right)$ is a sufficient

[25]Recall, in this section we consider a uniform markup, independent of the sector.

[26]After a simple calculation, one has $\frac{d[\tau_e(\theta_e)-\tau_w(\theta_w)]}{d\mu}=\frac{1}{\frac{\mu}{1-\tau_w(\theta_w)}}-\frac{1+[1-\bar{g}_e(\theta_e)]\frac{1+\varepsilon_e}{\varepsilon_e}\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{x_e'(\theta_e)}{x_e(\theta_e)}}{\left[\frac{\mu}{1-\tau_e(\theta_e)}\right]^2}$. It equals $\frac{1}{\frac{\mu}{1-\tau_w(\theta_w)}}-\frac{1}{\frac{\mu}{1-\tau_e(\theta_e)}}$ when

but not necessary condition for $\frac{d[\tau_e(\theta_e)-\tau_w(\theta_w)]}{d\mu} > 0$. Again, one can see the optimal tax wedges on skill are dependent on the technology.

Relative Net-tax Income Rate within Groups. Now we try to figure out how markup shapes optimal profit tax rate. Whether marginal tax rates for entrepreneurs increase or decrease with market power μ therefore depends on the technology χ. To see this, consider the case where productivity for workers and entrepreneurs $x_o(\theta)$ is invariant of type, i.e., $x_o' = 0$. Then we have

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

which is increasing in μ if and only if

$$\left[1-\bar{g}_e(\theta_e)\right]\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{\chi'\left(\theta_e\right)}{\chi\left(\theta_e\right)} > \left[1-\bar{g}_e(\theta_e')\right]\frac{1-F_e(\theta_e')}{f_e(\theta_e')}\frac{\chi'\left(\theta_e'\right)}{\chi\left(\theta_e'\right)}.$$

Notice that

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

and as a result, $\frac{1-\tau_e(\theta_e')}{1-\tau_e(\theta_e)}$ is increasing in μ if and only if $1-\tau_e\left(\theta_e'\right) > 1-\tau_e\left(\theta_e\right)$. This suggests that an increase in μ generally widens the relative difference between the net-tax income rates of different entrepreneur types.

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 formula are both determinants of the productivity of a firm. In fact, we can show that the $\chi\left(\theta_e\right)$ and $x_e\left(\theta_e\right)$ are substitute parameters in the sense that the equilibrium labor supply and sales incomes are only dependent on the composite productivity term $a(\theta_e) = x_e(\theta_e)^{\frac{\sigma-1}{\sigma}}\chi(\theta_e)$. However, $\chi\left(\theta_e\right)$ and $x_e\left(\theta_e\right)$ are not perfect substitutes in the sense that the equilibrium prices are dependent on the specific values of $\chi\left(\theta_e\right)$ and $x_e\left(\theta_e\right)$.[27] Since the entrepreneurial skill premium is directly related to the markups and prices, and the prices are dependent on the specific technology characterized by $\chi\left(\theta_e\right)$ and $x_e\left(\theta_e\right)$, the influence of markup on optimal taxation is dependent on the technology.

Technology, Monopoly Power, and Optimal Tax. It is of our interest to see how technology affect the optimal tax rate and we take the top tax rate as an illustration. Top tax rate is in the key of optimal income tax literature since the top earners account for the vast majority of income. Moreover, current change of technology biases top income individuals, thus it's crucial to see how technologies affect the top tax rates. Specificlly, we will focus on the influence of μ and ζ. Note that when $I=1$, markup is

$\frac{\chi'(\theta_e)}{\chi(\theta_e)} = 0$, and is larger than $\frac{1}{\frac{\mu}{1-\tau_w(\theta_w)}} - \frac{1}{\frac{\mu}{1-\tau_e(\theta_e)}}$ when $\frac{\chi'(\theta_e)}{\chi(\theta_e)} > 0$. Thus we have the following results.

[27] See Appendix A.5 for details about the influence of χ and x_e on equilibrium.

determined by the technology. Also note that ξ is a mirror of superstar effect considered by Scheuer and Werning (2017).

Denote by π_0 and μ_0 the profit and markup in the initial economy, and by $F_{\pi_0}(\pi_0)$ and $f_{\pi_0}(\pi_0)$ the corresponding CDF and PDF. Assume that there exist $\theta_e^* \in (\underline{\theta}_e, \overline{\theta}_e)$, so that initial profit tax rates on $\pi_0 \geq \pi_0(\theta_e^*)$ is constant, and $\frac{1-F_{\pi_0}(\pi_0(\theta_e))}{\pi_0(\theta_e) f_{\pi_0}(\pi_0(\theta_e))} = \overline{H}_{initial}$ is constant on $(\theta_e^*, \overline{\theta}_e)$.[28] We can establish the following Corollary 1:

Corollary 1 *Suppose that there exist $\theta_e^\star \in (\underline{\theta}_e, \overline{\theta}_e)$, so that for any $\theta_e \geq \theta_e^\star$, $\tau_e(\theta_e) = \overline{\tau}_e$ and $\bar{g}_e(\theta_e) = \overline{g}_e$ are constants. When $I = 1$, there exist π, so that for profits higher than π, the optimal profit tax rate is constant and equal to $\overline{\tau}_e$:*

$$\frac{1}{1-\overline{\tau}_e} = \frac{1 + (1-\overline{g}_e)\,\overline{H}_{initial}\left(\frac{1+\varepsilon_e}{\varepsilon_e}[\mu - \xi] - 1\right)}{\mu}. \tag{48}$$

Proof. See Appendix C.3. ■

Corollary 1 is powerful because it suggests that under reasonable assumptions, one can use original observable statistics including $\overline{H}_{initial}$ to derive the optimal top profit tax rate. It is worth noting that $\overline{H}_{initial}$ is the hazard ratio under initial tax policy. It won't change with the tax policies under the preconditions of Corollary 1, but it is dependent on the markup.

Formula (48) generalizes the familiar top tax rate result of Saez (2001) (in which $\xi = 0$ and $\mu = 1$) to a CES production function under monopoly competitive economy. When we compare our result to the Corollary 5 of Sachs et al. (2020), we see how market structure and technology affects the top tax rates given the hazard ratio of profit. However, while thise statistics-based optimal tax formulas facilitate tax design (it is a robust tax forumula in the sense of independent on technology), one should note that profit distribution is endogenous to the markup and when analysis how profit tax change with the raising markup, both the effects of markups on elasticity of profit and on profit distribution should be taken into consideration.

Combining Corollary 1 and Proposition 2 delivers interesting insights in the light of the findings by Scheuer and Werning (2017). When we look into the defined elasticity of profit, one can see that an increase in ξ (superstar effect) increases the elasticity of profit (39) while an increase in μ (markup) decrease the elasticity of profit. When χ is constant, Proposition 2 suggests that $\frac{1-\tau_e(\theta_e)}{1-\tau_w(\theta_w)}$ won't change with the markup for any θ_e. Thus, $\frac{1-\overline{\tau}_e}{1-\tau_w(\theta_w)}$ must not change with the markup, which suggests an increase in the markup decreases the elasticity as well as the hazard ratio of profit distribution. Moreover, these two effects will cancel each other out so that the relative net-tax income rate is unchanged. However, when χ varies with type but x_e is constant, the influence of hazard ratio will be relatively stronger, so that $\frac{1-\overline{\tau}_e}{1-\tau_w(\theta_w)}$ will be decreased with the increase of markup.

[28] Suppose that the profit distributed to a Pareto distribution when profit is high enough, and let $\Pi > 1$ denote the Pareto coefficient of the tail of the income distribution. That is, $1 - F_{\pi_o}(\pi_o) \sim c\pi_o^{-\Pi}$ as $\pi_o \to \infty$ for some constant c. Then $H_{\pi_o}^{top} = \frac{1}{\Pi}$.

The above findings show that how the optimal taxation changes with the markup and suggest that it is important to take both x and χ into consideration. What should be noted is that here the markup changing is due to the change of elasticity of substitution, and it's implication is generally different from the one frim markup change due to change of market structure. See case (iii) and (iv) for details.

(iii) Oligopolistic Competition with Uniform Markups

We now consider cases with $I > 1$, but still restrict the markup to be uniform, i.e., $\eta(\theta_e)$ is constant. This setting now introduces inter-firm strategic action but still abstracts from the effect of markup inequality between firms.

A Planner who intends to take the advantage of general equilibrium price effect and ease the incentive constraint would like to decrease the relative price of the goods produced by high-skilled entrepreneurs. However, whether the Planner should encourage the factor inputs of high-skilled entrepreneur ambiguous because there are two opposing forces. On one hand, raising labor inputs of competitors in the same submarket reduces the relative price of goods in the submarket; on the other hand, raising labor inputs of an agent increases its marginal productivity of effort.

To ascertain the optimal policy, we rewrite the general tax formula for the entrepreneurs in equation (41) for the case with uniform markups.

$$\frac{1}{1-\tau_e(\theta_e)} = \frac{1 + [1 - \bar{g}_e(\theta_e)] \frac{1-F_e(\theta_e)}{f_e(\theta_e)} \frac{1+\varepsilon_e}{\varepsilon_e} \left[\mu(\theta_e) \frac{\partial \ln P(Q(\theta_e),\theta_e)}{\partial \theta_e} + \frac{x_e'(\theta_e)}{x_e(\theta_e)}\right]}{\mu(\theta_e)} + IRE(\theta_e) \frac{\frac{\sigma}{\sigma-1}}{\frac{\sigma}{\sigma-1} - \xi} \tag{49}$$

Compared to the tax formula under monopolistic competition (47), 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.

$\frac{\frac{\sigma}{\sigma-1}}{\frac{\sigma}{\sigma-1}-\xi} = 1 - \xi \frac{\varepsilon^{\omega}_{l_e}(\theta_e)}{\varepsilon^{\omega}_{L_w}(\theta_e)}$ $\left(\frac{\varepsilon^{\omega}_{l_e}(\theta_e)}{\varepsilon^{\omega}_{L_w}(\theta_e)} = -\frac{1}{\frac{\sigma}{\sigma-1}-\xi}\right)$ 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 $\frac{\xi}{\frac{\sigma}{\sigma-1}-\xi}$ percent as a result of a one percent increase in $l_e(\theta_e)$. This 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 $IRE(\theta_e)$. Since $IRE(\theta_e)$ is generally positive for low θ_e and negative for high θ_e, it raises tax rate on low skills and decreases tax rates on high skills.[29] Intuitively, decreasing profit tax rate on high-skilled entrepreneur can

[29] Below 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 dominate the cumulative effect (e.g., see the numerical analysis in Cui et al. (2020)). Then one can see that the sign of $IRE(\theta_e)$ is mainly determined by $\varepsilon_{Q_{-ij}}(\theta_e)[1 - g_e(\theta_e)]$, which is positive when $g_e(\theta_e) > 1$.

enhance the output of intermediate goods in the submarket which in turn reduces the price of intermediate goods and improve the predistribution.

Market Structure, Indirect Redistribution, and Optimal Tax. One interesting finding of this paper is that the marker structure is crucial to the IRE (or supply-side effect) of taxation, which is omitted by previous studies. In previous studies with endogenous prices or wages, the direct tax has a first-order effect, and adjusting the relative prices thus can be used to ease the incentive constraint (e.g., see Naito (1999); Sachs et al. (2020)). Specifically, when the marginal productivity of 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 high-skilled agent's labor supply.

However, the agent's pricing action counteracts this IRE. In particular, when $I = 1$, $\frac{\partial \ln P\left(\overline{Q}(\theta_e),\theta_e\right)}{\partial \theta_e}$ is exogenous, and as a result, the incentive constraint can not be eased through the tax's effect on prices and the IRE of a profit tax disappears. This is because in a monopolistic market, the entrepreneur can choose the price at the intensive margin, which cancels out tax's first order effect on the price. This finding supports the argument of Saez (2004) – that the indirect distribution use of tax may fail – by providing a novel reason.

Our result suggests that IRE does matter when there are competitors in the sub-market, i.e., $I \geq 2$. This result thus supports Naito (1999) and Naito (2004) who argue that IRE is important if the agent can endogenously change the wage (price). Our setup provides a novel reason why IRE matters, namely imperfect competition in the output market. Our finding thus contributes to the debate on whether IRE is important in tax design by suggesting that its importance is dependent on the market structure.

(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 efficiency inducing reallocation of factors and enhance production efficiency.

For the workers, the tax formula (40) is unchanged compared to the case with uniform markups. The introduction of the heterogenous markups introduces the last change in the in the tax formula for the entrepreneurs (41), which is captured by the denominator of the right side of (41):

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

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 $[\mu(\theta_e) - \mu] W$ is the increase in 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 labor inputs of different firms are still uniform. On the other hand, $\frac{\xi}{\frac{\sigma}{\sigma-1} - \xi} = -\xi \frac{\varepsilon_{l_e}^{\omega}(\theta_e)}{\varepsilon_{L_w}^{\omega}(\theta_e)}$ is the percentage increase of labor demand $L_w(\theta_e)$ that ensures the marginal productivity of labor inputs to be uniform between firms when $l_e(\theta_e)$ is increased by one percent.

In conclusion, $[\mu(\theta_e) - \mu] \frac{\xi}{\frac{\sigma}{\sigma-1} - \xi}$ captures the aggregate output (in terms of labor inputs) increased with one percent increase of $l_e(\theta_e)$ and the resulting inter-firm re-allocation of workers' labor inputs. Our optimal tax formula suggests that the reallocation effect requires a lower (higher) tax rate on firms with markup higher (lower) than the average markup, since the labor inputs in firms with higher markup is relatively inefficient low. In the following analysis, we provide two special case to futher explain the optimal profit tax: (i) optimal tax under Utilitarian social welfare function; (ii) top tax rate.

Utilitarianism To make the effect of reallocation effect more transperant, we now analyze the special case where the social welfare function is Utilitarian. That is, we assume that $G(V_o(\theta_o)) = V_o(\theta_o)$ and $\widetilde{f}_o(\theta_o) = 1$, and as a result, $\lambda = 1$, and $g_o(\theta_o) = \bar{g}_o(\theta_o) = 1$. Under the above assumptions, we have

Corollary 2 *Denote by $\tau_w^{\mathbf{U}}$ and $\tau_e^{\mathbf{U}}$ the optimal labor income and profit tax rates under the Utilitarian social welfare function. Then the optimal tax wedges satisfy:*

$$1 - \tau_w^{\mathbf{U}}(\theta_w) = \mu, \tag{50}$$

$$1 - \tau_e^{\mathbf{U}}(\theta_e) = \mu(\theta_e) + [\mu(\theta_e) - \mu] \frac{\xi}{\frac{\sigma}{\sigma-1} - \xi}, \tag{51}$$

Proof. The above results can be derived by substituting the social welfare weights for the utilitarian social welfare function into the general optimal tax formulas (40) and (41). ■

First, optimal labor income tax rate is constant and negative, as in the case of a homogenous agents. Intuitively, the labor income tax is now only used to enhance production efficiency and correct the markup distortion, thus the optimal tax rate is only dependent on the average markup.

Second, (51) suggests $\tau_e^{\mathbf{U}}(\theta_e)$ increases with the average markup μ and decreases with the firm-level markup $\mu(\theta_e)$. Furthermore, when $\mu(\theta_e)$ is large enough (for example, $\mu(\theta_e) \geq \mu$), $\tau_e^{\mathbf{U}}(\theta_e)$ must be negative. Interestingly, $\tau_e'(\theta_e) = -\mu'(\theta_e) \frac{\frac{\sigma}{\sigma-1}}{\frac{\sigma}{\sigma-1} - \xi}$ is negative when markup increases with θ_e. 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)). It shows that the optimal profit tax will generally become more regressive markup inequality increases.

In addition, we find that $\tau_e^{\mathbf{U}}(\theta_e) > \tau_w^{\mathbf{U}}(\theta_w)$ if and only if $\mu > \mu(\theta_e)$:

$$\tau_e^{\mathbf{U}}(\theta_e) - \tau_w^{\mathbf{U}} = [\mu - \mu(\theta_e)] \frac{\frac{\sigma}{\sigma-1}}{\frac{\sigma}{\sigma-1} - \xi}.$$

It suggests that there exist θ_e^*, such that for any $\theta_e < \theta_e^*$, $\tau_e^{\mathbf{U}}(\theta_e) > \tau_w^{\mathbf{U}}$; and for any $\theta_e \geq \theta_e^*$, $\tau_e^{\mathbf{U}}(\theta_e) \leq \tau_w^{\mathbf{U}}$. This finding adds an explanation to why the effective tax rate on larger firms may be lower than the effective tax rate on workers.

There are two things worth noting. First, a decreasing marginal profit tax rate on high-skilled entrepreneur not necessary means that the average tax rate and tax burden on high-income entrepreneur is decreasing. To see this, define $\tau_\pi^{\mathbf{U}}(\pi(\theta_e)) = \tau_e^{\mathbf{U}}(\theta_e)$. One has

$$T_e(\pi(\theta_e)) = T_e(\pi(\underline{\theta}_e)) + \int_{\pi(\underline{\theta}_e)}^{\pi(\theta_e)} \tau_\pi^{\mathbf{U}}(\pi)\, d\pi,$$

where $T_e(\pi(\underline{\theta}_e))$ can be treated as a lump-sum tax (transfer if it is negative), since it is same for any entrepreneur. The tax burden of θ_e does not only depend on the marginal tax rate $\tau_\pi^{\mathbf{U}}(\pi(\theta_e))$, but also on the lump-sum tax and the marginal tax rates on profits below $\pi(\theta_e)$. Second, when social welfare weights are endogenous, whether the profit tax should decrease with the markup depends on the Planner's preference for equality. Note that the markup will shape the distribution of gross utilities, thus the social welfare weights.[30]

The Utilitarian social welfare function suggests that in the absence for a preference for equality, increased markup inequality will generally make the profit tax more regressive. This highlights the dilemma of tax design in a market exhibiting market power: production enhancement versus redistribution. Since using taxes to enhance factor inputs means lower marginal tax rate, the total lump-sum transfers generally becomes lower with the higher markups. Moreover, since large firms generally have higher market power, the dilemma also makes the redistribution within the group more expensive. This dilemma faced by tax design is more pronounced when there is market power than under perfect competition.

Lastly, with the definition of $\tau_e^{\mathbf{U}}$, we have

$$\frac{1}{1-\tau_e(\theta_e)} = \frac{\left[\begin{array}{c} [1-\bar{g}_e(\theta_e)] \frac{1-F_e(\theta_e)}{f_e(\theta_e)} \frac{1+\varepsilon_e}{\varepsilon_e} \left[\mu(\theta_e) \frac{\partial \ln P(Q(\theta_e),\theta_e)}{\partial \theta_e} + \frac{x_e'(\theta_e)}{x_e(\theta_e)}\right] \\ +1 + \mu(\theta_e)\, IRE(\theta_e) \frac{\frac{\sigma}{\sigma-1}}{\frac{\sigma}{\sigma-1} - \xi} \end{array}\right]}{1-\tau_e^{\mathbf{U}}(\theta_e)}$$

where $1-\tau_e^{\mathbf{U}}$ captures the Pigouvian part of optimal tax wedges.

[30] Note that when income (profit) or skill is finite, $\tau_o^{\mathbf{U}}(\underline{\theta}_o)$ and $\tau_o^{\mathbf{U}}(\bar{\theta}_o)$ is also the optimal tax rates on the boundary under the most general case (case 3). We can see that both $\tau_w^{\mathbf{U}}(\underline{\theta}_w)$ and $\tau_w^{\mathbf{U}}(\bar{\theta}_w)$ are negative. $\tau_e^{\mathbf{U}}(\bar{\theta}_e)$ is negative, while $\tau_e^{\mathbf{U}}(\underline{\theta}_e)$ may either be positive or negative. How raising markup affects the optimal tax on the boundary is dependent on both the firm-level markup and average markup, which is ambiguous.

Top Tax Rate. The general optimal profit tax formula is quite complex. As we have done above, under additional assumptions, we can deliver a simpler top tax formula:

Corollary 3 *Suppose that there exist $\theta_e^\star \in \left(\underline{\theta}_e, \overline{\theta}_e\right)$, so that for any $\theta_e \geq \theta_e^\star$, $\tau_e\left(\theta_e\right) = \overline{\tau}_e$, $\bar{g}_e(\theta_e) = \overline{g}_e$ and $\mu\left(\theta_e\right) = \overline{\mu}$ are constants. Then, for $\theta_e \geq \max\left\{\theta_e^*, \theta_e^\star\right\}$, we have*[31]

$$\frac{1}{1-\overline{\tau}_e} = \frac{1 + \left[1-\bar{g}_e\right]\overline{H}_{initial}\left[\frac{1+\varepsilon_e}{\varepsilon_e}\left(\overline{\mu}-\xi\right)-1\right]}{\overline{\mu}} + \left[1-\bar{g}_e\right]\left[1-\overline{H}_{initial}\right]\frac{\frac{\sigma}{\sigma-1}}{\frac{\sigma}{\sigma-1}-\xi}\varepsilon_{Q_{-ij}}(\theta_e) - \frac{\xi\left(1-\frac{\mu}{\overline{\mu}}\right)}{\frac{\sigma}{\sigma-1}-\xi}\frac{1}{1-\overline{\tau}_e}. \tag{52}$$

Proof. See Appendix C.5. ■

When $I = 1$, $\frac{\mu}{\overline{\mu}} = \frac{\sigma}{\sigma-1}\frac{1}{\overline{\mu}} = 1$, and the formula in Corollary 3 can be reduced to the formula in Corollary 1. How top profit tax rate changes with I is dependent on how average markup μ, top markup $\overline{\mu}$ and hazard ratio of profit $\overline{H}_{initial}$ change with I. Note that the second and third terms in the right side of (52) capture the RE and IRE, both of which are negative when $\overline{H}_{initial} < 1$.

5 Extension

In our benchmark model we consider an environment with uniform linear sales tax, which restricts $\tau_s\left(\theta_e\right)$ to be constant. In this section we remove the policy constraint. This extension helps us to understand the optimal profit tax under the policy constraint.

To do this, we allow the planner contracts with entrepreneurs on sales income $P\left(\theta_e\right)Q_{ij}\left(\theta_e\right)$ in addition to $\pi(\theta_e)$ and $c_e(\theta_e)$. An entrepreneur who reports θ_e' obtains $\pi(\theta_e')$ in profits, $S(\theta_e')$ in sales income, and $c_e(\theta_e')$ of after-tax profits. Worker's problem is as before. Then all incentive compatible allocations satisfying (33) and (35) are feasible as long as the resource constraints are also satisfied. The planner's problem is similar to the one that introduced before, except that the policy constraint $\frac{d\ln\varpi(\theta_e)}{d\theta_e} = 0$ is now relaxed. We use superscript B to denote values without policy constraint.

[31] Alternatively, we can say that there exist π, so that for profits higher than π, the optimal profit tax rate is constant and equal to $\overline{\tau}_e$: $1-\overline{\tau}_e = \frac{1+\left(1-\frac{\mu}{\overline{\mu}}\right)\frac{\xi}{\frac{\sigma}{\sigma-1}-\xi}}{\frac{1+(1-\bar{g}_e)\overline{H}_{initial}\left[\frac{1+\varepsilon_e}{\varepsilon_e}(\overline{\mu}-\xi)-1\right]}{\overline{\mu}} + \frac{1-\frac{\sigma}{\sigma-1}\frac{1}{\overline{\mu}}}{\frac{\sigma}{\sigma-1}-\xi}\left[1-\bar{g}_e\right]\left[1-\overline{H}_{initial}\right]}$, where we substitute the elasticities by specific parameters. Provided $\overline{H}_{initial}$, the first term in the denominator (i.e.,$\frac{1+(1-\bar{g})\overline{H}_{initial}\left[\frac{1+\varepsilon_e}{\varepsilon_e}(\overline{\mu}-\xi)-1\right]}{\overline{\mu}}$) decreases with $\overline{\mu}$. Provided $\overline{H}_{initial}$, the second term in the denominator (i.e., $\frac{1-\frac{\sigma}{\sigma-1}\frac{1}{\overline{\mu}}}{\frac{\sigma}{\sigma-1}-\xi}\left[1-\bar{g}\right]\left[1-\overline{H}_{initial}\right]$) generally increases with $\overline{\mu}$, because the empirical $\overline{H}_{initial}$ is generally lower than one. The numerator is larger than one if $\overline{\mu} > \mu$. It decreases in μ and increase in $\overline{\mu}$.

Theorem 2 *The optimal tax wedges without policy constraint (τ_s is free) satisfy:*

$$\frac{\tau_w^B(\theta_w)}{1-\tau_w^B(\theta_w)} = \frac{1-\widetilde{\mu}+\left[1-\bar{g}_w(\theta_w)\right]\frac{1-F_w(\theta_w)}{f_w(\theta_w)}\frac{x_w'(\theta_w)}{x_w(\theta_w)}\frac{1+\varepsilon_w}{\varepsilon_w}}{\widetilde{\mu}}, \tag{53}$$

$$\frac{\tau_e^B(\theta_e)}{1-\tau_e^B(\theta_e)} = \frac{1-\widetilde{\mu}+\left[1-\bar{g}_e(\theta_e)\right]\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu(\theta_e)\frac{\partial\ln P(Q(\theta_e),\theta_e)}{\partial\theta_e}+\frac{x_e'(\theta_e)}{x_e(\theta_e)}\right]}{\widetilde{\mu}}, \tag{54}$$

$$\frac{\tau_s^B(\theta_e)}{1-\tau_s^B(\theta_e)} = \overbrace{\left[\frac{\widetilde{\mu}}{\mu(\theta_e)}-1\right]}^{RE(\theta_e)}+\overbrace{\left[1-\tau_e(\theta_e)\right]\varepsilon_{Q_{-ij}}(\theta_e)\left[\begin{array}{c}\left[1-g_e(\theta_e)\right]-\frac{\left[1-\bar{g}_e(\theta_e)\right]\left[1-F_e(\theta_e)\right]}{f_e(\theta_e)}\\ \times\left[\frac{1+\varepsilon_e}{\varepsilon_e}\frac{l_e'(\theta_e)}{l_e(\theta_e)}+\frac{d\ln\left[\mu(\theta_e)\varepsilon_{Q_{-ij}}(\theta_e)\right]}{d\theta_e}\right]\end{array}\right]}^{IRE(\theta_e)}. \tag{55}$$

$RE(\theta_e)$ and $IRE(\theta_e)$ are generalized reallocation and indirect redistribution effects defined in Theorem 1, where

$$\widetilde{\mu}\equiv\frac{\int_{\theta_e}\frac{\mu(\theta_e)}{1-\tau_s^B(\theta_e)}L_w(\theta_e)f_e(\theta_e)\,d\theta_e}{\int_{\theta_e}L_w(\theta_e)f_e(\theta_e)\,d\theta_e} \tag{56}$$

is a modified average markup. If for any $\theta_e\in\Theta_e$, $\tau_s^B(\theta_e)=0$, $\widetilde{\mu}$ equals to μ.

Proof. See Appendix D.1 ■

Comparing Theorem 2 and 1, one can see that without $\tau_s^B(\theta_e)$, labor income and profit taxes are modified to mimic the role of $\tau_s^B(\theta_e)$. Specifically, $RE(\theta_e)\,\xi\frac{\varepsilon_{l_e}^{\varpi}(\theta_e)}{\varepsilon_{L_w}^{\varpi}(\theta_e)}$ and $IRE(\theta_e)\left[1-\xi\frac{\varepsilon_{l_e}^{\varpi}(\theta_e)}{\varepsilon_{L_w}^{\varpi}(\theta_e)}\right]$ in (41) capture the reallocation effect and indirect reallocation effect of profit tax, respectively. We now turn to the special cases to gradually build up our understanding of the optimal tax wedges in the most general case.

In the above analysis, we actually assume that the government has sufficient tax policies that can be used to intervene product market so as to impliment the $\tau_s^B(\theta_e)$. However it is hard to enforce the optimal product market intervention in the real economy. Thus, in the benchmark analysis, we focus on optimal taxation without such product market interventions (only linear sales tax is allowed), and see how can other tax policies be modifed to mimic the role of $\tau_s^B(\theta_e)$. Another advantage of the benchmark analysis is that there $\tau_o(\theta_o)$ reflect the effective (aggregate) tax rate on labor factors.

6 Numerical Analysis

Our general results depends the social preferences for redistribution. To see the overall impact of market power on optimal taxation, we numerically analyze and economy with a concave social welfare functions with $G(V)=\frac{V^{1-k}}{1-k}$. 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.

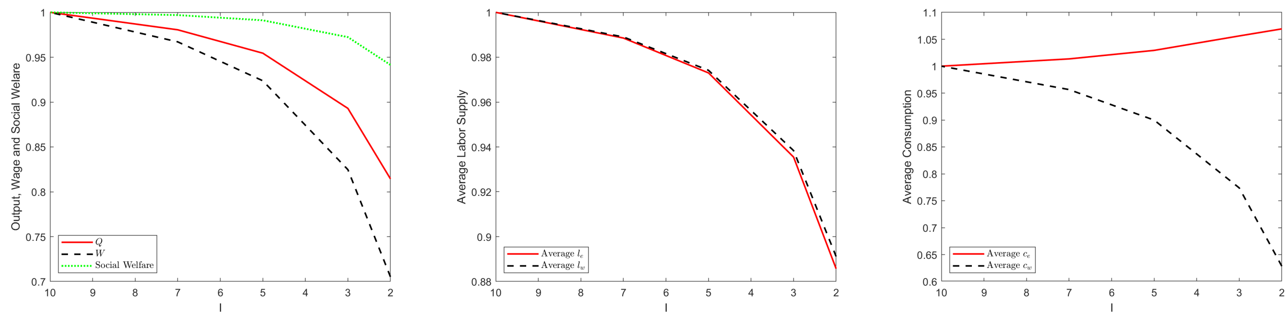
We maintain the following assumptions for the numerical analysis as follow. We treat θ_e and θ_w as the quantiles of $\pi(\theta_e)$ and $y(\theta_w)$, which means $f_o = 1$ is uniform on $\Theta_o = [0,1]$. Since the functions $x_o(\theta_o)$ and $\chi(\theta_e)$ are used to govern the heterogeneity,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

Table 1: Parameterization

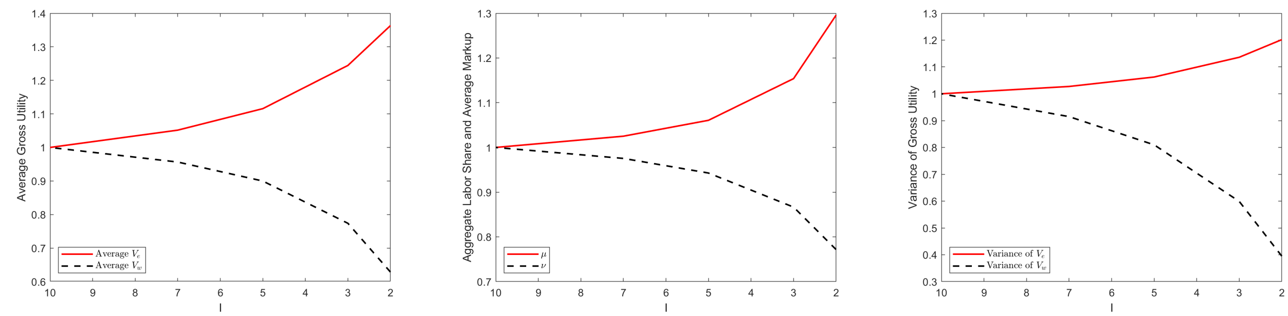
$G(V) = \frac{V^{1-k}}{1-k}$	social welfare function
$k \in \{1,3\}$	concavity of the social welfare function; $k = 1$ is benchmark
$f_o(\theta_o) = 1$	PDF of skills
$N_e = 0.2$	measure of entrepreneurs
$A = 10^4$	the TFP of final good production technology Q
$\xi = 0.85$	concavity of technology Q_{ij}
$\sigma = 1.5$	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))

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 decreasing order, to indicate increasing market power. Th 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, 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 a 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 entrepreneurs type distribution. Finally, markups are increasing and the labor share is decreasing in entrepreneur type.

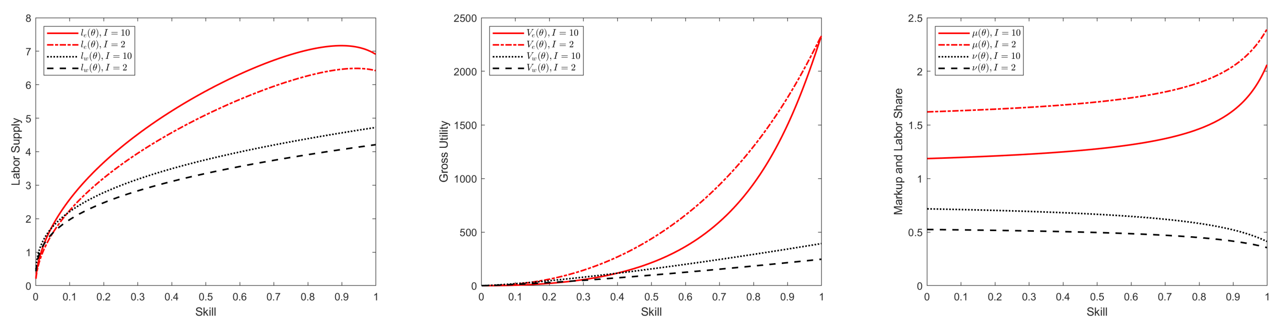


(a) Output, wage rate, social welfare | (b) Average Labor Supply | (c) Average Consumption



(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 $I = 10$



(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 represent 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

Table 2: Summary of Tax Measures

t_o	Lump-sum tax (depends on occupation, not on incomes)
$\tau_o\left(\theta_o\right)$	Marginal tax rate
$T_o\left(y(\theta_o)\right)$	Tax burden
$ATR_o\left(\theta_o\right) = \frac{T_o(y(\theta_o))}{y(\theta_o)}$	Average tax rate
$AVTR_o\left(\theta_o\right) = \frac{T_o(y(\theta_o))-t_o}{y(\theta_o)}$	Average variable tax rate
$TT_o = N_o \int_{\theta_o} T_o\left(y\left(\theta_o\right)\right) f_{\theta_o}\left(\theta_o\right) d\theta_o$	Total tax burden
$MMTR_o = \int_{\theta_o} \tau_o\left(\theta_o\right) f_{\theta_o}\left(\theta_o\right) d\theta_o$	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

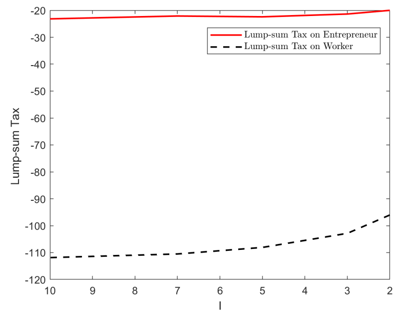
Note: we denote profits by $\pi(\theta_e) = y(\theta_e)$.

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

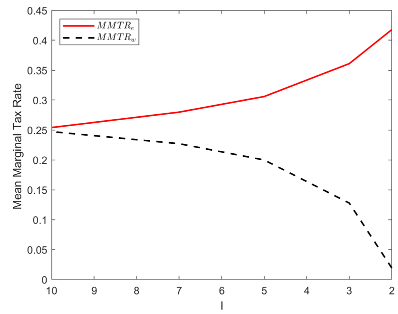
Next, in Figure 3, we report how optimal taxes vary by skill. The marginal tax rate is decreasing in skill, in order to provide incentives to exert effort (Figure 4a). This is the standard Mirrleesian incentive property. To provide incentives to the top entrepreneurs, a sharp decline until negative in 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 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 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.[32] Not surprisingly, social welfare is higher under optimal taxation than in Laissez-faire (Figure 5a). However, 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 supply of high-skill entrepreneur actually increases. Also, the firm-level labor supply suggests that although before tax labor supply of entrepreneur may decrease with the skill, it is

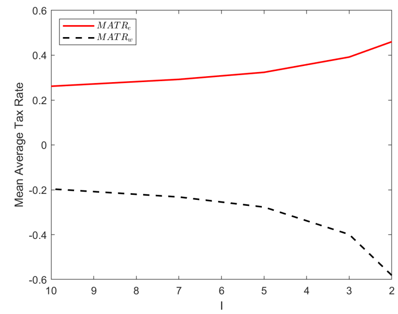
[32] Figure E2 in the Appendix reports the same results in ratios.



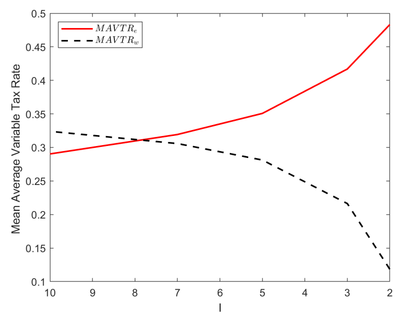
(a) Lump sum tax



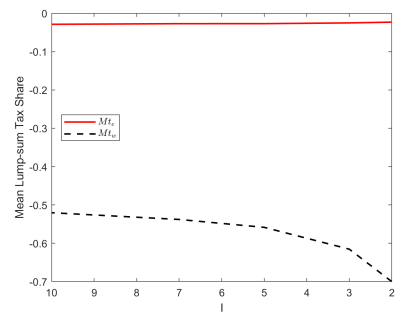
(b) Mean Marginal Tax Rate



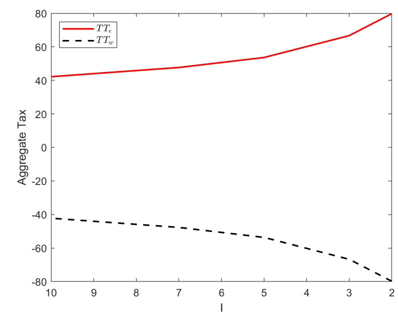
(c) Mean Average Tax Rate



(d) Mean Average Variable Tax Rate

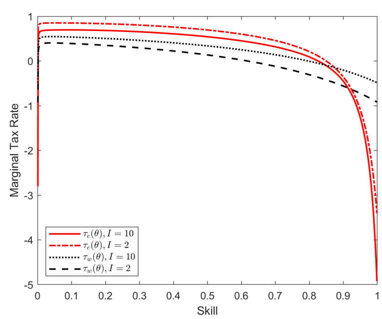


(e) Mean Lump-sum Tax Share

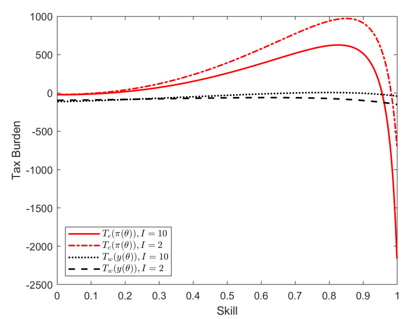


(f) Total Tax Burden

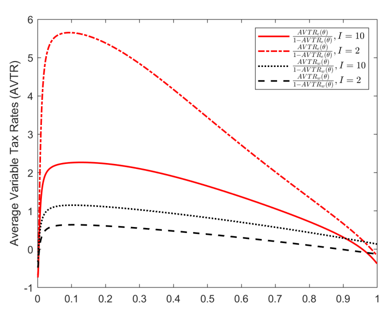
Figure 3: Optimal Taxation: Effect of market Power (number of competitors I)



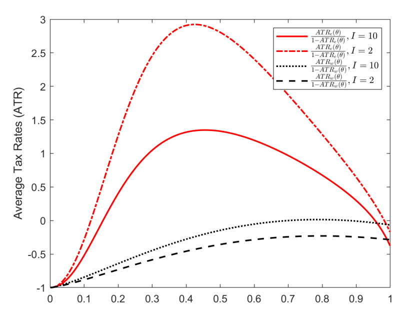
(a) Marginal Tax Rate



(b) Tax Burden



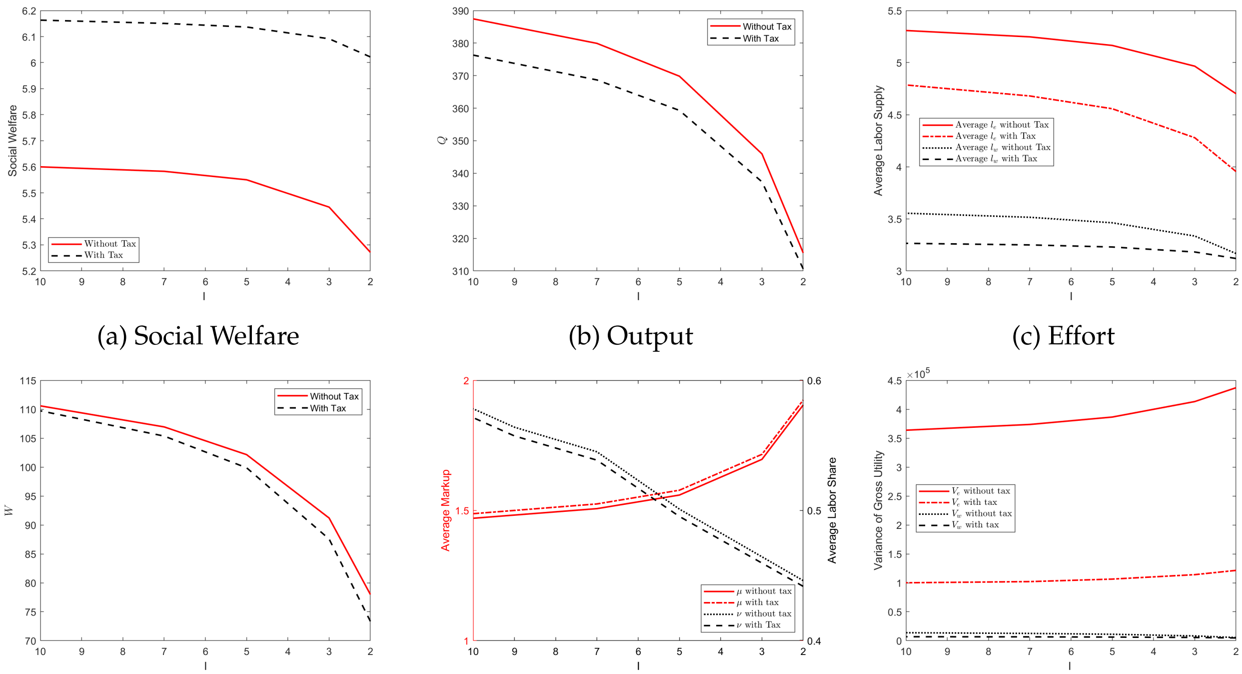
(c) Average Variable Tax Rate



(d) Average Tax Rate

Figure 4: Optimal Taxation: Variation by skill θ

increasing with the skill after the tax.



(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

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 the 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.

7 Conclusion

The best way to address market power is to cut out the root cause with 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 individuals' and final good producer's problmes, the equilibrium allocations are determined by (13) to (19) and individuals' budget constraints. Under the technology considered in this paper, and $\phi_o(l_o) = \frac{l_o^{1+\frac{1}{\varepsilon_o}}}{1+\frac{1}{\varepsilon_o}}$ we have the following conditions in the symmetric equilibrium:

1.First order conditions:

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

and

$$WL_w(\theta_e) = \frac{\xi(1-t_s)}{\mu(\theta_e)} P(\theta_e) Q_{ij}(\theta_e), \tag{A2}$$

and

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

and

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

2. Inverse demand function

$$P(Q_{ij},\theta_e) = \chi(\theta_e) A^{\frac{\sigma-1}{\sigma}} Q_{ij}^{-\frac{1}{\eta(\theta_e)}} I^{-\left[\frac{1}{\eta(\theta_e)}-\frac{1}{\sigma}\right]\frac{\eta(\theta_e)}{\eta(\theta_e)-1}} \begin{bmatrix} (I-1)\, Q_{ij}(\theta_e)^{\frac{\eta(\theta_e)-1}{\eta(\theta_e)}} \\ +Q_{ij}^{\frac{\eta(\theta_e)-1}{\eta(\theta_e)}} \end{bmatrix}^{\left[\frac{1}{\eta(\theta_e)}-\frac{1}{\sigma}\right]\frac{\eta(\theta_e)}{\eta(\theta_e)-1}} \left[\frac{Q}{N_e f_e(\theta_e)}\right]^{\frac{1}{\sigma}}, \tag{A5}$$

3. Labor market clear condition

$$N_w \int_{\theta_w} x_w(\theta_w)\, l_w(\theta_w)\, f_w(\theta_w) d\theta_w = N_w [W]^{\varepsilon_w} \int_{\theta_w} [\varkappa(\theta_w)]^{\varepsilon_w+1} [1 - t_w(\theta_w)]^{\varepsilon_w} f_w(\theta_w) d\theta_w \tag{A6}$$

4. Besides, in the equilibrium, we have

$$Q = \int_{\theta_e} N_e f_e(\theta_e) \left[P(\theta_e)Q_{ij}(\theta_e)\right] d\theta_e. \tag{A7}$$

The above part 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 allocation with individuals' budget constraints.

A.2 Laissez-faire Economy

Combination of (A2), (A4), and (A1) gives

$$l_e(\theta_e) = \left[\frac{A^{\frac{\sigma-1}{\sigma}} x_e(\theta_e)^{\frac{\sigma-1}{\sigma}} \chi(\theta_e)}{\mu(\theta_e)}\right]^{\frac{\sigma\varepsilon_e}{\varepsilon_e+\sigma}} L_w(\theta_e)^{\frac{\xi(\sigma-1)\varepsilon_e}{\varepsilon_e+\sigma}} \left[\frac{Q}{N_e f_e(\theta_e)}\right]^{\frac{\varepsilon_e}{\varepsilon_e+\sigma}} \tag{A8}$$

Substituting $P(\theta_e)$ and $Q_{ij}(\theta_e)$ in (A2) by (A1) and $Q_{ij}(\theta_e) = [x_e(\theta_e) l_e(\theta_e)] L_w(\theta_e)^{\xi}$, respectively, we have

$$\begin{aligned} L_w(\theta_e) &= \frac{\xi}{W\mu(\theta_e)}\chi(\theta_e)\left[x_e(\theta_e) l_e(\theta_e) L_w(\theta_e)^{\xi}\right]^{1-\frac{1}{\sigma}} \left[\frac{Q}{N_e f_e(\theta_e)}\right]^{\frac{1}{\sigma}} \\ &= \frac{\xi}{W}\left[\frac{A^{\frac{\sigma-1}{\sigma}} x_e(\theta_e)^{\frac{\sigma-1}{\sigma}} \chi(\theta_e)}{\mu(\theta_e)}\right]^{\frac{\sigma(\varepsilon_e+1)}{\varepsilon_e+\sigma}} L_w(\theta_e)^{\frac{\xi(\sigma-1)(\varepsilon_e+1)}{\sigma+\varepsilon_e}} \left[\frac{Q}{N_e f_e(\theta_e)}\right]^{\frac{\varepsilon_e+1}{\sigma+\varepsilon_e}}, \end{aligned} \tag{A9}$$

where we substitute $l_e(\theta_e)$ by (A8) in the second equation.

Rearranging the above equation gives

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

Substituting the above equation into (A8), we have

$$P(\theta_e)Q_{ij}(\theta_e) = \mu(\theta_e)\left(\frac{\xi}{W}\right)^{\frac{\xi(\sigma-1)(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}} \left[\frac{A^{\frac{\sigma-1}{\sigma}} \chi_2(\theta_e)^{\frac{\sigma-1}{\sigma}} \chi_1(\theta_e)}{\mu(\theta_e)}\right]^{\frac{\sigma(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}} \left[\frac{Q}{N_e f_e(\theta_e)}\right]^{\frac{\varepsilon_e+1}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}} \tag{A11}$$

and

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

Equation (A3) gives

$$l_w(\theta_w) = [W x_w(\theta_w)]^{\varepsilon_w}. \tag{A13}$$

The above three equations together with (A6) and (A7) 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 allocation with individuals' budget constraints. See below for details.

For later use, we define

$$\begin{aligned} A_1 &= \int_{\theta_e} N_e f_e(\theta_e)\,\mu(\theta_e) \left[\frac{A^{\frac{\sigma-1}{\sigma}} x_e(\theta_e)^{\frac{\sigma-1}{\sigma}} \chi(\theta_e)}{\mu(\theta_e)\,[N_e f_e(\theta_e)]^{\frac{1}{\sigma}}} \right]^{\frac{\sigma(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}} d\theta_e, \quad \text{(A14)} \\ A_2 &= \int N_e f_e(\theta_e) \left[\frac{A^{\frac{\sigma-1}{\sigma}} x_e(\theta_e)^{\frac{\sigma-1}{\sigma}} \chi(\theta_e)}{\mu(\theta_e)\,[N_e f_e(\theta_e)]^{\frac{1}{\sigma}}} \right]^{\frac{\sigma(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}} d\theta_e, \\ A_3 &= N_w \xi^{\varepsilon_w} \int_{\theta_w} x(\theta_w)^{\varepsilon_w+1} f_w(\theta_w) d\theta_w. \end{aligned}$$

Substituting $L_w(\theta_e)$ in (A2) by (A10), we have

$$P(\theta_e)Q_{ij}(\theta_e) = \mu(\theta_e) \left(\frac{\xi}{W}\right)^{\frac{\xi(\sigma-1)(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}} \left[\frac{A^{\frac{\sigma-1}{\sigma}} \chi_2(\theta_e)^{\frac{\sigma-1}{\sigma}} \chi_1(\theta_e)}{\mu(\theta_e)} \right]^{\frac{\sigma(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}} \left[\frac{Q}{N_e f_e(\theta_e)} \right]^{\frac{\varepsilon_e+1}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}} \quad \text{(A15)}$$

Substituting $P(\theta_e)Q_{ij}(\theta_e)$ in (A7) by (A15), we have

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

which gives the following equation by the definition of A_1:

$$Q = \left(\frac{\xi}{W}\right)^{\frac{\xi(\varepsilon_e+1)}{1-\xi(\varepsilon_e+1)}} A_1^{\frac{\sigma+\varepsilon_e-\xi(\sigma-1)(\varepsilon_e+1)}{1-\xi(\varepsilon_e+1)} \frac{1}{\sigma-1}}. \quad \text{(A16)}$$

Similarly, substituting $L_w(\theta_e)$ in (A6) by (A10), we have the aggregate labor demand

$$L^D \equiv \left(\frac{\xi}{W}\right)^{\frac{1}{1-\xi(\varepsilon_e+1)}} (A_1)^{\frac{\varepsilon_e+1}{(\sigma-1)[1-\xi(\varepsilon_e+1)]}} A_2. \quad \text{(A17)}$$

On the other hand, according to (A6) and (A13), we have the aggregate labor supply

$$L^S \equiv N_w [W]^{\varepsilon_w} \int_{\theta_w} x(\theta_w)^{\varepsilon_w+1} f_w(\theta_w) d\theta_w = \left[\frac{W}{\xi}\right]^{\varepsilon_w} A_3. \quad \text{(A18)}$$

Cmbination of (A17) and (A18) gives

$$\left[\frac{W}{\xi}\right]^{\varepsilon_w+\frac{1}{1-\xi(\varepsilon_e+1)}} = (A_1)^{\frac{\varepsilon_e+1}{(\sigma-1)[1-\xi(\varepsilon_e+1)]}} \frac{A_2}{A_3}, \quad \text{(A19)}$$

i.e.,

$$W = \xi \left[(A_1)^{\frac{\varepsilon_e+1}{(\sigma-1)[1-\xi(\varepsilon_e+1)]}} \frac{A_2}{A_3} \right]^{\frac{1}{\varepsilon_w+\frac{1}{1-\xi(\varepsilon_e+1)}}}. \quad \text{(A20)}$$

Lastly, substituting W in (A16) by (A20), we have

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

Then, we can derive $l_w(\theta_w)$, $L_w(\theta_e)$ and $l_e(\theta_e)$ by substituting Q and W into (A10), (A12), and (A13). Besides, by definitions, we have

$$c_e(\theta_e)=[\mu(\theta_e)-\xi]\left(\frac{\xi}{W}\right)^{\frac{\xi(\sigma-1)(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}}\left[\frac{A^{\frac{\sigma-1}{\sigma}}x_e(\theta_e)^{\frac{\sigma-1}{\sigma}}\chi(\theta_e)}{\mu(\theta_e)}\right]^{\frac{\sigma(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}}\left[\frac{Q}{N_e f_e(\theta_e)}\right]^{\frac{\varepsilon_e+1}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}}, \tag{A22}$$

and

$$Q_{ij}(\theta_e)=x_e(\theta_e)\left(\frac{\xi}{W}\right)^{\frac{\xi\sigma(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}}\left[\frac{A^{\frac{\sigma-1}{\sigma}}x_e(\theta_e)^{\frac{\sigma-1}{\sigma}}\chi(\theta_e)}{\mu(\theta_e)}\right]^{\frac{\sigma(\varepsilon_e+1)\xi+\sigma\varepsilon_e}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}}\left[\frac{Q}{N_e f_e(\theta_e)}\right]^{\frac{\xi(\varepsilon_e+1)+\varepsilon_e}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}}, \tag{A23}$$

and

$$P(\theta_e)=\frac{\mu(\theta_e)}{x_e(\theta_e)}\left(\frac{\xi}{W}\right)^{\frac{-\xi(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}}\left[\frac{A^{\frac{\sigma-1}{\sigma}}x_e(\theta_e)^{\frac{\sigma-1}{\sigma}}\chi(\theta_e)}{\mu(\theta_e)}\right]^{\frac{\sigma-\sigma(\varepsilon_e+1)\xi}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}}\left[\frac{Q}{N_e f_e(\theta_e)}\right]^{\frac{1-\xi(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}}, \theta_e\in\Theta. \tag{A24}$$

In addition, we have

$$V(\theta_e)=c_e(\theta_e)-l_e(\theta_e)^{\frac{\varepsilon_e+1}{\varepsilon_e}}=\left[\mu(\theta_e)-\xi-\frac{\varepsilon_e}{\varepsilon_e+1}\right]l_e(\theta_e)^{\frac{\varepsilon_e+1}{\varepsilon_e}}, \theta_e\in\Theta. \tag{A25}$$

A.2.1 Gross Utility

Notice that

$$V_e(\theta_e)\equiv\max_{l_e(\theta_e),L_w(\theta_e)} P(Q_{ij}(\theta_e),\theta_e)Q_{ij}(\theta_e)-WL_w(\theta_e)-\frac{[l_{e,ij}(\theta_e)]^{1+\frac{1}{\varepsilon_e}}}{1+\frac{1}{\varepsilon_e}},$$

we have the following equation by envelop theory:

$$\begin{aligned} V_e'(\theta_e) &= \left[\frac{P(Q_{ij}(\theta_e),\theta_e)Q_{ij}(\theta_e)}{\mu(\theta_e)}\frac{x_e'(\theta_e)}{x_e(\theta_e)}+\frac{\partial P(Q_{ij}(\theta_e),\theta_e)}{\partial\theta_e}Q_{ij}(\theta_e)\right] \\ &= \frac{P(Q_{ij}(\theta_e),\theta_e)Q_{ij}(\theta_e)}{\mu(\theta_e)}\left[\mu(\theta_e)\frac{\partial\ln P(Q_{ij}(\theta_e),\theta_e)}{\partial\theta_e}+\frac{x_e'(\theta_e)}{x_e(\theta_e)}\right], \end{aligned}$$

where according to the first order condition (A4)

$$\frac{P(Q_{ij}(\theta_e),\theta_e)Q_{ij}(\theta_e)}{\mu(\theta_e)}=l_e(\theta_e)\,\phi'(l_e(\theta_e)).$$

It can be seen that the gross utility under the equilibrium is consistent with the one we derived by mechanism design method, i.e.,

$$V_e'(\theta_e) = l_e(\theta_e)^{\frac{\varepsilon_e+1}{\varepsilon_e}} \left[\mu(\theta_e) \frac{\partial \ln P(Q_{ij}(\theta_e),\theta_e)}{\partial \theta_e} + \frac{x_e'(\theta_e)}{x_e(\theta_e)}\right], \forall \theta_e \in \Theta_e. \tag{A26}$$

Also note that by (A5), we have

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

where, by definition,

$$\begin{aligned} & \frac{d\ln Q_{ij}(\theta_e)}{d\theta_e} = \frac{d\ln x_e(\theta_e)}{d\theta_e} + \frac{d\ln l_e(\theta_e)}{d\theta_e} + \xi\frac{d\ln L_w(\theta_e)}{d\theta_e} \\ = & \frac{d\ln x_e(\theta_e)}{d\theta_e} + \frac{\sigma(\varepsilon_e+1)\xi + \sigma\varepsilon_e}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)} \left[\begin{array}{c} \frac{d\ln \chi(\theta_e)}{d\theta_e} - \frac{d\ln \mu(\theta_e)}{d\theta_e} \\ + \frac{\sigma-1}{\sigma}\frac{d\ln x_e(\theta_e)}{d\theta_e} \end{array}\right] \\ & - \frac{\varepsilon_e+\xi\varepsilon_e+\xi}{\sigma+\varepsilon_e-\xi(\sigma-1)(\varepsilon_e+1)}\frac{d\ln f_e(\theta_e)}{d\theta_e}. \end{aligned}$$

Thus,

$$\begin{aligned} & \frac{x_e'(\theta_e)}{x_e(\theta_e)} + \mu(\theta_e)\frac{\partial \ln P(Q_{ij},\theta_e)}{\partial \theta_e} \\ = & \frac{\sigma(\varepsilon_e+1)[\mu(\theta_e)-\xi] - \sigma\varepsilon_e}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}\frac{d\ln a(\theta_e)}{d\theta_e} \\ & + \frac{(\sigma-1)[\varepsilon_e+\xi(\varepsilon_e+1)]\left[\frac{\sigma}{\sigma-1}-\mu(\theta_e)\right]}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}\frac{d\ln\mu(\theta_e)}{d\theta_e} \\ & - \frac{[\mu(\theta_e)-\xi][1+\varepsilon_e]-\varepsilon_e}{\sigma+\varepsilon_e-\xi(\sigma-1)(\varepsilon_e+1)}\frac{d\ln f_e(\theta_e)}{d\theta_e} \end{aligned} \tag{A27}$$

where $a(\theta_e) = x_e(\theta_e)^{\frac{\sigma-1}{\sigma}}\chi(\theta_e)$.

Notice that $\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)$ is positive under condition (29), we can see that the entrepreneurial skill premium is increasing with $\mu(\theta_e)$ when $\frac{d\ln a(\theta_e)}{d\theta_e} > 0$.

Moreover, we can see that markup inequality raises the entrepreneurial skill premiums. Note that $\mu(\theta_e) \le \frac{\sigma}{\sigma-1}$, and $\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)$ is positive under condition (29). Thus, the coefficient of $\frac{d\ln\mu(\theta_e)}{d\theta_e}$ in the right side of (A27) is positive under condition (29), which suggests that with the introduce of markup inequality ($\frac{d\ln\mu(\theta_e)}{d\theta_e} > 0$), the entrepreneurial skill premium is increased.

A.3 Proof of Proposition 1

By (A21), (A20), and (A18), we have

$$\nu(I) \triangleq \frac{WL}{\xi Q} = \frac{A_2}{A_1}, \tag{A28}$$

where L is the aggregate labor inputs. Substituting A_1 and A_2 by (A14), we have

$$\nu(I) = \frac{\int f_e(\theta_e)\left[\frac{x_e(\theta_e)^{\frac{\sigma-1}{\sigma}}\chi(\theta_e)}{\mu(\theta_e)[f_e(\theta_e)]^{\frac{1}{\sigma}}}\right]^{\frac{\sigma(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}} d\theta_e}{\int_{\theta_e} f_e\left(\theta_e\right)\mu\left(\theta_e\right)\left[\frac{x_e(\theta_e)^{\frac{\sigma-1}{\sigma}}\chi(\theta_e)}{\mu(\theta_e)[f_e(\theta_e)]^{\frac{1}{\sigma}}}\right]^{\frac{\sigma(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}} d\theta_e}. \tag{A29}$$

For the convenience of analysis, define

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

Then, we have

$$\nu\left(I\right) = \int_{\theta_e} \frac{f_s(\theta_e, I)}{\mu\left(\theta_e\right)} d\theta_e. \tag{A30}$$

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 $\mu\left(\theta_e\right)$, 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 (A30), 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

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

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

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

Thus

$$\begin{aligned}&\int_{\theta_e\in\Omega_*} f_s(\theta_e, I)\left[\frac{1}{\mu\left(\theta_e\right)}-\nu\left(I\right)\right]\frac{d\ln\left[\frac{1}{\mu(\theta_e)}\right]}{d\ln I}d\theta_e \\ \geq\ &\left[1-\frac{\sigma-1}{\sigma}\nu\left(I\right)\right]\frac{I}{I-1}\int_{\theta_e\in\Omega_*} f_s(\theta_e, I)\left[\frac{1}{\mu\left(\theta_e\right)}-\nu\left(I\right)\right]d\theta_e.\end{aligned} \tag{A31}$$

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

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

Thus

$$\begin{aligned} & \int_{\theta_e \notin \Omega_*} f_s(\theta_e, I)\left[\frac{1}{\mu\left(\theta_e\right)}-\nu\left(I\right)\right]\frac{d\ln\left[\frac{1}{\mu(\theta_e)}\right]}{d\ln I}d\theta_e \\ \ge \quad & \left[1-\frac{\sigma-1}{\sigma}\nu\left(I\right)\right]\frac{I}{I-1}\int_{\theta_e \notin \Omega_*} f_s(\theta_e, I)\left[\frac{1}{\mu\left(\theta_e\right)}-\nu\left(I\right)\right]d\theta_e, \end{aligned} \tag{A32}$$

Combination of (A31) and (A32) gives

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

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

A.4 Proof of Proposition 2

Assume that markups are constant. (A20) and (A21) give

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

Substituting W and Q in (A10) by the above equations, we have

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

Similarly, we have

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

It's easy to see that under the conditions (28) and (29), $L_w(\theta_e)$, $S(\theta_e)$, $l_e(\theta_e)$, $\overline{Q}(\theta_e)$, $P(\theta_e)$ go down with the decrease of I. Besides, 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

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

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)), effort-augmenting and total-productivity-augmenting (non-effort-augmenting) elements (Ales et al. (2017)), which catches the difference between entrepreneur and worker.

The expressions for allocations and prices in Appendix A.1 show that $Q_{ij}(\theta_e)$ and $P(\theta_e)$ is generally dependent on the specific values of $x_e(\theta_e)$ and $\chi(\theta_e)$ instead of only dependent on the value of $a(\theta_e) = x_e(\theta_e)^{\frac{\sigma-1}{\sigma}}\chi(\theta_e)$. Specifically, according to (A23) and (A24), we have

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

and

$$\begin{aligned} P(\theta_e) &= A\chi(\theta_e)^{\frac{\sigma}{\sigma-1}}\left(\frac{W}{\xi}\right)^{\frac{\xi(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}}\left[\frac{Q}{N_e f_e(\theta_e)}\right]^{\frac{1-\xi(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}} \\ &\times\left[\frac{\mu(\theta_e)}{A^{\frac{\sigma-1}{\sigma}}a(\theta_e)}\right]^{\frac{(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)}\frac{\sigma}{\sigma-1}}\left[\frac{1}{\mu(\theta_e)}\right]^{\frac{1}{\sigma-1}}. \end{aligned}$$

On the other hand, given $a(\theta_e)$, $P(\theta_e)Q_{ij}(\theta_e)$, $L_w(\theta_e)$, $l_e(\theta_e)$, and $V_e(\theta_e)$ are independent on the specific values of $\chi(\theta_e)$ and $x_e(\theta_e)$. According to (A10) to (A12) and (A25), we have the following results:

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

and

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

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

Lastly, we find that given $\frac{d\ln a(\theta_e)}{d\theta_e}$, $\frac{V'(\theta_e)}{V(\theta_e)}$ is independent on the specific values of $\chi(\theta_e)$ and $x_e(\theta_e)$. Combination of (A26) and (A27) gives

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

Combination of (A25) and (A26) gives

$$\frac{V'(\theta_e)}{V(\theta_e)} = \frac{1}{\mu(\theta_e) - \xi - \frac{\varepsilon_e}{\varepsilon_e+1}} \left[\frac{d\ln\chi(\theta_e)}{d\theta_e} + \mu(\theta_e) \frac{\partial \ln P(Q_{ij}\left(\frac{\partial \ln P(Q_{ij},\theta_e)}{\partial\theta_e}\right), \theta_e)}{\partial\theta_e} \right] \tag{A33}$$

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

where the second equation is derived by (A27) and (A26). Specially, when markup and distribution function are constant, we have

$$\frac{V'(\theta_e)}{V(\theta_e)} = \frac{\sigma(\varepsilon_e+1)}{\varepsilon_e+\sigma-\xi(\sigma-1)(\varepsilon_e+1)} \frac{d\ln a(\theta_e)}{d\theta_e}, \forall\theta_e \in \Theta_e. \tag{A34}$$

Note that the coefficient of $\frac{d\ln a(\theta_e)}{d\theta_e}$ in the right side of (A33) is positive under condition (29).

■

A.6 Elasticities

A.6.1 Worker

Keeping the wage constant, labor income elasticity with respect to the net-of-tax rate along the linearized budget constraint (linear labor elasticity) is[33]

$$\varepsilon_w(\theta_w) \equiv \frac{\phi'_w(l_w(\theta_w))}{l_w(\theta_w)\phi''_w(l_w(\theta_w))}. \tag{A35}$$

To compute the elasticities along the non-linear budget constraint (named as non-linear elasticity), we introduce

$$\psi(y, x_w(\theta_w)W, d\tau) = \left[1 - T'_w(y) - d\tau\right] x_w(\theta_w) W - \phi'_w\left(\frac{y}{x_w(\theta_w)W}\right),$$

where $\psi(y, x_w(\theta_w)W, 0) = 0$ is exactly the first order condition to labor supply.[34] For the sake of conve-

[33] See, e.g., Saez (2001), Golosov, Tsyvinski and Werquin (2014), Lehmann, Simula and Trannoy (2014) and Sachs, Tsyvinski and Werquin (2016).

[34] See, e.g., Jacquet and Lehmann (2016), Scheuer and Werning (2016).

nience, we give the following terms:

$$\begin{aligned}
\psi_y(y, x_w(\theta_w)W, 0) &= -T^{w\prime\prime}(y)\theta_w W - \frac{\phi_w''\left(\frac{y}{x_w(\theta_w)W}\right)}{x_w(\theta_w)W}, \\
\psi_w(y, x_w(\theta_w)W, 0) &= 1 - T_w'(y) + y\frac{\phi_w''\left(\frac{y}{x_w(\theta_w)W}\right)}{(x_w(\theta_w)W)^2}, \\
\psi_\tau(y, x_w(\theta_w)W, 0) &= -x_w(\theta_w)W.
\end{aligned}$$

Under our utility function, which satisfies the single cross condition, Assumption 1 ensures that the first-order condition corresponds to a unique global maximum. Applying the implicit function theorem to $\psi(y(\theta_w), x_w(\theta_w)W, 0) = 0$ yields the nonlinear elasticities of labor income with respect to net-of-tax rate and wage in the equilibrium:

$$\begin{aligned}
\varepsilon_{1-\tau}^{y}(\theta_w) &\equiv \frac{\psi_\tau(y(\theta_w), x_w(\theta_w)W, 0)}{\psi_y(y(\theta_w), x_w(\theta_w)W, 0)}\frac{1 - T_w'(y(\theta_w))}{y(\theta_w)} \\
&= \frac{\varepsilon_w(\theta_w)}{\frac{T_w''(y(\theta_w))}{1-T_w'(y(\theta_w))}y(\theta_w)\varepsilon_w(\theta_w) + 1}
\end{aligned} \tag{A36}$$

and

$$\begin{aligned}
\varepsilon_w^y(\theta_w) &\equiv \frac{d\ln y(\theta_w)}{d\ln[x_w(\theta_w)W]} = -\frac{\psi_w(y(\theta_w), x_w(\theta_w)W, 0)}{\psi_y(y(\theta_w), x_w(\theta_w)W, 0)}\frac{x_w(\theta_w)W}{y(\theta_w)} \\
&= \frac{1 + \varepsilon_w(\theta_w)}{\frac{T^{w\prime\prime}(y(\theta_w))}{1-T_w'(y(\theta_w))}y(\theta_w)\varepsilon_w(\theta_w) + 1} \\
&= \varepsilon_{1-\tau}^{y}(\theta_w)\left[1 + \frac{1}{\varepsilon_w(\theta_w)}\right],
\end{aligned} \tag{A37}$$

respectively, where $\phi_w''\left(\frac{y(\theta_w)}{x_w(\theta_w)W}\right)$ is substituted by $\varepsilon_w(\theta_w) = \frac{\phi_w'(l_w(\theta_w))}{l_w(\theta_w)\phi''(l_w(\theta_w))}$. Note that given the tax policies, the nonlinear elasticities of labor income with respect to the wage in the equilibrium is the percentage change in an individual's labor income with respect to the 1% increase in wages in the equilibrium. Thus, $\varepsilon_w^y(y(\theta_w))$ is equal to $\frac{d\ln y(\theta_w)}{d\ln[x_w(\theta_w)W]}$.

A.6.2 Entrepreneur

We define the linear entrepreneural effort elasticity as

$$\varepsilon_e \equiv \frac{\phi_e'(l_e(\theta))}{l_e(\theta)\phi_e''(l_e(\theta))}. \tag{A38}$$

Note that we have assumed that the linear elasticity is constant in model setup to simplify the notation. We define the non-linear elasticity of profit with respect to net-tax income rate by (39). To understand the elasticity, consider the following tax reform.

Again, consider a small increase (i.e., $d\tau$) in the marginal tax rate faced by θ_e agent, which has no first order effects on aggregate values and the actions of other types. Then, by firm's problem, we have the following first order conditions:

$$WL_w = PQ_{ij}\frac{\xi}{\mu(\theta_e)},$$

and

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

where the second equation is derived by $WL_w = PQ_{ij}\frac{\xi}{\mu(\theta_e)}$. 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 to the net profit tax rate.

Suppose that a θ type firm treat $\mu(\theta)$, W and other firms' outputs (i.e., output other than $Q_{ij}(\theta)$) as given, it's reaction to the tax reform can be described by differential equations of the first order conditions. On one hand,

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

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

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

i.e.,

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

On the other hand, by $WL_w = PQ_{ij}\frac{\xi}{\mu(\theta_e)}$, we have

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

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

$$\frac{dL_w}{L_w} = \frac{1}{\mu(\theta_e)}\left[\frac{dl_e}{l_e} + \xi\frac{dL_w}{L_w}\right]. \tag{A40}$$

Combination of (A39) and (A40) gives

$$\frac{dL_w}{L_w} = \frac{-\frac{d\tau}{1-T_e'(\pi)}}{\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu\left(\theta_e\right)-\xi\right]-\left[1-\frac{\pi T_e''(\pi)}{1-T_e'(\pi)}\right]}. \tag{A41}$$

Lastly, by $\pi = PQ_{ij} - WL_w$ 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

$$\begin{aligned}
\frac{d\pi}{\pi} &= \frac{PQ_{ij}}{\pi}\frac{\xi}{\mu\left(\theta_e\right)}\frac{dL_w}{L_w} + \frac{PQ_{ij}}{\pi}\frac{1}{\mu\left(\theta_e\right)}\frac{dl_e}{l_e} - \frac{WL_w}{\pi}\frac{dL_w}{L_w} \\
&= \left[\frac{\frac{\xi}{\mu(\theta_e)}}{1-\frac{\xi}{\mu(\theta_e)}} - \frac{\frac{\xi}{\mu(\theta_e)}}{1-\frac{\xi}{\mu(\theta_e)}}\right]\frac{dL_w}{L_w} + \frac{\frac{1}{\mu(\theta_e)}}{1-\frac{\xi}{\mu(\theta_e)}}\frac{dl_e}{l_e} \\
&= \frac{\frac{1}{\mu(\theta_e)}}{1-\frac{\xi}{\mu(\theta_e)}}\frac{dl_e}{l_e} \\
&= \frac{dL_w}{L_w}
\end{aligned} \tag{A42}$$

where the last equation is derived by (A40). Equivalently, we can get the above equation by $WL_w = \left(\pi + WL_w\right)\frac{\xi}{\mu(\theta_e)}$, which is a combination of $\pi = PQ_{ij} - WL_w$ and $WL_w = PQ_{ij}\frac{\xi}{\mu(\theta_e)}$.

Combination of (A41) and (A42) gives

$$\frac{\frac{d\pi(\theta_e)}{\pi(\theta_e)}}{-\frac{d\tau}{1-T_e'(\pi(\theta_e))}} = \frac{1}{\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu\left(\theta_e\right)-\xi\right]-\left[1-\frac{\pi(\theta_e)T_e''(\pi(\theta_e))}{1-T_e'(\pi(\theta_e))}\right]} = \varepsilon^{\pi}_{1-\tau_e}\left(\pi\left(\theta_e\right)\right), \tag{A43}$$

where the last equation is derived by definition. Since the left side of the above equation reflect the elasticity of profit with respect to the tax reform, we making the definition of non-linear labor supply by (39).

A.6.3 Price Elasticity

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

$$P(\theta_e) = \chi(\theta_e)A^{\frac{\sigma-1}{\sigma}}Q_{ij}\left(\theta_e\right)^{-\frac{1}{\sigma}}Q^{\frac{1}{\sigma}}, \tag{A44}$$

and the inverse demand function

$$P(Q_{ij},\theta_e)=\chi(\theta_e)A^{\frac{\sigma-1}{\sigma}}Q_{ij}^{-\frac{1}{\eta(\theta_e)}}I^{-\left[\frac{1}{\eta(\theta_e)}-\frac{1}{\sigma}\right]\frac{\eta(\theta_e)}{\eta(\theta_e)-1}}\left[\begin{array}{c}(I-1)\,Q_{ij}\left(\theta_e\right)^{\frac{\eta(\theta_e)-1}{\eta(\theta_e)}}\\+Q_{ij}^{\frac{\eta(\theta_e)-1}{\eta(\theta_e)}}\end{array}\right]^{\left[\frac{1}{\eta(\theta_e)}-\frac{1}{\sigma}\right]\frac{\eta(\theta_e)}{\eta(\theta_e)-1}}Q^{\frac{1}{\sigma}}. \tag{A45}$$

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

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

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

$$\begin{aligned}&\frac{\partial\ln P(Q_{ij},\theta_e)}{\partial\theta_e}|_{Q_{ij}=Q_{ij}(\theta_e)} \\ =\ &\frac{d\ln P(\theta_e,Q_{ij}\left(\theta_e\right))}{d\theta_e}-\frac{\partial\ln P(\theta_e,Q_{ij}\left(\theta_e\right))}{\partial\ln Q_{ij}\left(\theta_e\right)}\frac{d\ln Q_{ij}\left(\theta_e\right)}{d\theta_e} \\ =\ &\frac{\chi(\theta_e)}{\chi(\theta_e)}-\frac{1}{\sigma}\frac{Q'_{ij}\left(\theta_e\right)}{Q_{ij}\left(\theta_e\right)}-\left[-\frac{1}{\eta\left(\theta_e\right)}+\frac{1}{I}\left(\frac{1}{\eta\left(\theta_e\right)}-\frac{1}{\sigma}\right)\right]\frac{Q'_{ij}\left(\theta_e\right)}{Q_{ij}\left(\theta_e\right)} \\ =\ &\frac{\chi(\theta_e)}{\chi(\theta_e)}+\left[\frac{1}{\eta\left(\theta_e\right)}-\frac{1}{\sigma}\right]\left(1-\frac{1}{I}\right)\frac{Q'_{ij}\left(\theta_e\right)}{Q_{ij}\left(\theta_e\right)} \\ =\ &\frac{\chi(\theta_e)}{\chi(\theta_e)}+\left[\frac{\sigma-1}{\sigma}-\frac{1}{\mu\left(\theta_e\right)}\right]\frac{d\ln Q_{ij}\left(\theta_e\right)}{d\theta_e},\end{aligned} \tag{A47}$$

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.6.4 Wage Elasticity

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

$$\varepsilon^{\varpi}_{L_w}(\theta_e)=\frac{\partial\ln\varpi(\theta_e)}{\partial\ln L_w(\theta_e)}-\varepsilon^{\varpi}_{L_w}(\theta_e,\theta_e),\quad\text{and}\quad\varepsilon^{\varpi}_{l_e}(\theta_e)=\frac{\partial\ln\varpi(\theta_e)}{\partial\ln l_e(\theta_e)}-\varepsilon^{\varpi}_{l_e}(\theta_e,\theta_e) \tag{A48}$$

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

$$\varepsilon_{L_w}^{\varpi}(\theta_e',\theta_e) = \begin{cases} \frac{\partial \ln \varpi(\theta_e')}{\partial \ln L_w(\theta_e)}, & \theta_e' \neq \theta_e, \\ \lim_{\theta_e' \to \theta_e} \frac{\partial \ln \varpi(\theta_e')}{\partial \ln L_w(\theta_e)}, & \theta_e' = \theta_e; \end{cases} \tag{A49}$$

$$\varepsilon_{l_e}^{\varpi}(\theta_e',\theta_e) = \begin{cases} \frac{\partial \ln \varpi(\theta_e')}{\partial \ln L_e(\theta_e)}, & \theta_e' \neq \theta_e, \\ \lim_{\theta_e' \to \theta_e} \frac{\partial \ln \varpi(\theta_e')}{\partial \ln l_e(\theta_e)}, & \theta_e' = \theta_e; \end{cases} \tag{A50}$$

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, $\varepsilon_{L_w}^{\varpi}(\theta_e',\theta_e)$ and $\varepsilon_{l_e}^{\varpi}(\theta_e',\theta_e)$ are not dependent on θ_e'. Specifically, we have

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

Then, by definition, we have

$$\varepsilon_{L_w}^{\varpi}(\theta_e) = \xi\left(1 - \frac{1}{\sigma}\right) - 1 < 0, \quad \text{and} \quad \varepsilon_{l_e}^{\varpi}(\theta) = 1 - \frac{1}{\sigma} > 0. \tag{A51}$$

Note that both $\varepsilon_{L_w}^{\varpi}(\theta_e)$ and $\varepsilon_{l_e}^{\varpi}(\theta)$ 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 $\lim_{\theta_e \to \theta_e'} \frac{\partial V_e(\theta_e'|\theta_e)}{\partial \theta_e'} = 0$, i.e.,

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

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, 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 $\theta_e' = \theta_e$, then decreases with θ_e'. In conclusion, there is unique local maximum point, which 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

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

we have

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

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

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

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

where $L_w(\theta_e'|\theta_e)$ is the optimal labor input given that θ_e entrepreneur reports θ_e'. The inverse demand function $P(Q_{ij}, \theta_e)$ is given by (A45). In the following proof of part (i), we refer to $P(Q_{ij}, \theta_e)$ and Q_{ij} as the short forms of $P\left(Q_{ij}\left(x_e(\theta_e)\, l_e(\theta_e'|\theta_e), L_w(\theta_e'|\theta_e)\right), \theta_e\right)$, and $Q_{ij}\left(x_e(\theta_e)\, l_e(\theta_e'|\theta_e), L_w(\theta_e'|\theta_e)\right)$, respectively. By (B4), we have

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

and

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

In addition, we have

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

and

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

By (A47), i.e.,

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

$\frac{\partial \ln P\left(Q_{ij},\theta_e\right)}{\partial \theta_e}$ is independent of Q_{ij} (note that the $\frac{Q'_{ij}(\theta_e)}{Q_{ij}(\theta_e)}$ in 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 the 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 needed to finish the θ'_e task by θ_e entrepreneur, 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, combination of $(B12)$, $(B13)$, and $(B14)$ suggests that

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

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 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 1-3. 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 utility, which means (33) and (35) should be satisfied (see the subsections on incentive compatibility).

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

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

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

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 income taxes. We taking the construction of $T_w(\cdot)$ as an example. To do this, we substituting

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

into

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

and show that there exist $T_w\left(y(\theta_w)\right)$ so that given allocation $\mathcal{A}$ and 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

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

According to the F.O.Cs of agent's problem, 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 is 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 incentive condition and the second equation is derived by the definition of $V_w(\theta)$. In conclusion, given the allocation, we can construct 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 now show that the allocation $\mathcal{A}$ and price $\mathcal{P}$ satisfying parts 1-3 and the constructed tax system $\mathcal{T}$ construct a SCCTE. Firstly, the allocation satisfies the incentive conditions (33) and (35). Thus, according to the analysis in subsections given before (see Lemma 1 for example), the allocation is consistent with agents' optimal choice. Secondly, the price $\mathcal{P}$ satisfying (16) and (17). Thirdly, 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 constructions of income taxes. In conclusion, the constructed tax system $\mathcal{T}$ together with the given allocation $\mathcal{A}$ and price $\mathcal{P}$ construct a SCCTE.

B.3 Proof of Lemma 3

We provide the following lemma, which is useful in expressing the optimal tax rate in sufficient statistics.

Lemma 3 *Suppose that the markup within the interval* $(\underline{\theta}_e^*,\overline{\theta}_e^*)\in\Theta_e^2$ *is constant, we have the following result under our benchmark model in the equilibrium:*

$$\frac{\pi'\left(\theta_e\right)}{\pi\left(\theta_e\right)}\left[1-\frac{\pi\left(\theta_e\right)T_e''\left(\pi\left(\theta_e\right)\right)}{1-T_e'\left(\pi\left(\theta_e\right)\right)}\right]=\frac{l_e'\left(\theta_e\right)}{l_e\left(\theta_e\right)}\frac{1+\varepsilon_e}{\varepsilon_e}, \tag{B17}$$

and

$$\varepsilon_{1-\tau_e}^{\pi}\left(\pi\left(\theta_e\right)\right)=\frac{\frac{d\ln\pi(\theta_e)}{d\theta_e}}{\frac{1+\varepsilon_e}{\varepsilon_e}\left[\frac{x_e'(\theta_e)}{x_e(\theta_e)}+\mu\left(\theta_e\right)\frac{\partial\ln P\left(Q_{ij}(\theta_e),\theta_e\right)}{\partial\theta_e}\right]},\forall\theta_e\in(\underline{\theta}_e^*,\overline{\theta}_e^*). \tag{B18}$$

Proof: By first order conditions

$$1-\tau_e(\theta_e)=\frac{\phi_e'(l_e(\theta_e))}{\frac{P(\theta_e)}{\mu(\theta_e)}\frac{\partial Q_{ij}(\theta_e)}{\partial l_e(\theta_e)}},$$

and

$$\begin{aligned}\frac{P(\theta_e)}{\mu(\theta_e)}\frac{\partial Q_{ij}(\theta_e)}{\partial[x_e(\theta_e)\,l_e(\theta_e)]} &= \frac{\pi(\theta_e)}{\mu(\theta_e)}\frac{P\left(Q_{ij}(\theta),\theta\right)Q_{ij}(\theta)}{[x_e(\theta_e)\,l_e(\theta_e)]\,\pi(\theta_e)}\frac{\partial\ln Q_{ij}(\theta_e)}{\partial\ln[x_e(\theta_e)\,l_e(\theta_e)]}\\ &= \frac{\pi(\theta_e)}{\mu(\theta_e)}\frac{1}{\left[1-\frac{\xi}{\mu(\theta_e)}\right](1-t_s)}\frac{1}{[x_e(\theta_e)\,l_e(\theta_e)]},\end{aligned}$$

we have

$$\left[1-T_e'(\pi(\theta_e))\right]\pi(\theta_e)=[\mu(\theta_e)-\xi]\,\phi_e'(l_e(\theta_e))\,l_e(\theta_e),$$

where we have substitute $1-\tau_e(\theta_e)$ by $[1-T_e'(\pi(\theta_e))]$.

Taking derivation of both sides of the above equation with respect to θ_e gives

$$\begin{aligned}&\left[1-T_e'(\pi(\theta_e))\right]\pi'(\theta_e)-T_e''(\pi(\theta_e))\,\pi(\theta_e)\,\pi'(\theta_e)=\\ &[\mu(\theta_e)-\xi]\,\phi_e'(l_e(\theta_e))\left[1+\frac{\phi_e''(l_e(\theta_e))\,l_e(\theta_e)}{\phi_e'(l_e(\theta_e))}\right]\frac{l_e'(\theta_e)}{l_e(\theta_e)}+\mu'(\theta_e)\,\phi_e'(l_e(\theta_e))\,l_e(\theta_e).\end{aligned}$$

Dividing the left side by $[1-T_e'(\pi(\theta_e))]\,\pi(\theta_e)$ and the right side by $[\mu(\theta_e)-\xi]\,\phi_e'(l_e(\theta_e))\,l_e(\theta_e)$ gives

$$\frac{\pi'(\theta_e)}{\pi(\theta_e)}\left[1-\frac{\pi(\theta_e)\,T_e''(\pi(\theta_e))}{1-T_e'(\pi(\theta_e))}\right]=\frac{l_e'(\theta_e)}{l_e(\theta_e)}\frac{1+\varepsilon_e}{\varepsilon_e}+\frac{\mu'(\theta_e)}{\mu(\theta_e)}\frac{1}{1-\xi/\mu(\theta_e)},\tag{B19}$$

where $\frac{\phi_e''(l_e(\theta_e))l_e(\theta_e)}{\phi_e'(l_e(\theta_e))}$ is substituted by $\frac{1}{\varepsilon_e}$ and $\frac{\pi T_e''(\pi)}{1-T_e'(\pi)}$ catches the progressivity of profit income tax.

On the other hand, notice that $\frac{\pi(\theta_e)}{P\left(Q_{ij}(\theta),\theta\right)Q_{ij}(\theta)(1-t_s)}=1-\frac{\xi}{\mu(\theta_e)}$, we have

$$\begin{aligned}\pi'(\theta) &= \left[\frac{\partial P\left(Q_{ij}(\theta),\theta\right)}{\partial\theta}Q_{ij}(\theta)+\frac{\partial\left[P\left(Q_{ij}(\theta),\theta\right)Q_{ij}(\theta)\right]}{\partial Q_{ij}(\theta)}\frac{\partial Q_{ij}(\theta)}{\partial(x_e(\theta)\,l_e(\theta))}\frac{d(x_e(\theta)\,l_e(\theta))}{d\theta}\right](1-t_s)\\ &\quad+\left[\frac{\partial\left[P\left(Q_{ij}(\theta),\theta\right)Q_{ij}(\theta)\right]}{\partial Q_{ij}(\theta)}\frac{\partial Q_{ij}(\theta)}{\partial L_w(\theta)}(1-t_s)-W\right]L_w'(\theta)\\ &= \left[\frac{\partial P\left(Q_{ij}(\theta),\theta\right)}{\partial\theta}Q_{ij}(\theta)+\frac{\partial\left[P\left(Q_{ij}(\theta),\theta\right)Q_{ij}(\theta)\right]}{\partial Q_{ij}(\theta)}\frac{\partial Q_{ij}(\theta)}{\partial(x_e(\theta)\,l_e(\theta))}\frac{d(x_e(\theta)\,l_e(\theta))}{d\theta}\right](1-t_s)\\ &= \left[\frac{\partial P\left(Q_{ij}(\theta),\theta\right)}{\partial\theta}Q_{ij}(\theta)+\frac{P(\theta)}{\mu(\theta)}\frac{\partial Q_{ij}(\theta)}{\partial(x_e(\theta)\,l_e(\theta))}\frac{d(x_e(\theta)\,l_e(\theta))}{d\theta}\right](1-t_s),\end{aligned}$$

thus

$$\frac{\pi'(\theta_e)}{\pi(\theta_e)}=\frac{\partial\ln P\left(Q_{ij}(\theta_e),\theta_e\right)}{\partial\theta_e}\frac{\mu(\theta_e)}{\mu(\theta_e)-\xi}+\frac{1}{\mu(\theta_e)-\xi}\frac{d\ln[x_e(\theta_e)\,l_e(\theta_e)]}{d\theta_e}.\tag{B20}$$

Combination of (B20) and (B19) gives

$$\mu\left(\theta_e\right)\frac{\partial \ln P\left(Q_{ij}(\theta_e),\theta_e\right)}{\partial \theta_e}+\frac{x_e'\left(\theta_e\right)}{x_e\left(\theta_e\right)} = \frac{\pi'\left(\theta_e\right)}{\pi\left(\theta_e\right)}\left[\mu\left(\theta_e\right)-\xi-\left[1-\frac{\pi\left(\theta_e\right)T_e''\left(\pi\left(\theta_e\right)\right)}{1-T_e'\left(\pi\left(\theta_e\right)\right)}\right]\frac{\varepsilon_e}{1+\varepsilon_e}\right] + \frac{\mu'\left(\theta_e\right)}{\mu\left(\theta_e\right)-\xi}\frac{\varepsilon_e}{1+\varepsilon_e}. \tag{B21}$$

If $\mu\left(\theta_e\right)$ is constant on $\theta_e\in(\underline{\theta}_e^*,\overline{\theta}_e^*)$, then for any $\theta_e\in(\underline{\theta}_e^*,\overline{\theta}_e^*)$, we have (B17) and

$$\mu\left(\theta_e\right)\frac{\partial \ln P\left(Q_{ij}(\theta_e),\theta_e\right)}{\partial \theta_e}+\frac{x_e'\left(\theta_e\right)}{x_e\left(\theta_e\right)}=\frac{\pi'\left(\theta_e\right)}{\pi\left(\theta_e\right)}\left[\mu\left(\theta_e\right)-\xi-\left[1-\frac{\pi\left(\theta_e\right)T_e''\left(\pi\left(\theta_e\right)\right)}{1-T_e'\left(\pi\left(\theta_e\right)\right)}\right]\frac{\varepsilon_e}{1+\varepsilon_e}\right]. \tag{B22}$$

Combination of (39) and (B22) gives (B18).

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 problemn.[35] The Lagrangian function for planner's problem is

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

where $\frac{\chi'(\theta_e)}{\chi(\theta_e)} + \varepsilon_{Q_{-ij}}(\theta_e)\delta\left(\theta_e\right) = \frac{\partial \ln P\left(Q_{ij}(\theta_e),\theta_e\right)}{\partial\theta_e}$. 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 $\ln Q_{ij}\left(\theta_e\right)$ 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 partially integrals yields:

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

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

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

[35]See Luenberger (1969) for details about the Lagrangian techniques. See Mirrlees (1976), Golosov (2016), Findeisen and Sachs (2017) for its application in the field of public economics.

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

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

Thus

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

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

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

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

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

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

and

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

C.1.2 Social Welfare Weight

Unless otherwise specified, the following equations in this sub-section are derived for any $\theta_o \in \Theta_o$. According to $\frac{\partial \pounds}{\partial V_o(x)}$ 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 monetary marginal social welfare weight for θ_o agent of o occupation. Set

$$\bar{g}_o(\theta_o) = \frac{\int_{\theta_o}^{\bar{\theta}_o} g(x) \widetilde{f}_o\left(x\right) dx}{1 - F_o(\theta_o)} \tag{C11}$$

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

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

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

Taking integration and utilize the boundary conditions gives

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

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

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

where the second equation is derived by

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

In addition, we have

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

Substituting $\psi_e(\theta_e)$ and $\psi_e'(\theta_e)$ by (C12) and (C13), we have

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

and

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

C.2 Proof of Theorem 1

(i) Unless otherwise specified, the following equations in this sub-section are derived for any $\theta_o \in \Theta_o$. According to $\frac{\partial £}{\partial L_w(\theta_e)}$, one has:

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

where $\int_{\theta_e} \varphi'(\theta'_e) \frac{\partial \ln \varpi(\theta'_e)}{\partial \ln L_w(\theta_e)} dx' = \varphi'(\theta_e) \varepsilon^{\varpi}_{L_w}(\theta_e)$ since $\varepsilon^{\varpi}_{L_w}(\theta'_e, \theta_e)$ is independent on 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 $\frac{W\mu(\theta)}{1-t_s}$ gives

$$\frac{W\mu(\theta)}{1-t_s} = \frac{\lambda'}{\lambda} - \frac{\kappa'(\theta_e)\xi}{\lambda L_w(\theta_e) N_e f_e(\theta_e)} + \frac{\varphi'(\theta_e)\varepsilon^{\varpi}_{L_w}(\theta_e)}{\lambda L_w(\theta_e) N_e f_e(\theta_e)}. \tag{C19}$$

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

$$W\int_{\theta_e} \frac{\mu(\theta_e)}{1-t_s} \frac{L_w(\theta_e) N_e f_e(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)} d\theta_e = \frac{\lambda'}{\lambda}\int_{\theta_e} \frac{L_w(\theta_e) N_e f_e(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)} d\theta_e - \int_{\theta_e} \frac{\kappa'(\theta_e)}{\lambda \varepsilon^{\varpi}_{L_w}(\theta_e)} \xi d\theta_e,$$

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

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

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

Define

$$\varepsilon^{Q_{ij}}_{1-\tau}(\theta_e) \equiv \frac{\frac{\xi}{\varepsilon^{\varpi}_{L_w}(\theta_e)}}{\int_{\theta_e} \frac{\mu(\theta_e)}{1-t_s} \frac{WL(\theta_e)N_e f_e(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)} d\theta_e} \tag{C21}$$

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

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

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

According to $\frac{\partial \pounds}{\partial l_w(\theta_w)}$, 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 $\frac{\phi_w'(l_w(\theta_w))}{x_w(\theta_w)}$ by $[1-\tau_w(\theta_w)]\,W$ gives:

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

Secondly, combination of (C24), (C13), and (C22) gives

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

In addition, by the definitions of income elasticities, i.e., (A36) and (A37), we have $\frac{d\ln y(\theta_w)}{d\ln[x_w(\theta_w)W]} = \varepsilon_{1-\tau}^{y}(\theta_w)\left[1+\frac{1}{\varepsilon_w}\right]$. Thus

$$\frac{1}{1-\tau_w(\theta_w)} = \frac{1}{\frac{\mu}{1-t_s}}\left[1+[1-\bar{g}_w(\theta_w)]\frac{1-F_y(y(\theta_w))}{f_y(y(\theta_w))y(\theta_w)}\frac{1}{\varepsilon_{1-\tau}^{y}(\theta_w)}\right]. \tag{C26}$$

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

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

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

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

Then, one has:

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

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

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

or

$$\frac{1}{1-\widetilde{\tau}_e(\theta_e)} = 1-\frac{\psi_e(\theta_e)}{\lambda N_e f_e(\theta_e)}\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu(\theta_e)\frac{\partial \ln P(Q(\theta_e),\theta_e)}{\partial\theta_e}+\frac{x_e'(\theta_e)}{x_e(\theta_e)}\right] - \frac{1}{1-\widetilde{\tau}_e(\theta_e)}\frac{1}{P(\theta_e)Q_{ij}(\theta_e)}\left[\frac{\kappa'(\theta_e)}{\lambda N_e f_e(\theta_e)}-\frac{\varphi'(\theta_e)}{\lambda N_e f_e(\theta_e)}\varepsilon_{l_e}^{\varpi}(\theta_e)\right]. \tag{C29}$$

Combination of (C28) and (C13) gives

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

Using (C19)[36] to substitute $\frac{\varphi'(\theta_e)\varepsilon_{l_e}^{\varpi}(\theta_e)}{\lambda N_e f_e(\theta_e)}$, we have

$$\frac{\widetilde{\tau}_e(\theta_e)}{1-\widetilde{\tau}_e(\theta_e)} = [1-\bar{g}_e(\theta_e)]\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu(\theta_e)\frac{\partial \ln P(Q(\theta_e),\theta_e)}{\partial\theta_e}+\frac{x_e'(\theta_e)}{x_e(\theta_e)}\right] - \frac{1}{1-\widetilde{\tau}_e(\theta_e)}\frac{1}{P(\theta_e)Q_{ij}(\theta_e)}\frac{\kappa'(\theta_e)}{\lambda N_e f_e(\theta_e)}\left[1-\xi\frac{\varepsilon_{l_e}^{\varpi}(\theta_e)}{\varepsilon_{L_w}^{\varpi}(\theta_e)}\right] - \frac{1}{1-\widetilde{\tau}_e(\theta_e)}\frac{L_w(\theta_e)}{P(\theta_e)Q_{ij}(\theta_e)}\frac{\lambda'}{\lambda}\left[1-\frac{\lambda W}{\lambda'}\mu(\theta_e)\right]\frac{\varepsilon_{l_e}^{\varpi}(\theta_e)}{\varepsilon_{L_w}^{\varpi}(\theta_e)}. \tag{C30}$$

We now transform the three terms in the right side of the above equations one by one. Firstly, by (B21),

$$\begin{aligned} &[1-\bar{g}_e(\theta_e)]\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu(\theta_e)\frac{\partial \ln P(Q(\theta_e),\theta_e)}{\partial\theta_e}+\frac{x_e'(\theta_e)}{x_e(\theta_e)}\right] \\ =\ &[1-\bar{g}_e(\theta_e)]\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{\pi'(\theta_e)}{\pi(\theta_e)}\left[\frac{1+\varepsilon_e}{\varepsilon_e}[\mu(\theta_e)-\xi]-\left[1-\frac{\pi(\theta_e)T_e''(\pi(\theta_e))}{1-T_e'(\pi(\theta_e))}\right]\right] \\ &+[1-\bar{g}_e(\theta_e)]\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{\partial \ln\mu(\theta_e)}{\partial\theta_e}\frac{\mu(\theta_e)}{\mu(\theta_e)-\xi}, \end{aligned} \tag{C31}$$

[36] Which suggests $\frac{\varphi'(\theta_e)\gamma_{own}^{\varpi,L_e}(\theta_e)}{\lambda_1 N_e f_e(\theta_e)} = \left[\left[W\mu(\theta_e)-\frac{\lambda_2}{\lambda_1}\right]L_w(\theta_e)+\frac{\chi_1(\theta_e)}{\lambda_1 N_e f_e(\theta_e)}\xi\right]\frac{\gamma_{own}^{\varpi,L_e}(\theta_e)}{\gamma_{own}^{\varpi,L_w}(\theta_e)}$.

where $\frac{\mu(\theta)}{\mu(\theta)-\xi} = \frac{P(\theta_e)Q_{ij}(\theta_e)(1-t_s)}{\pi(\theta)}$. Moreover, by the definition of non-linear elasticity of profit (39), we have

$$\begin{aligned} & [1-\bar{g}_e(\theta_e)] \frac{1-F_e(\theta_e)}{f_e(\theta_e)} \frac{\pi'(\theta_e)}{\pi(\theta_e)} \left[\frac{1+\varepsilon_e}{\varepsilon_e} [\mu(\theta_e) - \xi] - \left[1 - \frac{\pi(\theta_e) T_e''(\pi(\theta_e))}{1-T_e'(\pi(\theta_e))} \right] \right] \\ = \; & [1-\bar{g}_e(\theta_e)] \frac{1-F_\pi(\pi(\theta_e))}{\pi(\theta_e) f_\pi(\pi(\theta_e))} \frac{1}{\varepsilon^{\pi}_{1-\tau_e}(\pi)}. \end{aligned} \tag{C32}$$

Secondly, substituting $\kappa'(\theta_e)$ by (C18), we have the following equation:

$$\begin{aligned} & -\frac{1}{1-\widetilde{\tau}_e(\theta_e)} \frac{1}{P(\theta_e) Q_{ij}(\theta_e)} \frac{\kappa'(\theta_e)}{\lambda N_e f_e(\theta_e)} \\ = \; & \frac{1-\tau_e(\theta_e)}{1-\widetilde{\tau}_e(\theta_e)} \varepsilon_{Q_{-ij}}(\theta_e) \left[\begin{array}{c} [1-g_e(\theta_e)] - \\ \frac{[1-\bar{g}_e(\theta_e)][1-F_e(\theta_e)]}{f_e(\theta_e)} \left[\begin{array}{c} \frac{1+\varepsilon_e}{\varepsilon_e} \frac{l_e'(\theta_e)}{l_e(\theta_e)} \\ + \frac{d \ln\left[\mu(\theta_e)\varepsilon_{Q_{-ij}}(\theta_e)\right]}{d\theta_e} \end{array} \right] \end{array} \right]. \end{aligned} \tag{C33}$$

At last, notice that $\frac{L_w(\theta_e)W}{P(\theta_e)Q_{ij}(\theta_e)[1-\tau_s(\theta_e)]} = \frac{\xi}{\mu(\theta_e)}$ and $\frac{\lambda'}{\lambda W \frac{\mu}{1-t_s}} = 1$, the last term of (C30) equals

$$\begin{aligned} & -\frac{1}{1-\widetilde{\tau}_e(\theta_e)} \frac{L_w(\theta_e)}{P(\theta_e) Q_{ij}(\theta_e)} \frac{\lambda'}{\lambda} \left[1 - \frac{\lambda W}{\lambda'} \mu(\theta_e) \right] \frac{\varepsilon^{\varpi}_{l_e}(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)} \\ = \; & -\frac{1-t_s}{1-\widetilde{\tau}_e(\theta_e)} \frac{\xi}{\mu(\theta_e)} \frac{\mu}{1-t_s} \left[1 - \frac{\mu(\theta_e)}{\frac{\mu}{1-t_s}} \right] \frac{\varepsilon^{\varpi}_{l_e}(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)} \\ = \; & -\frac{1-t_s}{1-\widetilde{\tau}_e(\theta_e)} \xi \left[\frac{\frac{\mu}{1-t_s}}{\mu(\theta_e)} - 1 \right] \frac{\varepsilon^{\varpi}_{l_e}(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)}. \end{aligned} \tag{C34}$$

Combination of equations (C32) to (C34) gives

$$\begin{aligned} & \frac{\widetilde{\tau}_e(\theta_e)}{1-\widetilde{\tau}_e(\theta_e)} \\ = \; & [1-\bar{g}_e(\theta_e)] \frac{1-F_e(\theta_e)}{f_e(\theta_e)} \frac{1+\varepsilon_e}{\varepsilon_e} \left[\mu(\theta_e) \frac{\partial \ln P\left(Q_{ij}(\theta_e), \theta_e\right)}{\partial \theta_e} + \frac{x_e'(\theta_e)}{x_e(\theta_e)} \right] \\ & + \frac{1-\tau_e(\theta_e)}{1-\widetilde{\tau}_e(\theta_e)} \varepsilon_{Q_{-ij}}(\theta_e) \left[\begin{array}{c} [1-g_e(\theta_e)] - \\ \frac{[1-\bar{g}_e(\theta_e)][1-F_e(\theta_e)]}{f_e(\theta_e)} \left[\begin{array}{c} \frac{1+\varepsilon_e}{\varepsilon_e} \frac{l_e'(\theta_e)}{l_e(\theta_e)} \\ + \frac{d \ln\left[\mu(\theta_e)\varepsilon_{Q_{-ij}}(\theta_e)\right]}{d\theta_e} \end{array} \right] \end{array} \right] \left[1 - \xi \frac{\varepsilon^{\varpi}_{l_e}(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)} \right] \\ & + \frac{1-t_s}{1-\widetilde{\tau}_e(\theta_e)} \left[1 - \frac{\widetilde{\mu}}{\mu(\theta_e)} \right] \xi \frac{\varepsilon^{\varpi}_{l_e}(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)}. \end{aligned} \tag{C35}$$

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

$$\begin{aligned}
&\frac{1}{1-\tau_e(\theta_e)} \\
= \; &\left[1+[1-\bar{g}_e(\theta_e)]\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu(\theta_e)\frac{\partial \ln P\left(Q_{ij}(\theta_e),\theta_e\right)}{\partial\theta_e}+\frac{x_e'(\theta_e)}{x_e(\theta_e)}\right]\right]\frac{1-t_s}{\mu(\theta_e)} \\
&+\varepsilon_{Q_{-ij}}(\theta_e)\left[\begin{array}{c}[1-g_e(\theta_e)]- \\ \frac{[1-\bar{g}_e(\theta_e)][1-F_e(\theta_e)]}{f_e(\theta_e)}\left[\begin{array}{c}\frac{1+\varepsilon_e}{\varepsilon_e}\frac{l_e'(\theta_e)}{l_e(\theta_e)} \\ +\frac{d\ln\left[\mu(\theta_e)\varepsilon_{Q_{-ij}}(\theta_e)\right]}{d\theta_e}\end{array}\right]\end{array}\right]\left[1-\xi\frac{\varepsilon^{\varpi}_{l_e}(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)}\right] \\
&+\frac{1-t_s}{1-\tau_e(\theta_e)}\left[1-\frac{\widetilde{\mu}}{\mu(\theta_e)}\right]\xi\frac{\varepsilon^{\varpi}_{l_e}(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)}.
\end{aligned} \tag{C36}$$

Utilizing

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

and

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

we have

$$\begin{aligned}
\frac{1}{1-\tau_e(\theta_e)} \; = \; &\frac{1+[1-\bar{g}_e(\theta_e)]\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu(\theta_e)\frac{\partial\ln P(Q(\theta_e),\theta_e)}{\partial\theta_e}+\frac{x_e'(\theta_e)}{x_e(\theta_e)}\right]}{\mu(\theta_e)\,/\,(1-t_s)} \\
&+IRE(\theta_e)\;\overbrace{\left[1-\xi\frac{\varepsilon^{\varpi}_{l_e}(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)}\right]}^{\text{Elasticity of } Q_{ij} \text{ w.r.t } l_e} \\
&+\frac{1-t_s}{1-\tau_e(\theta_e)}\xi RE(\theta_e)\quad\overbrace{\left[-\frac{\varepsilon^{\varpi}_{l_e}(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)}\right]}^{\text{Elasticity of } L_w \text{ w.r.t } l_e}\quad,
\end{aligned} \tag{C37}$$

i.e., $\forall\theta_e\in\Theta_e$,

$$\frac{1}{1-\tau_e(\theta_e)}=\frac{\frac{1+[1-\bar{g}_e(\theta_e)]\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu(\theta_e)\frac{\partial\ln P(Q(\theta_e),\theta_e)}{\partial\theta_e}+\frac{x_e'(\theta_e)}{x_e(\theta_e)}\right]}{\mu(\theta_e)/(1-t_s)}+IRE(\theta_e)\left[1-\xi\frac{\varepsilon^{\varpi}_{l_e}(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)}\right]}{1+(1-t_s)\,RE(\theta_e)\,\xi\frac{\varepsilon^{\varpi}_{l_e}(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)}}. \tag{C38}$$

Substituting the elasticities by parameters and t_s by 0, we have the results given in Theorem 1.

C.3 Proof of Corollary 1

We use subscript 0 to represent the original variables. For $\theta_e \geq \theta_e^*$,

$$\frac{1-F_{\pi_0}\left(\pi_0\left(\theta_e\right)\right)}{\pi_0\left(\theta_e\right) f_{\pi_0}\left(\pi_0\left(\theta_e\right)\right)}=\frac{1-F_{\theta_e}\left(\theta_e\right)}{f_{\theta_e}\left(\theta_e\right)} \frac{\pi_0^{\prime}\left(\theta_e\right)}{\pi_0\left(\theta_e\right)} \tag{C39}$$

is constant. To derive $\frac{\pi_0(\theta_e)}{\pi_0(\theta_e)}$, we utilizing the first order conditions:

$$W=\frac{P(\theta_e)\overline{Q}\left(\theta_e\right)}{\mu\left(\theta_e\right)} \frac{\xi}{L_w\left(\theta_e\right)},$$

and

$$\frac{P_0(\theta_e)Q_{ij0}\left(\theta_e\right)}{\mu\left(\theta_e\right)}\left[1-\tau_0\left(\theta_e\right)\right]=l_{e,0}\left(\theta_e\right) \phi_e^{\prime}\left(l_{e,0}\left(\theta_e\right)\right)=l_{e,0}\left(\theta_e\right)^{1+\frac{1}{\varepsilon_e}},$$

where

$$P_0\left(\theta_e\right) Q_{ij0}\left(\theta_e\right)=\chi(\theta_e) A^{\frac{\sigma-1}{\sigma}} Q_{ij0}\left(\theta_e\right)^{1-\frac{1}{\sigma}} Q^{\frac{1}{\sigma}}=\chi(\theta_e) A^{\frac{\sigma-1}{\sigma}}\left[x_e(\theta_e) l_{e,0}\left(\theta_e\right) \cdot L_{w,0}\left(\theta_e\right)^{\xi}\right]^{1-\frac{1}{\sigma}} Q^{\frac{1}{\sigma}}, \theta_e \geq \theta_e^* .$$

Take the logarithm of both sides of the above equations, and then take the derivative of θ_e on both sides of the equations:

$$0=\frac{\chi^{\prime}(\theta_e)}{\chi(\theta_e)}-\frac{d \ln \mu\left(\theta_e\right)}{d \theta_e}+\frac{\sigma-1}{\sigma}\left[\frac{x_e^{\prime}(\theta_e)}{x_e(\theta_e)}+\frac{l_{e,0}^{\prime}\left(\theta_e\right)}{l_{e,0}\left(\theta_e\right)}+\xi \frac{L_{w,0}^{\prime}\left(\theta_e\right)}{L_{w,0}\left(\theta_e\right)}\right]-\frac{L_{w,0}^{\prime}\left(\theta_e\right)}{L_{w,0}\left(\theta_e\right)}$$

and

$$\left(1+\frac{1}{\varepsilon_e}\right) \frac{l_{e,0}^{\prime}\left(\theta_e\right)}{l_{e,0}\left(\theta_e\right)}=\frac{\chi^{\prime}(\theta_e)}{\chi(\theta_e)}-\frac{d \ln \mu\left(\theta_e\right)}{d \theta_e}+\frac{\sigma-1}{\sigma}\left[\frac{x_e^{\prime}(\theta_e)}{x_e(\theta_e)}+\frac{l_{e,0}^{\prime}\left(\theta_e\right)}{l_{e,0}\left(\theta_e\right)}+\xi \frac{L_{w,0}^{\prime}\left(\theta_e\right)}{L_{w,0}\left(\theta_e\right)}\right], \theta_e \geq \theta_e^* .$$

When $\tau_0\left(\theta_e\right)$ is constant for $\theta_e \geq \theta_e^*$, we can express $\frac{l_{e,0}^{\prime}(\theta_e)}{l_{e,0}(\theta_e)}$ and $\frac{L_w^{\prime}(\theta_e)}{L_w(\theta_e)}$ in exogenous parameters:

$$\begin{aligned}
\frac{l_{e,0}^{\prime}\left(\theta_e\right)}{l_{e,0}\left(\theta_e\right)} & =\frac{\frac{d \ln \mu(\theta_e)}{d \theta_e}-\frac{\chi^{\prime}(\theta_e)}{\chi(\theta_e)}-\frac{\sigma-1}{\sigma} \frac{x_e^{\prime}(\theta_e)}{x_e(\theta_e)}}{\frac{\sigma-1}{\sigma}\left[1+\xi\left(1+\frac{1}{\varepsilon_e}\right)\right]-\left(1+\frac{1}{\varepsilon_e}\right)}, \\
\frac{L_{w,0}^{\prime}\left(\theta_e\right)}{L_{w,0}\left(\theta_e\right)} & =\left(1+\frac{1}{\varepsilon_e}\right) \frac{l_{e,0}^{\prime}\left(\theta_e\right)}{l_{e,0}\left(\theta_e\right)}, \theta_e \geq \theta_e^* .
\end{aligned} \tag{C40}$$

Lastly, notice that

$$\pi_0(\theta)=P_0\left(\theta_e\right) Q_{ij0}\left(\theta_e\right)-W L_{w,0}\left(\theta_e\right)=P_0\left(\theta_e\right) Q_{ij0}\left(\theta_e\right)\left[1-\frac{\xi}{\mu\left(\theta_e\right)}\right],$$

we have

$$\begin{aligned}\frac{\pi_0'(\theta_e)}{\pi_0(\theta_e)} &= \frac{d\ln\left[P_0(\theta_e)Q_{ij0}(\theta_e)\right]}{d\theta_e}+\frac{d\ln\left[1-\frac{\xi}{\mu(\theta_e)}\right]}{d\theta_e} \\ &= \frac{\chi'(\theta_e)}{\chi(\theta_e)}+\frac{\sigma-1}{\sigma}\left[\frac{x_e'(\theta_e)}{x_e(\theta_e)}+\frac{l_{e,0}'(\theta_e)}{l_{e,0}(\theta_e)}+\xi\frac{L_{w,0}'(\theta_e)}{L_{w,0}(\theta_e)}\right]+\frac{d\ln\left[1-\frac{\xi}{\mu(\theta_e)}\right]}{d\theta_e}, \forall\theta_e\geq\theta_e^*.\end{aligned} \tag{C41}$$

One can see that $\frac{1-F_{\theta_e}(\theta_e)}{f_{\theta_e}(\theta_e)}\frac{\pi'(\theta_e)}{\pi(\theta_e)}$ is constant on $\theta_e \geq \theta_e'$ as long as the profit tax wedge is constant on $\theta_e \geq \theta_e'$ and $\theta_e' \geq \theta_e^*$. This is because for any $\theta_e \geq \theta_e^*$, $\frac{l_{e,0}'(\theta_e)}{l_{e,0}(\theta_e)}+\xi\frac{L_{w,0}'(\theta_e)}{L_{w,0}(\theta_e)}$ can be expressed in exogenous parameters when the profit tax rate is constant on $\theta_e \geq \theta_e^*$. Moreover, $\frac{1-F_{\theta_e}(\theta_e)}{f_{\theta_e}(\theta_e)}\frac{\pi'(\theta_e)}{\pi(\theta_e)}$ equals to $\overline{H}_{initial}$ when (i) $\theta_e \geq \theta_e'$; (ii) the profit tax wedge is constant on $\theta_e \geq \theta_e'$; and (iii) $\theta_e' \geq \theta_e^*$.

Suppose that there exist $\theta_e^\star \in \left(\underline{\theta}_e,\overline{\theta}_e\right)$, so that for any $\theta_e \geq \theta_e^\star$, $\tau_e(\theta_e)=\tau_e$ is constant, then for any $\theta_e \geq \theta_e^\star$, $\frac{1-F_{\theta_e}(\theta_e)}{f_{\theta_e}(\theta_e)}\frac{\pi'(\theta_e)}{\pi(\theta_e)}$ equals to $\overline{H}_{initial}$. When for any $\theta_e \geq \theta_e^\star$, $\bar{g}_e(\theta_e)=\bar{g}_e$, we have (48) for any $\theta_e \geq \{\theta_e^*,\theta_e^\star\}$.

■

C.4 Proof of Corollary 2

When social welfare function is Utilitarian, one has $\bar{g}_o(\theta_o)=1$ and $\widetilde{g}(\theta_e)=0$. Then the general optimal tax formulas given in Theorem 1 can be simplified to the formulas given in Corollary 2.

■

C.5 Proof of Corollary 3

Combination of equations (C32) to (C34) also gives

$$\begin{aligned}&\frac{\widetilde{\tau}_e(\theta_e)}{1-\widetilde{\tau}_e(\theta_e)} \\ =\ &[1-\bar{g}_e(\theta_e)]\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu(\theta_e)\frac{\partial\ln P(Q(\theta_e),\theta_e)}{\partial\theta_e}+\frac{x_e'(\theta_e)}{x_e(\theta_e)}\right] \\ &+\frac{1-\tau_e(\theta_e)}{1-\widetilde{\tau}_e(\theta_e)}\varepsilon_{Q_{-ij}}(\theta_e)\left[\begin{array}{c}[1-g_e(\theta_e)]- \\ \frac{[1-\bar{g}_e(\theta_e)][1-F_e(\theta_e)]}{f_e(\theta_e)}\left[\begin{array}{c}\frac{1+\varepsilon_e}{\varepsilon_e}\frac{l_e'(\theta_e)}{l_e(\theta_e)} \\ +\frac{d\ln\left[\mu(\theta_e)\varepsilon_{Q_{-ij}}(\theta_e)\right]}{d\theta_e}\end{array}\right]\end{array}\right]\left[1-\xi\frac{\varepsilon_{l_e}^{\varpi}(\theta_e)}{\varepsilon_{L_w}^{\varpi}(\theta_e)}\right] \\ &-\frac{1}{1-\widetilde{\tau}_e(\theta_e)}\left[1-\frac{\mu}{\mu(\theta_e)}\right]\xi\left[-\frac{\varepsilon_{l_e}^{\varpi}(\theta_e)}{\varepsilon_{L_w}^{\varpi}(\theta_e)}\right].\end{aligned} \tag{C42}$$

where

Substituting $\frac{1}{1-\tilde{\tau}_e(\theta_e)}$ by $\frac{\mu(\theta_e)}{1-\tau_e(\theta_e)}$, we have

$$\begin{aligned}
& \frac{1}{1-\tau_e(\theta_e)} \\
= \quad & \frac{1+[1-\bar{g}_e(\theta_e)]\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu(\theta_e)\frac{\partial \ln P(Q(\theta_e),\theta_e)}{\partial\theta_e}+\frac{x_e'(\theta_e)}{x_e(\theta_e)}\right]}{\mu(\theta_e)} \\
& +\varepsilon_{Q_{-ij}}(\theta_e)\left[\begin{array}{c}[1-g_e(\theta_e)]- \\ \frac{[1-\bar{g}_e(\theta_e)][1-F_e(\theta_e)]}{f_e(\theta_e)}\left[\begin{array}{c}\frac{1+\varepsilon_e}{\varepsilon_e}\frac{l_e'(\theta_e)}{l_e(\theta_e)} \\ +\frac{d\ln\left[\mu(\theta_e)\varepsilon_{Q_{-ij}}(\theta_e)\right]}{d\theta_e}\end{array}\right]\end{array}\right]\left[1-\xi\frac{\varepsilon_{l_e}^{\varpi}(\theta_e)}{\varepsilon_{L_w}^{\varpi}(\theta_e)}\right] \\
& -\left[1-\frac{\mu}{\mu(\theta_e)}\right]\xi\left[-\frac{\varepsilon_{l_e}^{\varpi}(\theta_e)}{\varepsilon_{L_w}^{\varpi}(\theta_e)}\right],
\end{aligned} \tag{C43}$$

where

$$\begin{aligned}
\frac{\partial \ln P(Q_{ij},\theta_e)}{\partial\theta_e}|_{Q_{ij}=Q_{ij}(\theta_e)} &= \frac{\chi'(\theta_e)}{\chi(\theta_e)}+\left[\frac{\sigma-1}{\sigma}-\frac{1}{\mu(\theta_e)}\right]\frac{d\ln Q_{ij}(\theta_e)}{d\theta_e} \\
\varepsilon_{Q_{-ij}}(\theta_e) &= \frac{\sigma-1}{\sigma}-\frac{1}{\mu(\theta_e)} \\
\frac{\varepsilon_{l_e}^{\varpi}(\theta_e)}{\varepsilon_{L_w}^{\varpi}(\theta_e)} &= \frac{1}{\xi-\frac{\sigma}{\sigma-1}}, \forall\theta_e\in\Theta.
\end{aligned} \tag{C44}$$

Now suppose that the tax rate is constant in the interval $\left(\underline{\theta}_e^*,\overline{\theta}_e^*\right)$. According to the proof of Corollary 1, one can see that

$$\begin{aligned}
\frac{l_e'(\theta_e)}{l_e(\theta_e)} &= \frac{\frac{d\ln\mu(\theta_e)}{d\theta_e}-\frac{\chi'(\theta_e)}{\chi(\theta_e)}-\frac{\sigma-1}{\sigma}\frac{x_e'(\theta_e)}{x_e(\theta_e)}}{\frac{\sigma-1}{\sigma}\left[1+\xi\left(1+\frac{1}{\varepsilon_e}\right)\right]-\left(1+\frac{1}{\varepsilon_e}\right)}, \\
\frac{L_w'(\theta_e)}{L_w(\theta_e)} &= \left(1+\frac{1}{\varepsilon_e}\right)\frac{l_e'(\theta_e)}{l_e(\theta_e)}, \\
\frac{Q_{ij}'(\theta_e)}{Q_{ij}(\theta_e)} &= \frac{x_e'(\theta_e)}{x_e(\theta_e)}+\frac{l_e'(\theta_e)}{l_e(\theta_e)}+\xi\frac{L_w'(\theta_e)}{L_w(\theta_e)}, \forall\theta_e\in\left(\underline{\theta}_e^*,\overline{\theta}_e^*\right).
\end{aligned} \tag{C45}$$

Futher assume that the markup is constant in the interval $\left(\underline{\theta}_e^*,\overline{\theta}_e^*\right)$, then in this interval, $\frac{d\ln\left[\mu(\theta_e)\varepsilon_{Q_{-ij}}(\theta_e)\right]}{d\theta_e}=0$, and $\frac{l_e'(\theta)}{l_e(\theta)}\frac{1+\varepsilon_e(\theta)}{\varepsilon_e(\theta)}=\frac{\pi'(\theta)}{\pi(\theta)}$ and $\varepsilon_{1-\tau_e}^{\pi}(\pi(\theta_e))=\frac{\frac{d\ln\pi(\theta_e)}{d\theta_e}}{\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu(\theta_e)\frac{\partial\ln P\left(Q_{ij}(\theta_e),\theta_e\right)}{\partial\theta_e}+\frac{x_e'(\theta_e)}{x_e(\theta_e)}\right]}=\frac{1}{\frac{1+\varepsilon_e}{\varepsilon_e}[\mu(\theta_e)-\xi]-1}$ (e.g., see

Lemma 3). Thus

$$\begin{aligned}
&\frac{1}{1-\tau_e\left(\theta_e\right)} \\
=\ & \frac{1+\left[1-\bar{g}_e(\theta_e)\right]\frac{1-F_\pi(\pi(\theta_e))}{\pi(\theta_e)f_\pi(\pi(\theta_e))}\left[\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu\left(\theta_e\right)-\xi\right]-1\right]}{\mu\left(\theta_e\right)}+ \\
& \varepsilon_{Q_{-ij}}(\theta_e)\left[\begin{array}{c}\left[1-g_e(\theta_e)\right]- \\ \left[1-\bar{g}_e(\theta_e)\right]\frac{1-F_\pi(\pi(\theta_e))}{\pi(\theta_e)f_\pi(\pi(\theta_e))}\end{array}\right]\left[1-\xi\frac{\varepsilon^{\varpi}_{l_e}(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)}\right] \\
& -\frac{1}{1-\tau_e\left(\theta_e\right)}\left[1-\frac{\mu}{\mu\left(\theta_e\right)}\right]\xi\left[-\frac{\varepsilon^{\varpi}_{l_e}(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)}\right],\forall\theta_e\in\left(\underline{\theta}_e^*,\overline{\theta}_e^*\right).
\end{aligned} \tag{C46}$$

Equivalently

$$\frac{1}{1-\tau_e\left(\theta_e\right)}=\frac{\left[\begin{array}{c}\frac{1}{\mu(\theta_e)}\left[1+\left[1-\bar{g}_e(\theta_e)\right]\frac{1-F_\pi(\pi(\theta_e))}{\pi(\theta_e)f_\pi(\pi(\theta_e))}\left[\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu\left(\theta_e\right)-\xi\right]-1\right]\right]+ \\ \varepsilon_{Q_{-ij}}(\theta_e)\left[\left[1-g_e(\theta_e)\right]-\left[1-\bar{g}_e(\theta_e)\right]\frac{1-F_\pi(\pi(\theta_e))}{\pi(\theta_e)f_\pi(\pi(\theta_e))}\right]\left[1-\xi\frac{\varepsilon^{\varpi}_{l_e}(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)}\right]\end{array}\right]}{1+\left[1-\frac{\mu}{\mu(\theta_e)}\right]\xi\left[-\frac{\varepsilon^{\varpi}_{l_e}(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)}\right]}, \tag{C47}$$

where $\theta_e\in\left(\underline{\theta}_e^*,\overline{\theta}_e^*\right)$. According to the definitions of elasticities, we have

$$\varepsilon_{Q_{-ij}}(\theta_e)\left[1-\xi\frac{\varepsilon^{\varpi}_{l_e}(\theta_e)}{\varepsilon^{\varpi}_{L_w}(\theta_e)}\right]=\left[\frac{\sigma-1}{\sigma}-\frac{1}{\mu\left(\theta_e\right)}\right]\left[1-\frac{\xi}{\xi-\frac{\sigma}{\sigma-1}}\right]=\frac{1-\frac{1}{\mu(\theta_e)}\frac{\sigma}{\sigma-1}}{\frac{\sigma}{\sigma-1}-\xi}. \tag{C48}$$

In conclusion, we have, $\forall\pi\in\left(\pi\left(\underline{\theta}_e^*\right),\pi\left(\overline{\theta}_e^*\right)\right)$,

$$\frac{1}{1-T'_\pi\left(\pi\right)}=\frac{\left[\begin{array}{c}\frac{1}{\mu_\pi(\pi)}\left[1+\left[1-\bar{g}_\pi(\pi)\right]\frac{1-F_\pi(\pi)}{\pi f_\pi(\pi)}\left[\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu\left(\theta_e\right)-\xi\right]-1\right]\right]+ \\ \frac{1-\frac{\sigma}{\sigma-1}\frac{1}{\mu_\pi(\pi)}}{\frac{\sigma}{\sigma-1}-\xi}\left[\left[1-g_\pi(\pi)\right]-\left[1-\bar{g}_\pi(\pi)\right]\frac{1-F_\pi(\pi)}{\pi f_\pi(\pi)}\right]\end{array}\right]}{1-\left[1-\frac{\mu}{\mu_\pi(\pi)}\right]\frac{\xi}{\xi-\frac{\sigma}{\sigma-1}}},$$

where $\mu_\pi\left(\pi\left(\theta_e\right)\right)\equiv\mu\left(\theta_e\right)$.

For top tax rate, we have

$$\frac{1}{1-\overline{\tau}_e}=\frac{\frac{1+(1-\bar{g})\overline{H}_{initial}\left[\frac{1+\varepsilon_e}{\varepsilon_e}(\overline{\mu}-\xi)-1\right]}{\overline{\mu}}+\frac{1-\frac{\sigma}{\sigma-1}\frac{1}{\overline{\mu}}}{\frac{\sigma}{\sigma-1}-\xi}\left[1-\bar{g}\right]\left[1-\overline{H}_{initial}\right]}{1-\left(1-\frac{\mu}{\overline{\mu}}\right)\frac{\xi}{\xi-\frac{\sigma}{\sigma-1}}},$$

which suggests (52).■

D Extensions and Discussions

D.1 Proof of Theorem 2

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

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

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

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

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

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

Substituting $P(\theta_e)\frac{\partial Q_{ij}(\theta_e)}{\partial L_w(\theta_e)}$ by $\frac{W\mu(\theta)}{1-\tau_s(\theta_e)}$ gives

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

where the third equation is derived by (C18).

In particular, if within the interval $(\underline{\theta}^*_e, \overline{\theta}^*_e) \in \Theta^2_e$ tax rates and markups are constant, we have

$\frac{d\ln\left[\mu(\theta_e)\varepsilon_{Q_{-ij}}(\theta_e)\right]}{d\theta_e} = 0$ and the following result (e.g., see Lemma 3):

$$\frac{\pi'(\theta_e)}{\pi(\theta_e)} = \frac{l_e'(\theta_e)}{l_e(\theta_e)} \frac{1+\varepsilon_e}{\varepsilon_e}. \tag{D2}$$

Thus, for any $(\underline{\theta}_e^*, \overline{\theta}_e^*) \in \Theta_e^2$,

$$\frac{1}{1-\tau_s(\theta_e)} = \frac{\widetilde{\mu}}{\mu(\theta)} + [1-\tau_e(\theta_e)]\,\varepsilon_{Q_{-ij}}(\theta_e)\left[[1-g_e(\theta_e)] - [1-\bar{g}_e(\theta_e)]\frac{1-F_\pi(\pi(\theta_e))}{f_\pi(\pi(\theta_e))\,\pi(\theta_e)}\right].$$

(ii) According to $\frac{\partial \pounds}{\partial l_w(\theta_w)}$, we have

$$\frac{1}{\frac{\phi_w'(l_w(\theta_w))}{x_w(\theta_w)}} = \frac{\lambda}{\lambda'}\left[1 - \frac{x_w'(\theta_w)}{x_w(\theta_w)} \frac{\psi_w(\theta_w)}{\lambda N_w f_w(\theta_w)} \frac{1+\varepsilon_w}{\varepsilon_w}\right]. \tag{D3}$$

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

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

(iii) According to $\frac{\partial \pounds}{\partial l_e(\theta_e)}$, we have

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

Using $\frac{\mu(\theta)}{1-\tau_s(\theta_e)} = \widetilde{\mu} - \frac{\kappa'(\theta_e)\xi}{\lambda WL_w(\theta_e) N_e f_e(\theta_e)}$ to substitute $\frac{\kappa'(\theta_e)\xi}{\lambda WL_w(\theta_e) N_e f_e(\theta_e)}$, and $\frac{\xi[1-\tau_s(\theta_e)]}{\mu(\theta)}$ to substitute $\frac{WL_w(\theta_e)}{P(\theta_e)Q_{ij}(\theta_e)}$ we have

$$\begin{aligned}\frac{1}{1-\widetilde{\tau}_e(\theta_e)} \;=\;& 1 + [1-\bar{g}_e(\theta_e)]\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu(\theta_e)\frac{\partial \ln P(Q(\theta_e),\theta_e)}{\partial\theta_e} + \frac{x_e'(\theta_e)}{x_e(\theta_e)}\right] \\ & + \frac{1}{1-\widetilde{\tau}_e(\theta_e)}\left[1 - \widetilde{\mu}\frac{1-\tau_s(\theta_e)}{\mu(\theta)}\right].\end{aligned} \tag{D4}$$

Lastly, substituting $1-\widetilde{\tau}_e(\theta_e)$ by $\frac{[1-\tau_e(\theta_e)][1-\tau_s(\theta_e)]}{\mu(\theta_e)}$, we have

$$\frac{1}{1-\tau_e(\theta_e)} = \frac{1 + [1-\bar{g}_e(\theta_e)]\frac{1-F_e(\theta_e)}{f_e(\theta_e)}\frac{1+\varepsilon_e}{\varepsilon_e}\left[\mu(\theta_e)\frac{\partial \ln P(Q(\theta_e),\theta_e)}{\partial\theta_e} + \frac{x_e'(\theta_e)}{x_e(\theta_e)}\right]}{\widetilde{\mu}}. \tag{D5}$$

E Additional Figures

E.1 Comparison economy under Laissez-faire and optimal tax: ratios

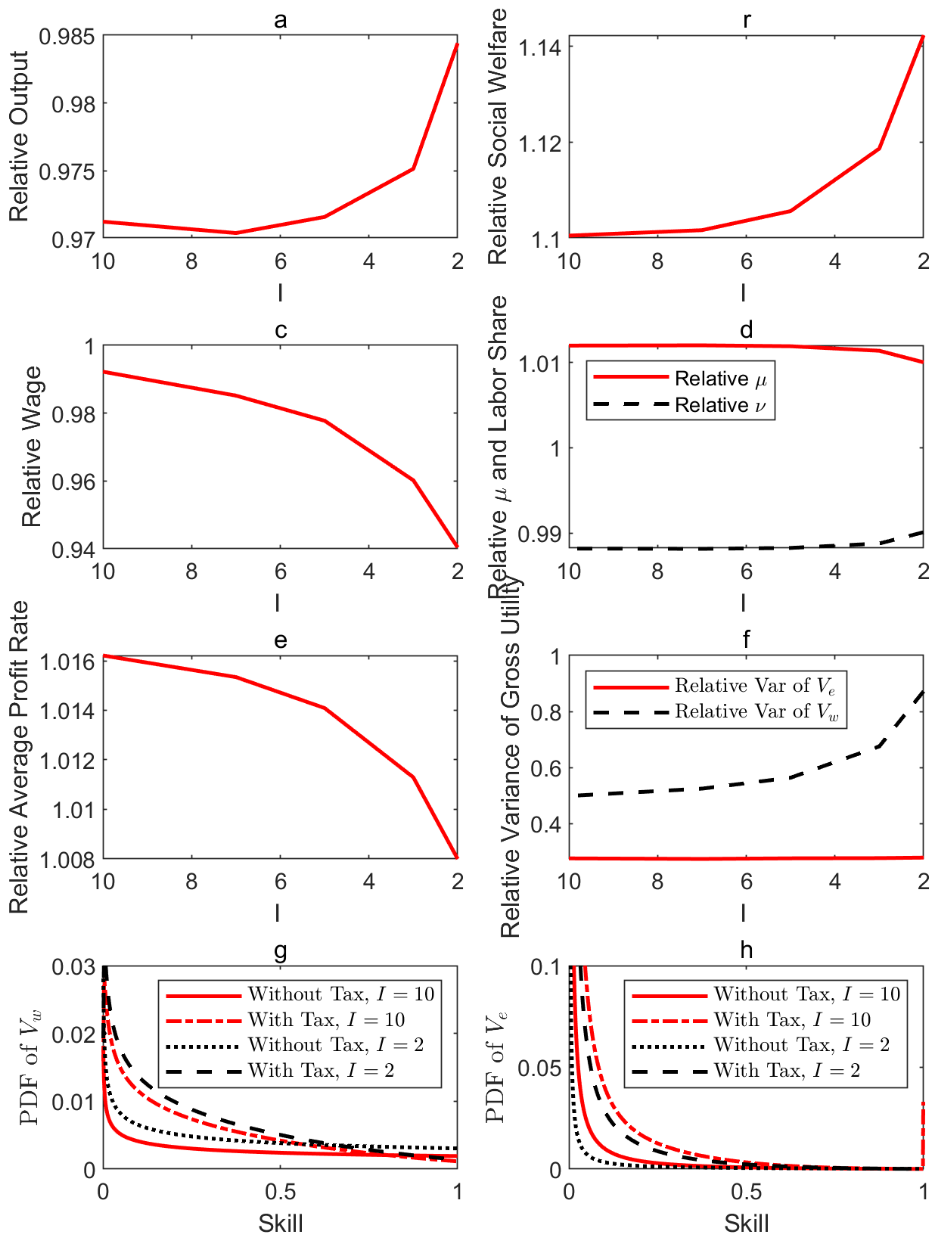


Figure E2: Comparing variables under zero tax and optimal tax

E.2 Optimal tax rates under different distributions of markup.

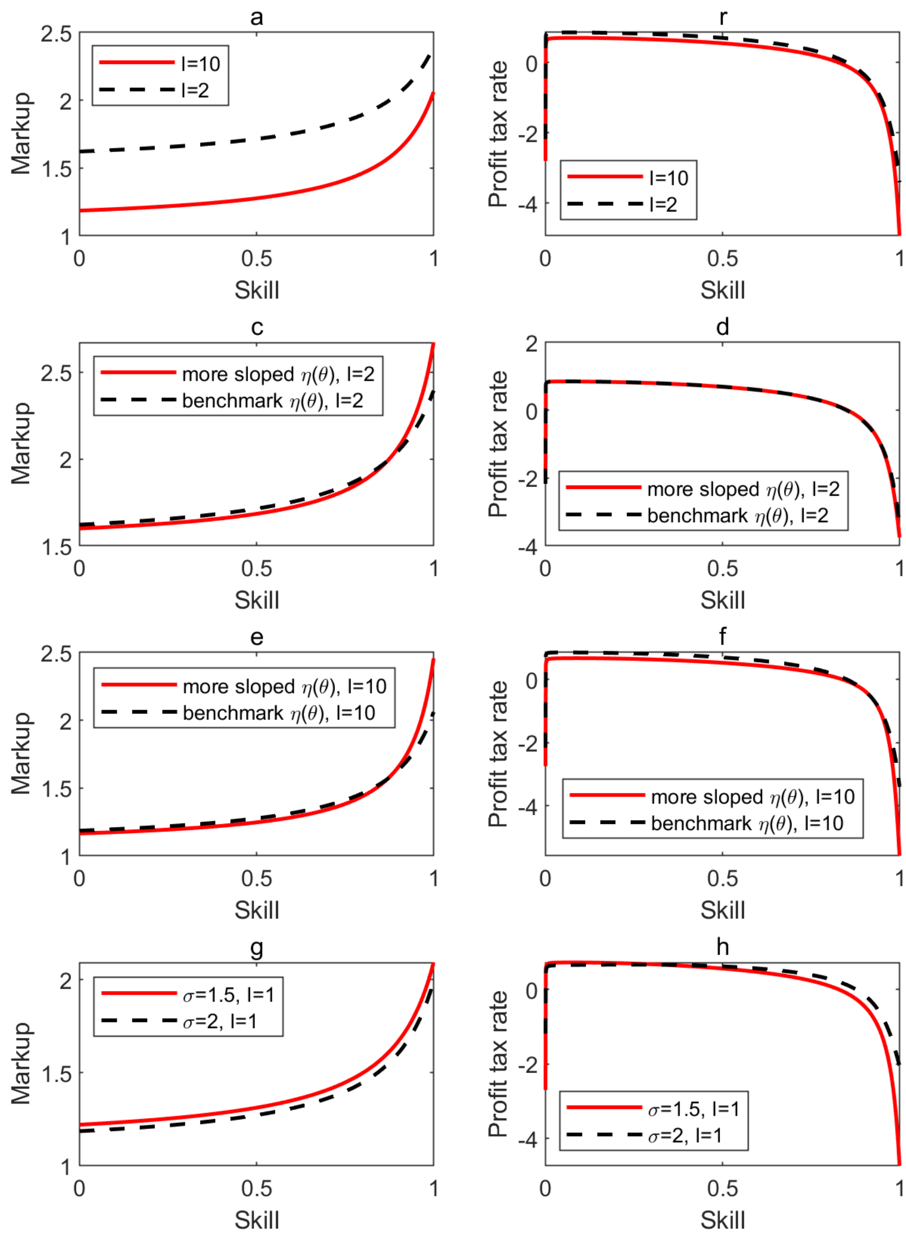


Figure E3: Optimal profit tax rates under different distributions of markup

References

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

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

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

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

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

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

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

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

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

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

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

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

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

CHETTY, R. (2012): "Bounds on Elasticities With Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply," *Econometrica*, 80, 969–1018, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA9043.

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 (2020): "Optimal Supply-side Capital Income Tax with Endogenous Wage Inequality," Working paper.

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.

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.

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

GRUBER, J. AND E. SAEZ (2002): "The elasticity of taxable income: evidence and implications," *Journal of public Economics*, 84, 1–32.

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.

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.

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.

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.

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.