Optimal Firm Regulation with Labor Market Monopsony

Hanvit Kim* Yena Park†

March 20, 2025

Abstract

We study the optimal firm regulation in the presence of imperfect competition in the labor market. More precisely, we investigate how to regulate firms for workers' redistribution in a model where firms have monopsonistic wage setting power due to the workers' idiosyncratic preferences for different firms. Labor market monopsony provides forces for compressing wage schedules across firms—subsidizing more productive firms less, but the constrained planner who cannot provide direct insurance against idiosyncratic preferences of workers should balance the trade-off between redistribution of workers across different firms and allocative efficiency. The size of efficiency cost and thus the optimal amount of redistribution depends on the mechanism of firm premium (rent sharing vs compensating differential). The optimal markdown—the ratio of wage relative to the marginal product of labor—schedule is not necessarily monotone in firm size. The shape of the optimal markdown depends on the sources of sorting between firms and workers. It can be inversely U-shaped if the demand-side sorting dominates.

Keywords: Firm Regulation, Imperfect Labor Market, Monopsony, Redistribution Policy

*University of Chicago, Department of Economics, hanbrian@snu.ac.kr

†Seoul National University, Department of Economics, yena.park@snu.ac.kr

1 Introduction

What is the optimal redistribution policy when firms have wage setting power in the labor market? Standard framework for studying the design of redistribution policy is a competitive labor market where firms take market wages as given. However, there has been accumulating evidence on the role of firms in wage inequality, and the recent empirical studies have an agreement that firms have some wage-setting power. Bonhomme, Holzheu, Lamadon, Manresa, Mogstad, and Setzler (2022) show that firms explain 30% to 50% of wage inequality in developed countries.[1] This firm's effect combines the firm-specific wage premium and sorting of high-skilled workers to productive firms. Many studies have used a model of Abowd, Kramarz, and Margolis (1999) (hereafter AKM) to quantify the role of each effect, and they show that firm premiums cause about 5–20% of wage inequality.[2,3]

With this firm's role in wage inequality, this paper asks the following questions. How should we design optimal firm regulation in an economy with firms' labor market power when we consider its impact on redistribution across workers?[4] What is the key trade-off in setting the policy? How do the trade-off and optimal policy depend on the different mechanisms of firms' effect on wage inequality and sources of sorting? Can the regulation of firms improve redistribution of workers even in the presence of income tax-transfer policy?

[1]This description is based on the within-between decomposition of wage variation by using employer-employee data from the United States and several European countries. Card, Heining, and Kline (2013) also carry out this decomposition to show that between-firm variation explains larger share of rising West Germany wage inequality over time. Similarly, the decomposition of Song, Price, Guvenen, Bloom, and von Wachter (2019) using US tax data shows that two-thirds of the rise in earnings inequality occurred due to a rise in the dispersion between firms, whose role is much larger than within-firm variation.

[2]Abowd, Kramarz, and Margolis (1999) make an identification using movers—the wage change around a move between firms reflects the difference in firm premiums of the two firms. Due to the small number of movers, the firm premium effects tend to be upward biased, and the recent studies show smaller firm premium effects (5–15%) after correcting this bias (e.g. Bonhomme, Lamadon, and Manresa (2019) and Kline, Saggio, and Sølvsten (2020)).

[3]There is also a related literature estimating the impact of exogenous shocks to firm on wages of workers, and studies find wage-productivity elasticities in the range of 0.05–0.15. For example, see Abowd and Lemieux (1993), Van Reenen (1996), and Van Reenen (1996).

[4]There is relatively small literature which studies the normative implications of firm size-wage gap. Most of these studies either focus on its implications on the optimal labor income tax or for policies affecting labor demand exclusively concentrate on the optimal minimum wage.

We answer these questions by embedding a monopsonistic labor market model into an optimal tax framework. We present a simple model with imperfect competition in the labor market which stems from the idiosyncratic preferences of workers across firms, building on the work of Manning (2013), Card, Cardoso, Heining, and Kline (2017) and Lamadon, Mogstad, and Setzler (2022). Workers have heterogeneous preferences over the amenities and work environment (such as firm location and job characteristics) of firms. Due to this preference heterogeneity of workers, firms face upward-sloping labor supply curves and have some wage setting power. With no regulation on firms, wages are marked down relative to the marginal revenue products of labor. Firms with higher productivity enlarge production by posting higher wages, resulting in identically skilled workers working at more productive firms earning more compared to those at other firms. Thus the dispersion of marginal revenue products of labor across firms is transmitted to the wage dispersion. If this inequality is too high from the perspective of a planner, the planner wants to provide redistribution across workers by regulating firms.

We consider a constrained planner who maximizes social welfare—weighted sum of workers' and entrepreneurs' (firm owners') utilities. This planner is constrained so that she cannot provide direct insurance against the idiosyncratic preference shock of the workers and thus cannot directly control the assignment of workers across firms.

We first highlight the key trade-off of the planner and how the optimal balance of the trade-off depends on the mechanisms of the firm's effect. For expositional clarity, consider a simple model with only a single skill type of workers.[5] The constrained planner wants to provide redistribution across workers at different firms, but it is impossible to reduce wage inequality across firms without incurring losses in allocative efficiency as the planner should consider the worker's workplace decision. The workers decide where to work given the redistribution, and the planner's wage compression shifts workers from high- to low-productivity firms. The optimal degree of redistribution is determined by balancing this trade-off. We show that the amount of redistribution is characterized by the negative slope of the markdown schedule across firms' productivities, where the markdown is the ratio of the wage relative to the

[5]The analysis of this simple model is essentially the same with the analysis of skill-dependent policy in an economy with multiple skill types.

marginal revenue product of labor. The planner with stronger redistributive motive reduces wage inequality with wider markdown of wage at higher productive firms.

Importantly, the severity of the trade-off and the degree of optimal redistribution depend not just on the wage inequality caused by the firms but on the mechanisms of the firm's effect. That is, low estimates of the firm's effects in the empirical literature do *not* necessarily imply low demand for the firm regulation. The dispersion of firm premium can be generated by either rent sharing of differently productive firms or compensating differential of heterogeneous amenities across firms. Lamadon, Mogstad, and Setzler (2022) find that the wage variation caused by firms' premiums is small because high productivity firms tend to provide more amenities, which lowers wages productive firms have to pay. Under this mechanism, lower wage inequality calls for more redistribution as the positive correlation between firms' productivity and amenity lowers the efficiency cost of redistribution.

Next, we characterize the optimal redistribution—markdown schedule—with multiple skill types. In the real world, skill-dependent redistribution policy rarely exists. Thus we analyze optimal markdown under the assumption that the planner can only use skill-independent policy. Then the planner who wants to increase the wages of a low-skilled worker at a particular firm can do this by also raising the wages of a high-skilled worker at that firm. We formalize this by introducing the no-discrimination constraint (NDC) to the planner.

Under the NDC, the optimal markdown schedule is shaped by two forces: (i) redistribution of workers across firms and (ii) redistribution between skill types. We have already discussed the wage compression force from the first redistribution channel in the analysis with only single type. On the other hand, in the presence of the complementarity between the firm's productivity and worker's skill, the same (first) redistribution motive generates additional force for the markdown schedule whose direction is not monotone. More dispersed marginal products of labor across firms still calls for more redistribution, but higher productivity-skill complementarity can either increase or decrease average marginal revenue product of a firm depending on the employment share of skill types. For firms with a large (small) share of low-skilled workers, increase in complementarity lowers (increases) the average marginal product, providing a force to subsidize higher productivity firms more (less).

The optimal markdown schedule should also reflect the second redistribution motive between the skill types. The planner who does not have a direct tool for redistribution from the high-skilled to the low-skilled can indirectly achieve this by providing more subsidy to a firm which has a larger labor share of low-skilled workers as the proportion of subsidies transmitted to a particular type of worker is determined by the employment share of that worker type within a firm. If the employment share of high-skilled workers is higher at more productive firms, the redistribution motive between the skill types provides additional force of wage compression—wider markdown of wage at more productive firms.

Our theoretical decomposition implies that the shape of the markdown schedule is not necessarily monotone in a firm's productivity. Thus we quantitatively simulate the optimal markdown schedule across firms. Our numerical example shows that it can be actually inversely U-shaped especially at the low productivity region in the firm's distribution. Both the redistribution against the firm's productivity dispersion and the redistribution between the skill types with positive sorting result in wage compression—wider markdown at productive firms. However, at the low region of firms' productivity distribution, the channel of redistribution across firms through the complementarity dominates, thus the planner subsidizes more productive firms more. Importantly, the shape depends on the sources of sorting. When the positive firm-worker sorting is mainly driven by the complementarity of firm productivity and worker skill, the complementarity channel becomes stronger, making the inverse U-shape of the markdown schedule more pronounced. On the other hand, if workers' heterogeneous taste is the main source of the positive sorting, the markdown schedule becomes monotone.

Our results are robust to the extension with more direct redistributive income tax policy. As long as income tax does not depend on the employer of the worker, regulating firm's markdown with the subsidy to the firms still plays a role for the redistribution of workers across different firms. In the presence of progressive income tax-transfer, the amount of redistribution through the firm's regulation becomes quantitatively smaller because (i) there is direct substitution by the income tax, and (ii) progressive income tax-transfer schedule can relax the no-discrimination constraint of the planner. Finally, we also extend the analysis to the economy with nonzero labor market share and to the economy with product market power to show the robustness of our results.

Literature Studies on optimal redistribution policy for workers mostly take a competitive labor market framework and thus abstract from the firm's role in wage inequality. In such a setting, labor income tax-transfer policy is the dominant tool for redistribution. (e.g. Heathcote, Storesletten, and Violante (2017)). This paper departs from the competitive labor market and study optimal redistribution policy in the presence of monopsony power in the labor market. Closely related studies are Berger, Herkenhoff, Mongey, and Mousavi (2024) and Berger, Herkenhoff, and Mongey, Simon Oppenheimer (2024). They take Ramsey style normative analysis in an economy where firms have labor market power, restricting their attention to the income tax function which is the same as in Heathcote, Storesletten, and Violante (2017).[6] Our paper investigates the constrained planner's problem with no restrictions on the policy function and studies the implications for the firm regulation.

Our paper is connected to the literature on firm's wage setting power. Brown and Medoff (1989) and Oi and Idson (1999) summarizes the literature documenting the existence of firm size-wage gap. Manning (2013) argued that imperfect competition in the labor market is the underlying reason for identically skilled workers earning different wages across firms. Using administrative matched employer-employee datasets, Card, Heining, and Kline (2013) and Song, Price, Guvenen, Bloom, and von Wachter (2019) show that both a rise in the dispersion of firm pay premiums, as well as an increased sorting of high-skilled workers on high-wages firms, are quantitatively important in characterizing the role of firms in rising wage inequality. Recently Card, Cardoso, Heining, and Kline (2017) and Lamadon, Mogstad, and Setzler (2022) suggested workers' idiosyncratic preferences across firms as a potential source of firm's wage setting power, the setup which our baseline model closely follows. Berger, Herkenhoff, and Mongey (2022a) proceed further to develop an oligopsony model which allows the firm's strategic interaction.

Our paper is also connected to the relatively small literature studying the normative implications of imperfect competition on optimal policy. Cahuc and Laroque (2014) and da Costa and Maestri (2019) investigates how optimal labor income taxes are affected by labor mar-

[6]Berger, Herkenhoff, Mongey, and Mousavi (2024) show the the amplified distortions of progressive taxation and Berger, Herkenhoff, and Mongey, Simon Oppenheimer (2024) study how labor market power affects optimal income tax policy.

ket monopsony. Hurst, Kehoe, Pastorino, and Winberry (2022), Ahlfeldt, Roth, and Seidel (2022) and Berger, Herkenhoff, and Mongey (2022b) analyze the optimal minimum wage when firms have wage setting power. Edmond, Midrigan, and Xu (2022), Boar and Midrigan (2019) studies the impact of product market concentration on optimal firm size regulation. Finally, Ales and Sleet (2022) studies optimal taxation under discrete choice environments.

Layout The remainder of the paper is organized as follows. In Section 2, we introduce our baseline model. In Section 3, we highlight the key trade-off using the simple model. Section 4 characterizes optimal redistribution policy in the full model, and Section 5 provides a quantitative analysis. Section 6 shows the robustness of our results in the presence of labor income tax-transfer, and Section 7 provides extension analysis. Section 8 concludes.

2 Model—Economy without Regulation

In this section we present the environment and define the equilibrium under labor market monopsony with no redistribution policy. The description of the model closely follows Lamadon, Mogstad, and Setzler (2022).

2.1 Environment

An economy is populated by J firms and two types of workers, high- and low-skilled. Each firm $j \in \{1, \ldots, J\}$ posts skill-specific wages $(w_{hj}, w_{\ell j})$ and provides amenities. We denote the value of the amenities that workers of skill type s are expected to get from firm j by g_{sj}. After observing the distribution of wages and amenities across firms, each worker chooses which firm to work for. The indirect utility of a worker with skill type $s \in \{h, \ell\}$ when working at firm j is given by

$$v_s(\varepsilon, j) = \log w_{sj} + \log g_{sj} + \varepsilon_j \tag{1}$$

where ε_j is the idiosyncratic preference of a worker for working at firm j. Working at a particular firm is valued differently by different workers. Thus a firm doesn't lose all of its employees when it lowers the wage, which gives rise to firm's wage setting power. We assume

that each worker's $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_J) \in \mathbb{R}^J$ is independently drawn from a Gumbel distribution with zero location and scale $1/\beta$. We denote joint distribution of idiosyncratic preference shock across workers as $F(\varepsilon)$ and its density as $f(\varepsilon)$.

Each worker selects a firm to work at to maximize his utility, taking the distribution of skill-specific wages and amenities across firms as given. A s-skill type worker with preference shock ε solves

$$j_s(\varepsilon) = argmax_k \quad v_s(\varepsilon, k). \tag{2}$$

Then the probability of worker with type s working at firm j is

$$P_{sj} = \mathbb{P}_\varepsilon[j = argmax_k \;\; v_s(\varepsilon, k)] = \frac{(w_{sj} g_{sj})^\beta}{I_s}, \quad I_s = \sum_k (w_{sk} g_{sk})^\beta,$$

where I_s is wage index for the s-type workers. The number of s-type workers at firm j is

$$L_{sj} = N_s P_{sj} = N_s \frac{(w_{sj} g_{sj})^\beta}{I_s},$$

where N_s is the mass of workers with skill level s. The higher the wage and amenity level that a firm provides to type s workers, the larger the employment a firm can utilitize to produce. The labor supply elasticity with respect to wages is β, which is the inverse of the scale for idiosyncratic preference distribution. Lower β implies that workers' preferences across firms are more dispersed, thus workers respond less to changes in wages.

All firms produce identical consumption good which we take as numeraire. Firms are heterogeneous with respect to their production technologies. The production function of firm j takes the following form:

$$y_j = z_j (\alpha_j L_{hj}^{\frac{\sigma-1}{\sigma}} + (1 - \alpha_j) L_{\ell j}^{\frac{\sigma-1}{\sigma}})^{\frac{\sigma}{\sigma-1}} = z_j \tilde{y}_j,$$

where y_j is output, z_j is productivity, α_j is skill intensity, $\tilde{y}_j = (\alpha_j L_{hj}^{\frac{\sigma-1}{\sigma}} + (1 - \alpha_j) L_{\ell j}^{\frac{\sigma-1}{\sigma}})^{\frac{\sigma}{\sigma-1}}$ is employment index, and σ is elasticity of substitution between different types of workers.

Each firm owner employs workers, produces output, provides skill-specific wages and amenities, and consumes remaining profit. We assume that firms cannot observe each worker's idiosyncratic preferences across firms. Thus firms pay same wages to its employees with

identical skills as they cannot discriminate prices across workers with different idiosyncratic preferences. An owner of firm j solves the follwing problem

$$\begin{aligned} \pi_j = max_{w_{hj},w_{\ell j},L_{hj},L_{\ell j}} \quad & z_j(\alpha_j L_{hj}^{\frac{\sigma-1}{\sigma}} + (1-\alpha_j)L_{\ell j}^{\frac{\sigma-1}{\sigma}})^{\frac{\sigma}{\sigma-1}} - (w_{hj}L_{hj} + w_{\ell j}L_{\ell j}) \\ s.t. \quad & L_{sj} = L_{sj}^f(w_{sj}) = N_s \frac{(w_{sj}g_{sj})^\beta}{I_s}, \quad s = h, \ell, \end{aligned} \tag{3}$$

where π_j is profit which the firm owner consumes, and $L_{hj}^f(w_{hj})$ and $L_{\ell j}^f(w_{\ell j})$ are upward-sloping labor supply curves faced by the firm j. We assume that there is no strategic interaction. Each firm is small—labor market share of each firm is near zero—thus firms ignore the impact of their wage choice on the wage index I_s when they set wages. We relax this in Section 7.1.

Solving the firm's problem (3) yields firm j's optimal wage choice

$$w_{hj} = \frac{\beta}{1+\beta} MR_{hj}, \quad w_{ellj} = \frac{\beta}{1+\beta} MR_{\ell j} \tag{4}$$

where $MR_{hj} = z_j \alpha_j \left(\frac{L_{hj}}{\tilde{y}_j}\right)^{-\frac{1}{\sigma}}$ and $MR_{\ell j} = z_j(1-\alpha_j)\left(\frac{L_{\ell j}}{\tilde{y}_j}\right)^{-\frac{1}{\sigma}}$ are marginal revenue products of high- and low-skilled workers at firm j. The higher the labor supply elasticity β, the less the wages of each type of worker are markdowned from the marginal revenue products.

2.2 Equilibrium without Regulation

We define the equilibrium in our environment.

Definition 1. *A monopsonistic competitive equilibrium is defined as a set of worker decisions $j_s(\varepsilon)$, wage indices $\{I_s\}$, labor demand $\{L_{sj}\}$, firm specific labor supply curves $\{L_{sj}^f(w_{sj})\}$ and wages $\{w_{sj}\}$ such that*

1. *Each worker's decision of firm $j_s(\varepsilon)$ maximizes his utility, defined in* (1)
2. *Firms facing upward sloping labor supply curves $\{L_{sj}^f(w_{sj})\}$ choose labor demand $\{L_{sj}\}$ by posting wages $\{w_{sj}\}$ to maximize profits, as described in* (3)
3. *Wage indices $\{I_s\}$ with $I_s = \sum_j (w_{sj}g_{sj})^\beta$ are generated from workers' decision $j_s(\varepsilon)$.*

Empirical evidences document the existence of firm size-wage gap, i.e. large firms pay higher wages. One explanation of the firm's premium is that high-skilled workers are systematically sorted into large firms with high productivity. An alternative view is that identically skilled workers, who are perfectly replacable in terms of the job they are performing, are paid different wages depending on which firm they work for. Our model incorporates both of these explanations.

In our environment, firm productivity heterogeneity gives rise to across-firm wage inequality for identically skilled workers. High-productivity firms have higher marginal revenue product of labor than low-productivity firms holding level of employment constant, which is transmitted to wages as in (4). Thus workers employed at high-productivity firms earn more.

Our model also permits systematic sorting of worker skills across firms. Following lemma explicitly shows the two forces of sorting (see Appendix A for the proof).

Lemma 2. *Under monopsonistic competitive equilibrium* $\frac{d\log\left(\frac{L_{hj}}{L_{lj}}\right)}{dz_j} = \frac{\sigma\beta}{\sigma+\beta}\frac{d\log\left(\frac{\alpha_j}{1-\alpha_j}\right)}{dz_j} + \frac{\sigma\beta}{\sigma+\beta}\frac{d\log\left(\frac{g_{hj}}{g_{lj}}\right)}{dz_j}$.

First, sorting may arise due to the skill intensity heterogeneity. In the presence of complementarity between firm productivity and worker skill ($\frac{d\alpha_j}{dz_j} > 0$), high-productivity firms have a larger demand for high-skilled workers relative to less productive firms. Second, the fact that differently skilled workers value amenities differently is another potential source of sorting. If firms with higher productivity provide more amenities, then the high skilled workers sort into more productive firms ($\frac{d\left(\frac{g_{hj}}{g_{lj}}\right)}{dz_j} > 0$).

3 Simple Model: Redistribution Policy within Skill Type

From now on, we study the optimal firm regulation from the perspective of the planner with redistributive concerns. In this section, we begin with the optimal regulation analysis in the simple model where there is only one type of worker, which allows us to understand the key equity-efficiency trade-off in a monopsonistic labor market.[7]

[7]The analysis is essentially the same with the case with multiple types if we allow the planner to use a type-dependent policy.

We simplify the environment of the benchmark model in this section. There is only one type of worker with mass N. The indirect utility of worker working at firm j is $v(\varepsilon, j) = \log w_j + \log g_j + \varepsilon_j$, where w_j is wage of workers at a firm j and g_j is amenties which firm j provides to workers. From the worker's optimization, we derive the labor supply curve of a firm j $L_j = N\frac{(w_j g_j)^\beta}{I}$, where $I = \sum_k (w_k g_k)^\beta$ is wage index.

Firm j produce consumption goods using a production technology $y_j = z_j L_j^\eta$. Throughout this section, the default assumption is linear production with $\eta = 1$. When we extend to a nonlinear production, we will explicitly mention it. In equilibrium without regulation, firms choose wages and employment to maximize profits subject to the labor supply constraint $L^f(w_j; g_j) = N\frac{(w_j g_j)^\beta}{I}$, taking wage index I as given. The equilibrium wage of firm j is markdown of marginal product z_j: $w_j = \frac{\beta}{1+\beta} z_j$, thus the dispersion of marginal product across firms is transmitted to the wage inequality with proportion $\frac{\beta}{1+\beta}$.

3.1 Constrained Planner's Problem

We first analyze the optimal allocation by solving the constrained planner's problem, then we show that the decentralization can be achieved with revenue-dependent subsidies to firms. The welfare function is the weighted sum of workers' and firm owners (entrepreneurs)' utilities. Denote the utility of a worker with preference draw ε by $V(\varepsilon; \{c_j\}) = max_j \log c_j + \log g_j + \varepsilon_j$ given the distribution of consumption allocation $\{c_j\}$ across firms chosen by the planner. Note that the function V takes into account workers' optimal choice of firms. The social welfare function is given by

$$\mathscr{W} = N \int_\varepsilon V(\varepsilon; \{c_j\}) + \psi_f \left(\sum_j \log c_{fj} \right)$$

where ψ_f is welfare weight on entrepreneurs relative to workers and c_{fj} is the consumption allocation of the entrepreneurs working at firm j. We analyze the optimal regulation from the perspective of the planner with redistributive concerns, not only between workers and firm owners but also across workers.

Our planner is constrained in a sense that given the planner's choice of consumption allocation, she should respect the optimal workplace decision of workers. When calculating

welfare, the planner takes into account that once she has set consumption distribution $\{c_j\}$, each worker chooses where to work to maximize utility. Thus the employment distribution across firms should also be consistent with the workers' optimal decisions, and the constrained planner faces the same upward sloping labor supply curve as a firm does.

We define our social planner's problem.

Definition 3. *The social planner's problem with one type of labor is defined as the program*

$$\begin{aligned} max_{\{c_j,c_{fj}\},L_j,I} \quad & N\int_{\varepsilon} V(\varepsilon;\{c_j\}) + \psi_f \left(\sum_j \log c_{fj}\right) \\ s.t. \quad (\lambda) \quad & \sum_j (c_j L_j + c_{fj}) \leq \sum_j z_j L_j \qquad (5) \\ (\nu) \quad & I = \sum_j (c_j g_j)^{\beta} \qquad (6) \\ & L_j = N\frac{(c_j g_j)^{\beta}}{I}, \quad \forall j, \qquad (7) \end{aligned}$$

where (5) *is resource constraint,* (6) *is formula for wage index, and* (7) *is employment distribution across firms as function of* $\{c_j\}$.

3.2 Optimal Allocation

By solving the planner's problem in Definition 3, we obtain the optimal consumption allocation. Next proposition characterizes the optimal consumption schedule (see Appendix A for the proof).

Proposition 4. *The optimal consumption allocation of workers at firm* j, c_j, *is given by*

$$c_j = \frac{\beta}{1+\beta} \times z_j + \Delta, \quad \Delta = \frac{1}{\lambda(1+\beta)}\left(1 - \beta\nu\frac{I}{N}\right) \qquad (8)$$

where λ *is Lagrange multiplier on the resource constraint* (5), *and* ν *is Lagrange multiplier on the equation for wage index* (6). *Lagrange multipliers satisfies* $\frac{1}{\lambda} = \frac{c_f}{\psi_f} = \frac{\sum_j c_j L_j}{N} = \frac{1}{N+\psi_f J}\sum_j y_j$ *and* $\nu\frac{I}{N} = \frac{\psi_f J}{N}$. *There exists a cutoff Pareto weight* $\bar{\psi}_f$ *such that for* $\psi_f < \bar{\psi}_f$, $1 - \beta\nu\frac{I}{N} > 0$ *and thus* $\Delta > 0$.

Equation (8) characterizes the optimal consumption allocation for workers working at firm j, which is the sum of a fraction of marginal product of labor $\frac{\beta}{1+\beta} z_j$ and the common transfer term $\mathbf{\Delta}$. The size of $\mathbf{\Delta}$ measures both (i) the degree of redistribution from the entrepreneurs toward workers and (ii) the degree of redistribution across workers at different firms. The first redistribution motive is determined by the size of Pareto weight on entrepreneurs ψ_f—Lower ψ_f implies more redistribution toward workers. We focus on the case with $\psi_f < \bar{\psi}_f$ so that $\mathbf{\Delta}$ is positive. The second redistribution motive is achieved as the consumption inequality of workers working at different firms becomes smaller by the common transfer term.

In the monopsonistic labor market, the key trade-off the planner faces is the trade-off between redistribution across workers working at different firms and allocative efficiency. The workers decide where to work given the redistribution provided by the planner. The constrained planner is subject to the constraint that once the consumption allocation has been set, she should respect the resulting optimal workplace decisions of workers and the corresponding employment distribution across firms. Then to close the consumption gap between workers working at different firms, the constrained planner should necessarily shift the employment from high- to low-productivity firms, which reduces the overall level of production. The optimal allocation is determined to optimally balance this trade-off.

Allocative Inefficiency in comparison with the First-Best Case

We can better understand the trade-off of our planner by comparing out planner to an alternative First-Best planner who can independently choose consumption and employment distribution of workers across firms, as in Mongey and Waugh (2024). For simplicity, we assume $g_j = 1, \forall j$. The First-Best planner solves

$$\begin{aligned} max_{\{x_j(\varepsilon)\},\{c_j(\varepsilon),c_{fj}\}} \quad & N \int_\varepsilon \left(\sum_j x_j(\varepsilon)(\log c_j(\varepsilon) + \varepsilon_j) \right) dF(\varepsilon) + \psi_f \sum_j \log c_{fj} \\ s.t. \quad (\lambda) \quad & N \int_\varepsilon \left(\sum_j x_j(\varepsilon) c_j(\varepsilon) \right) dF(\varepsilon) + \sum_j c_{fj} \leq \sum_j z_j \left(N \int_\varepsilon x_j(\varepsilon) dF(\varepsilon) \right) \end{aligned}$$

where $x_j(\varepsilon)$ is an indicator function mapping a worker with preference draw ε into 1 if the worker is assinged to firm j and zero otherwise, $c_j(\varepsilon)$ is consumption of the worker who is assigned to firm j, and c_{fj} is consumption of owners of firm j. Note that the key difference

from the constrained planner's problem in Definition 3 is that the First-Best planner can assign workers without considering workers' decision of workplaces.

Optimal $x_j(\varepsilon)$ is given by

$$x_j^s(\varepsilon) = \begin{cases} 1 & \text{if} \quad \log c_j(\varepsilon) + \varepsilon_j + \lambda(z_j - c_j(\varepsilon)) \geq max_k[\log c_k(\varepsilon) + \varepsilon_k + \lambda(z_k - c_k(\varepsilon))] \\ 0 & \text{otherwise} \end{cases}$$

where λ is lagrangian multiplier on the resource constraint. For consumption of workers, we have $\frac{1}{\lambda} = c_j(\varepsilon)$, i.e. consumption level is equal across all workers regardless of the firm that each worker is assigned to. Thus the formula for optimal $x_j(\varepsilon)$ simplifies to

$$x_j^s(\varepsilon) = \begin{cases} 1 & \text{if} \quad \lambda z_j + \varepsilon_j \geq max_k[\lambda z_k + \varepsilon_k] \\ 0 & \text{otherwise.} \end{cases}$$

Without any cost of redistribution, the First-Best planner can provide full insurance for workers against idiosyncratic preference shock. Then allocating worekrs is determined to achieve two goals: (i) enhancing production efficiency and (ii) allocating workers with heterogeneous preferences to their more preferred workplaces. On the other hand, our constrained planner reduces the uninsurable component of worker's earnings by compressing the wage distribution across firms. This inevitably harms the allocative efficiency of workers to firms.

Determinants of the Optimal Balance between the Trade-off

We investigate how the optimal balance between the equity and efficiency changes with key parameters of monopsony labor market. First important factor is the labor supply elasticity β. Higher β implies that workers are more responsive to the wage compensation, which implies that redistribution is more costly. When the planner tries to compress wage schedule, workers are more likely to move to lower productivity firms under higher β. Thus, the efficiency cost of additional redistribution goes up, and the amount of redistribution decreases. As a result, consumption inequality increases. By measuring the consumption inequality with its variance, we can directly see that higher β increases the optimal consumption inequality:

$$var(c_j) = \frac{\beta^2}{(1+\beta)^2} var(z_j).$$

The following lemma also shows that when β increases, the planner has more concerns about the efficiency in the optimal balance of the trade-off (see Appendix A for the proof).

Lemma 5. *Assume $J = 2$ with $z_1 < z_2$ and $g_1 = g_2$. Then the employment level at each firm under the optimal allocation is given by*

$$L_1 = N\frac{\gamma_c^\beta}{\gamma_c^\beta + 1}, \quad L_2 = N\frac{1}{\gamma_c^\beta + 1}, \text{ where } \quad \gamma_c^\beta \equiv \frac{c_2}{c_1} \approx \frac{\beta}{1+\beta} \times \frac{N + J\psi_f}{N} \times (\gamma_z - 1) + 1$$

and $\gamma_z = \frac{z_2}{z_1}$. Then $\frac{d\gamma_c^\beta}{d\beta} > 0$, $\frac{dL_2/L_1}{d\beta} > 0$ and $\frac{d\Delta}{d\beta} < 0$.

Under higher β, the planner provides less redistribution across workers by lowering Δ due to the higher efficiency cost given redistribution. Lemma 5 also shows that under higher β, the optimal employment distribution is more skewed towards high-productivity firms, increasing allocative efficiency.

Lastly, we show that with a nonlinear and concave production $y_j = z_j L_j^\eta$, $\eta < 1$, the equity-efficiency trade-off becomes severe. With a nonlinear production function, optimal consumption allocation of workers at firm j is given by

$$c_j = \frac{\beta}{1+\beta} MR_j + \Delta, \text{ where } MR_j = \eta z_j L_j^{\eta-1}.$$

Then the consumption inequality across firms is determined by the dispersion of $\frac{\beta}{1+\beta}$ fraction of marginal products of labor. Providing additional redistribution by compressing wage inequality increases the fraction of workers working at lower productivity firms. Then the marginal product of lower productivity firms decreases, while the marginal product of labor increases for higher productivity firms. This leads to an additional dispersion of marginal products of labor and reduction in the optimal redistribution.

3.3 Optimal Markdown and Subsidy Characterization

We now characterize the optimal markdown schedule which determines the schedule of the policy implementing the constrained planner's solution. Following the convention in the literature, we define the markdown of a firm j by

$$\mu_j \equiv \frac{w_j}{MR_j}, \tag{9}$$

which is the ratio of wage relative to the marginal product of labor. Note that lower μ_j implies wider markdown of wage relative to MR_j. Since $c_j = w_j$ without income tax, the markdown is also expressed as $\mu_j = \frac{c_j}{MR_j}$.

We then show that the markdown schedule essentially shapes the subsidy schedule which implements the optimal allocation. Decentralization of the optimal allocation can be achieved by the revenue-dependent subsidies to firms. In the benchmark analysis, we focus on the decentralization with only firm regulations only, but in Section 6, we show that the property of optimal firm regulation still applies to the decentralization with income tax-transfer policy.

Suppose that the government sets subsidy schedule $\{S_j(y_j)\}$ which satisfies the government's budget constraint: $\sum_j S_j(y_j) = 0$. Then $\{S_j\}$, although named as subsidies, are actually taxes for some firms. Firm j solves the following problem given S_j:

$$\begin{aligned} \pi_j = max_{y_j, w_j, L_j} \quad & y_j + S_j(y_j) - w_j L_j \\ s.t. \quad & y_j = z_j L_j, \quad L_j = L^f(w_j; g_j) = N \frac{(w_j g_j)^\beta}{I}. \end{aligned}$$

The optimal wage choice of firm j is $w_j = (1 + S_j'(y_j))\frac{\beta}{1+\beta} z_j$, thus the marginal subsidy rate and markdown has the following relationship:

$$\frac{\beta}{1+\beta}(1 + S_j'(y_j)) = \mu_j,$$

which shows that wider markdown of wage relative to the marginal revenue product of labor—lower μ_j—corresponds to lower subsidy rate S_j' in our decentralization.

We then characterize the optimal markdown and subsidy schedule together.

Proposition 6. *We assume $\psi_f < \bar{\psi}_f$. The optimal markdown is given by*

$$\mu_j = \frac{\beta}{1+\beta} + \frac{\Delta}{z_j}, \quad \textit{where } \Delta = \frac{1}{\lambda(1+\beta)} \left(1 - \beta\nu \frac{I}{N}\right). \tag{10}$$

Then the slope of the optimal markdown schedule is given by

$$\frac{d\mu_j}{dz_j} = -\frac{\Delta}{z_j^2} < 0. \tag{11}$$

Thus the marginal subsidy rate $S_j'(y_j)$ which decentralize the optimal allocation decreases in firm's productivity z_j.

The planner achieves redistribution by setting wider markdown—lower μ_j—at more productive firms. That is, the planner subsidizes low-productivity firms at higher rates, encouraging such firms to pay higher wages and employ more workers. When the planner cares more about workers' utility, Δ goes up. Then the planner has stronger motive to redistribute across workers, thus increases the gap of subsidy rates between high- and low-productivity firms.

3.4 Inequality Mechanism and Optimal Trade-off

Recent studies following the AKM models have tried to decompose the sources of wage inequality. Lamadon, Mogstad, and Setzler (2022) have investigated whether the AKM estimates of wage inequality across firms are driven by firm's rent sharing with workers or compensating differentials due to the amenity differences, and they highlighted that the size of the firm premium crucially depends on the mechanisms of inequality. We now show that these mechanisms are also important for the optimal redistribution. Although the redistribution motive of the planner arises due to the wage variation caused by the firms, larger wage inequality across firms does not necessarily call for more redistribution of the planner depending on the sources of the firm's premium.

Our model also allows two potential reasons why identically skilled workers working at different firms earn different wages. First, when firms set wages with monopsony power, workers working at high-productivity firms are paid higher wages than those of the workers working at low-productivity firms. Second, workers working at firms which provide lower level of amenities g_j earn higher wages. Firms which provide bad working conditions must pay compensating differential to attract workers. Empirical analysis in Lamadon, Mogstad, and Setzler (2022) show that although there is a substantial variance of TFP and amenities across firms, the estimated firm effects on wage inequality are relatively small because there is a positive correlation between firm productivity and amenities (i.e. more productive firms provide better amenities to workers $g_j'(z_j) > 0$).[8]

[8]Lamadon, Mogstad, and Setzler (2022) decompose the variance of firm effects identified from the AKM regression into three components : (i) variance of total factor productivity, (ii) variance of amenities provision and (iii) covariance between amenities and productivity.

The presence of positive correlation between firm's productivity and provision of amenities reduces the total wage variation caused by firms. Given this observation, one might conjecture that lower wage inequality would lead to less wage compression (and less redistribution) of the planner. Next proposition investigates the role of the positive correlation and show that the result is the opposite to the conjecture (see Appendix A for the proof).

Proposition 7. *Assume $\psi_f = 0$. Starting from the economy with $g_j'(z_j) = 0$, we increase $\frac{g_j'(z_j)z_j}{g_j}$. Then both the common transfer Δ and $\frac{dL_j}{dz_j}\frac{z_j}{L_j}$ increase in the optimal allocation.*

Proposition 7 shows that when the positive correlation between firm productivity and amenity ($\frac{g_j'(z_j)z_j}{g_j} > 0$) is introduced, the planner provides more redistribution across workers at different firms with larger common transfer to the workers Δ. Further redistribution can be achieved as the equity-efficiency trade-off is relaxed. Note that Proposition 7 also shows that the employment distribution of workers is more skewed towards high-productivity firms (i.e. larger $\frac{dL_j}{dz_j}\frac{z_j}{L_j}$), which reflects the improved allocative efficiency of workers to firms. If more productive firms provide better amenities to workers, then firms with higher productivity employ larger workforce given the same redistribution provided by the planner. As the planner can achieve the same level of redistribution at lower efficiency costs, she provides more redistribution of consumption across workers working at different firms.

One important lesson from this investigation is that what is important for determining the optimal amount of redistribution is not simply the degree of wage variation caused by firms. The mechanisms of the wage variation matter for the efficiency cost of redistribution and thus for the optimal amount of redistribution.

Numerical Example

Although Proposition 7 shows the role of inequality mechanisms in optimal redistribution, it only focuses the local analysis near the zero correlation between productivity and amenity. We now provide numerical example to investigate the role of inequality mechanisms more realistically. We set labor supply elasticity $\beta = 4.50$ and returns to scale $\eta = 0.75$.[9] We

[9]In this example, β is set lower than β in the baseline quantitative model because with only single type of workers, it is hard to generate wage variation in the data.

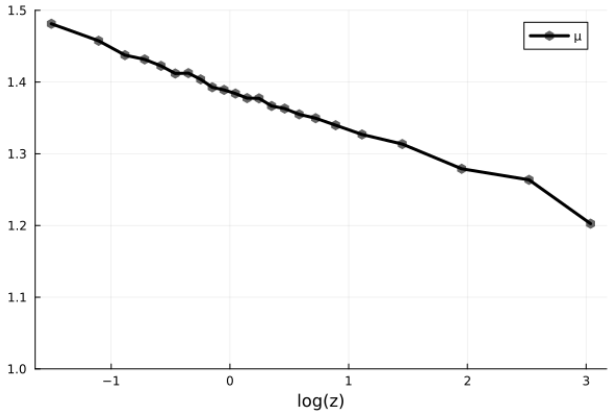


Figure 1: Optimal markdown

assume that firm productivity z_j and amenity g_j follow bivariate lognormal distribution

$$\begin{pmatrix} \log z_j \\ \log g_j \end{pmatrix} \sim N(0, \Sigma), \quad \text{where} \quad \Sigma = \begin{pmatrix} \sigma_z^2 & \rho_{zg}\sigma_z\sigma_g \\ \rho_{zg}\sigma_z\sigma_g & \sigma_g^2 \end{pmatrix}.$$

In an economy with only single type of worker, the log wage in an equilibrium can be additively decomposed into the firm's premium caused by productivity ψ_{zj} and the premium due to amenity ψ_{gj}.[10] Then the variance of the log wage ψ_j can be decomposed into

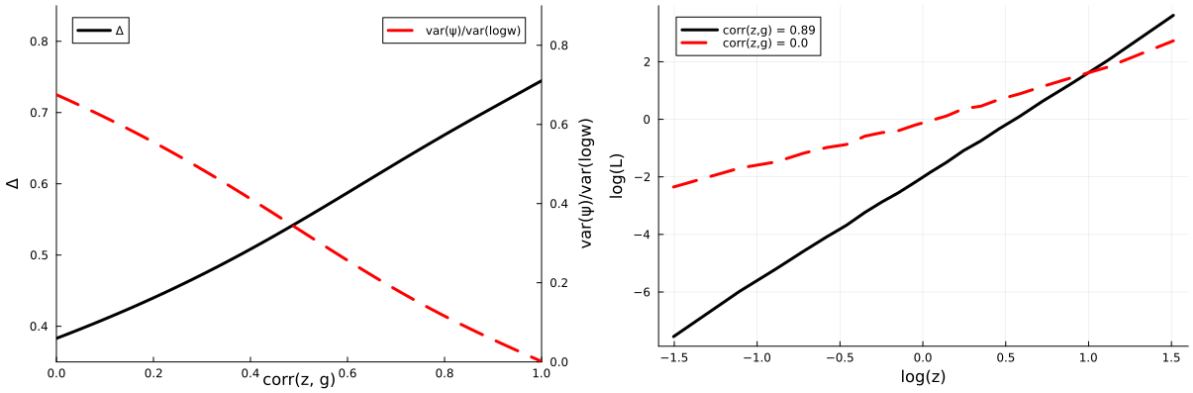
$$Var(\psi_j) = Var(\psi_{zj}) + Var(\psi_{gj}) + 2Cov(\psi_{zj}, \psi_{gj}).$$

We calibrate $\sigma_z = 0.76$, $\sigma_g = 0.64$ and $\rho_{zg} = 0.89$ to match the variance decomposition of the firm pay premiums in Lamadon, Mogstad, and Setzler (2022) ($Var(\psi_{zj}) = 0.17$, $Var(\psi_{gj}) = 0.13$ and $2Cov(\psi_{zj}, \psi_{gj}) = -0.27$). Note that the resulting parameter $\rho_{zg} = 0.89$ implies a substantially high level of correlation between firm productivity and amenities. Figure 1 shows the optimal μ-schedule which is downward sloping in firm productivity. The

[10]The equilibrium log wage in the model has the following decomposition

$$\log w_j = \underbrace{\underbrace{\frac{1}{1+\beta(1-\eta)}\log z_j}_{\psi_{zj}} + \underbrace{\log\frac{\beta\eta}{1+\beta} - (1-\eta)\log H_j}_{\psi_{gj}}}_{\psi_j}$$

where H_j is a term which positively depends on the firm's amenity terms g_j but is independent of firm's productivity z_j. (See appendix for the detail.)



(a) (b)

Figure 2: Sources of sorting

planner assigns wider markdowns of wage relative to the marginal products—lower μ_j—to more productive firms. This optimal markdown schedule compresses wage inequality providing redistribution among identically skilled workers working at different firms.

To investigate the importance of sources of inequality in determining the optimal markdown schedule, we conduct the following exercise. Given fixed σ_z and σ_g, we analyze how the planner's optimal policy changes as the correlation parameter ρ_{zg} varies. Figure 2a shows the result. For each value of ρ_{zg}, the red dashed line plots the variance of firm pay premium $Var(\psi_j)$ as a fraction of the total wage variance $Var(\log w) = 0.44$ observed in the data. As the correlation ρ_{zg} increases, the firm's premium plays a smaller role in explaining the overall wage inequality, but the planner achieves *more* redistribution by providing higher lump sum transfers $\mathbf{\Delta}$, as shown in the black solid line.

The economic intuition behind this result is illustrated in Figure 2b which plots the employment distribution across firms at the optimal allocation for the baseline case with $\rho_{zg} = 0.89$ and the zero-correlation case. Observe that when there is a positive correlation between productivity and amenity, the number of workers working at productive firms is larger, and the planner faces lower efficiency costs. As a result, the planner sets steeper markdown schedule across firms and achieves more redistribution at the optimum.

4 Optimal Regulation

We now analyze optimal policy in the full model with multiple types of workers with different skills—the baseline mode presented in Section 2. The trade-off between redistribution and efficiency investigated in the simple model with single skill type is still an important mechanism of shaping the optimal policy in the full model. In the presence of multiple types of workers, however, there is additional motive of redistribution between different skill types. Since a typical firm hires multiple types of workers together, to separate this redistribution between skill types from the redistribution across workers at different firms within skill type, we need a type-specific policy for each firm. With the type-dependent subsidy, the characterization of optimal allocation with multiple types is the same as in the economy with only single type (See Online Appendix A for more details).

However, the assumption of type-specific policy is surely unrealistic. We commonly observe size-dependent policies for firms in many countries, but the skill-dependent subsidy to the firm's labor cost is rarely used. Moreover, our definition of skill includes not only education level but also worker's general income generating ability.[11] Thus a skill-dependent policy is hard to implement either because of the government's administrative cost or because of the firm's response to avoid tax (such as relabeling of labor type).

With this realistic concerns, an additional constraint naturally arises. We assume that the constrained planner can only use type-independent policy. The planner's ability is restricted in a sense that if the planner wants to increase the wages of a low-skilled worker working at a particular firm, the planner should also raise the wages of workers with different skills at the same firm. We formalize this restriction by introducing no-discrimination constraint to our constrained planner's problem to study the optimal redistribution policy.

[11] This broadly defined definition of skill is as in the standard Mirrleesian labor income tax literature.

4.1 Constrained Planner's Problem

Given the distribution of type-s workers' consumption allocation $\{c_{sj}\}$ across firms, we denote the indirect utility of a type s worker with idiosyncratic preference draw ε by

$$V_s(\varepsilon; \{c_{sj}\}) = max_j \ \log c_{sj} + \log g_{sj} + \varepsilon_j.$$

Note that the function $V_s(\varepsilon; \{c_{sj}\})$ takes into account workers' optimal workplace decisions.

The welfare function is the weighted sum of workers' and firm owners' utilities, given by

$$\mathscr{W} = N_h \int_\varepsilon V_h(\varepsilon; \{c_{hj}\}) dF(\varepsilon) + N_\ell \int_\varepsilon V_\ell(\varepsilon; \{c_{\ell j}\}) dF(\varepsilon) + \psi_f \cdot \sum_j \log c_{fj},$$

where ψ_f is relative Pareto weight on firm owners and $\{c_{fj}\}$ is consumption of firm owners. The welfare function captures the planner's motive to redistribute not only between workers and firm owners but also also across workers. The redistribution across workers combines both the redistribution across different firms within skill type and the redistribution between skill types. Using the indirect utility $V_s(\varepsilon; \{c_{sj}\})$ as an input in the welfare function captures the notion that once the planner has set the consumption allocation $\{c_{sj}\}$, the planner should respect the optimal firm choice of workers.

The planner who cannot use skill-specific policy faces additional constraints which we call *no-discrimination constraints* (hereafter NDC):

$$\frac{c_{hj}}{MR_{hj}} = \frac{c_{\ell j}}{MR_{\ell j}}, \quad \forall j = 1, \ldots, J. \tag{12}$$

This constraint implies that to increase the consumption allocation of a low-skilled worker working at firm j, the planner should also raise the consumption of a high-skilled worker working at the same firm. In other words, when subsidizing a firm, the planner cannot direct all of its resources to a particular type of worker hired at that firm. Even though our planner can create wedges between marginal products of labor and marginal costs for each firm, she is subject to the constraint that such wedges should be equal across different types of workers within a firm.

The form of the no-discrimination constraint (12) is motivated from the decentralization with skill-independent policy. Suppose that the government can set subsidy schedule $\{S_j(y_j)\}$

such that $\sum_j S_j(y_j) = 0$. Firm j solves the following problem given S_j:

$$\begin{aligned}\pi_j = max_{w_{hj},w_{\ell j},L_{hj},L_{\ell j}} \quad & y_j + S_j(y_j) - (w_{hj}L_{hj} + w_{\ell j}L_{\ell j}) \\ s.t. \quad & y_j = z_j(\alpha_j L_{hj}^{\frac{\sigma-1}{\sigma}} + (1-\alpha_j)L_{\ell j}^{\frac{\sigma-1}{\sigma}})^{\frac{\sigma}{\sigma-1}} \\ & L_{hj} = L_{hj}^f(w_{lj}) = N_h\frac{(w_{hj}g_{hj})^{\beta_h}}{I_h}, \quad L_{lj} = L_{lj}^f(w_{\ell j}) = N_l\frac{(w_{\ell j}g_{\ell j})^{\beta_l}}{I_\ell},\end{aligned}$$

The firm's optimality condition shows the wedges in equilibrium:

$$w_{hj} = (1 + S_j'(y_j))\frac{\beta}{1+\beta}MR_{hj}, \quad w_{\ell j} = (1 + S_j'(y_j))\frac{\beta}{1+\beta}MR_{\ell j}.$$

In particular, note that

$$(1 + S_j'(y_j))\frac{\beta}{1+\beta} = \frac{w_{hj}}{MR_{hj}} = \frac{w_{\ell j}}{MR_{\ell j}},$$

which implies our no-discrimination constraint (12). The specific form of the NDC actually depends on the policy tools the planner can use in the decentralization. The NDC (12) is derived from the decentralization with subsidy to the firms only, thus it implicitly assumes that there is no income tax-transfer policy in the decentralization. In Section 6, we extend our analysis to the case with income tax-transfer policy to show the robustness of our results.

We now define the constrained planner's problem with the NDC.

Definition 8. *Constrained planer's problem with no-discrimination constraints is defined as*

$$\begin{aligned}\max_{\{c_{hj},c_{\ell j},c_{fj}\},\{L_{hj},L_{\ell j}\},I_h,I_\ell} \quad & N_h\int_\varepsilon V_h(\varepsilon;\{c_{hj}\})dF(\varepsilon) + N_\ell\int_\varepsilon V_\ell(\varepsilon;\{c_{\ell j}\})dF(\varepsilon) + \psi_f\cdot\sum_j \log c_{fj} \\ s.t. \quad & \sum_j (c_{hj}L_{hj} + c_{\ell j}L_{\ell j} + c_{fj}) \le \sum_j z_j(\alpha_j L_{hj}^{\frac{\sigma-1}{\sigma}} + (1-\alpha_j)L_{\ell j}^{\frac{\sigma-1}{\sigma}})^{\frac{\sigma}{\sigma-1}} \quad (13) \\ & \frac{c_{hj}}{MR_{hj}} = \frac{c_{\ell j}}{MR_{\ell j}}, \quad \forall j = 1,\dots,J \quad (14) \\ & I_h = \sum_j (c_{hj}g_{hj})^\beta, \quad I_\ell = \sum_j (c_{\ell j}g_{\ell j})^\beta \quad (15) \\ & L_{hj} = N_h\frac{(c_{hj}g_{hj})^\beta}{I_h}, \quad L_{\ell j} = N_\ell\frac{(c_{\ell j}g_{\ell j})^\beta}{I_\ell}, \quad \forall j = 1,\dots,J, \quad (16)\end{aligned}$$

As in the planner's problem in Section 3, the employment level of high- and low-skilled workers at each firm should also be consistent with workers' optimal decision, which is guaranteed by (15) and (16).

4.2 Optimal Allocation

By solving the constrained planner's problem in Definition 8, we obtain the optimal allocation. Next proposition characterizes the optimal consumption allocation (see Appendix A for the proof).

Proposition 9. *The optimal consumption allocation c_{sj} for each type of workers at firm j is*

$$c_{hj} = \frac{1}{\Omega_{hj}} \times \left(\frac{\beta}{1+\beta} MR_{hj} + \Delta_h \right), \quad c_{\ell j} = \frac{1}{\Omega_{\ell j}} \times \left(\frac{\beta}{1+\beta} MR_{\ell j} + \Delta_\ell \right) \tag{17}$$

where Ω_{hj} and $\Omega_{\ell j}$ are

$$\Omega_{hj} = 1 - \frac{\zeta_j}{\lambda(1+\beta)} \frac{1}{MR_{hj} L_{hj}} \left(1 + \frac{\beta}{\sigma} \right), \quad \Omega_{\ell j} = 1 + \frac{\zeta_j}{\lambda(1+\beta)} \frac{1}{MR_{\ell j} L_{\ell j}} \left(1 + \frac{\beta}{\sigma} \right),$$

respectively, and Δ_h, Δ_ℓ are

$$\Delta_h = \frac{1}{\lambda(1+\beta)} \left(1 - \beta \nu_h \frac{I_h}{N_h} \right), \quad \Delta_\ell = \frac{1}{\lambda(1+\beta)} \left(1 - \beta \nu_\ell \frac{I_\ell}{N_\ell} \right).$$

λ is Lagrange multiplier on the resource constraint (13)*, ζ_j is Lagrange multiplier on the no-discrimination constraint* (14)*, ν_h, ν_ℓ are Lagrange multipliers on the index equations* (15)*.*

The optimal consumption formula (17) closely resembles the optimal formula in the simple model with single type worker, apart from the modification done by Ω_{sj}'s. Note that when the no-discrimination constraint is not binding, and $\zeta_j = 0$ implies $\Omega_{sj} = 1$, which exactly goes back to the result with type-specific policy (in Online Appendix A). The planner still faces the trade-off between redistributing consumption across identically skilled workers at different firms and enhancing efficiency. The planner who has more concerns for workers relative to the entrepreneurs sets positive Δ_s to transfer resources from the firm owners to the workers and, at the same time, to reduce consumption inequality of identically skilled workers working at different firms.

In the presence of no-discrimination constraint, however, the optimal consumption allocation is modified by Ω_{sj}. Note that the equality constraints (14) can be actually rewritten as the inequality:

$$\frac{c_{hj}}{MR_{hj}} \geq \frac{c_{\ell j}}{MR_{\ell j}}, \quad \forall j = 1, \cdots, J,$$

and thus the Lagrange multipler ξ_j is nonnegative. That is, the planner who wants to redistribute from the high skill type to the low skill type within a firm j wants to decrease consumption of high skill type c_{hj} and increase the consumption of low skill type $c_{\ell j}$ at the same firm, but the planner cannot do this with the NDC. If the NDC is binding, then $\zeta_j > 0$ and thus we have $\Omega_{hj} < 1$ and $\Omega_{\ell j} > 1$. With this modification, the consumption allocation for the high-skilled workers working at firm j is adjusted upward (relative to the consumption without no-discrimination constraint), and the consumption of the low-skilled at the same firm is lowered. Then how does the planner achieve redisdribution across skill types? We answer this question by characterizing the optimal markdown schedule.

4.3 Optimal Markdown and Subsidy Characterization

Despite the equity concerns between differently skilled workers working at the same firm, there is no direct tool to achieve this for the constrained planner. The planner who can only use the same markdown for both skill types of each firm can indirectly provide redistribution between the skill types by adjusting the markdown schedule across firms according to the firm's labor share of the skill types. Moreover, with the same markdown for both skill types, the motive of redistribution across workers at different firms does not necessarily lead to a simple wage compression against the firm's productivity dispersion. The planner should consider how the average marginal revenue products—average between the skill types withing a firm—varies across firms with different productivities. How do these considerations shape the optimal markdown schedule?

We define the optimal (type-independent) markdown at firm j by

$$\mu_j = \frac{c_{sj}}{MR_{sj}}, \quad s \in \{h, \ell\},$$

and we now characterize the optimal markdown schedule in the next proposition (see Appendix A for the proofs).

Proposition 10. *Optimal markdown of a firm j is given by*

$$\mu_j - \frac{\beta}{1+\beta} = S_{hj} \times \frac{\Delta_h}{MR_{hj}} + S_{\ell j} \times \frac{\Delta_\ell}{MR_{\ell j}}$$

where the weight S_{sj} of type s worker within firm j is

$$S_{hj} = \frac{\alpha_j L_{hj}^{\frac{\sigma-1}{\sigma}}}{\alpha_j L_{hj}^{\frac{\sigma-1}{\sigma}} + (1-\alpha_j) L_{\ell j}^{\frac{\sigma-1}{\sigma}}}, \quad S_{\ell j} = \frac{(1-\alpha_j) L_{\ell j}^{\frac{\sigma-1}{\sigma}}}{\alpha_j L_{hj}^{\frac{\sigma-1}{\sigma}} + (1-\alpha_j) L_{\ell j}^{\frac{\sigma-1}{\sigma}}} = 1 - S_{h_j}.$$

The slope of the optimal markdown schedule $\mu_j = \mu(z_j)$ is decomposed into two terms:

$$\frac{d\mu_j}{dz_j} = F1 + F2, \tag{18}$$

$$where \quad F1 = \left.\frac{d\mu_j}{dz_j}\right|_{\frac{L_{hj}}{L_{\ell j}} \ given} \tag{19}$$

$$F2 = \frac{\partial \mu_j}{\partial \left(\frac{L_{hj}}{L_{\ell j}}\right)} \times \frac{d\left(\frac{L_{hj}}{L_{\ell j}}\right)}{dz_j}. \tag{20}$$

Proposition 10 shows that the net markdown $\mu_j - \frac{\beta}{1+\beta}$ (optimal markdown net of the markdown in an equilibrium without regulation) is given by the weighted sum of $\frac{\Delta_h}{MR_{hj}}$ and $\frac{\Delta_\ell}{MR_{\ell j}}$, with weights given by S_{hj} and $S_{\ell j}$. Thus the optimal markdown schedule across firms is determined by the variation of $\frac{\Delta_s}{MR_{sj}}$ and S_{sj} across firms.

The slope of optimal markdown schedule $\frac{d\mu_j}{dz_j}$ expressed in Proposition 10 characterizes the subsidy schedule across firms with different productivities, which is shaped by two distinct channels. The $F1$-term (19) captures the planner's motive to redistribute consumption across identically skilled workers working at different firms, while the $F2$-term (20) reflects the planner's motive to redistribute between workers with different skill types. We investigate each term one by one (see Appendix A for the proof).

Proposition 11. *The $F1$ term in the slope of the optimal markdown $\mu(z_j)$ is given by*

$$\begin{aligned} F1 &= F_{1z} + F_{1\alpha} \\ where \quad F_{1z} &= -\left[S_{hj}\frac{\Delta_h}{MR_{hj}^2}\frac{\partial MR_{hj}}{\partial z_j} + S_{\ell j}\frac{\Delta_\ell}{MR_{\ell j}^2}\frac{\partial MR_{\ell j}}{\partial z_j}\right] < 0, \\ F_{1\alpha} &= \begin{cases} \frac{\sigma}{1-\sigma} \cdot \frac{(L_{hj}^{\frac{\sigma-1}{\sigma}} - L_{\ell j}^{\frac{\sigma-1}{\sigma}})}{\alpha_j L_{hj}^{\frac{\sigma-1}{\sigma}} + (1-\alpha_j) L_{\ell j}^{\frac{\sigma-1}{\sigma}}} \times \left[S_{hj}\frac{\Delta_h}{MR_{hj}} + S_{\ell j}\frac{\Delta_\ell}{MR_{\ell j}}\right] \times \alpha'(z_j), & \sigma \neq 1 \\ -\log\left(\frac{L_{hj}}{L_{\ell j}}\right) \times \left(\alpha_j \frac{\Delta_h}{MR_{hj}} + (1-\alpha_j)\frac{\Delta_\ell}{MR_{\ell j}}\right) \times \alpha'(z_j), & \sigma = 1 \end{cases} \end{aligned} \tag{21}$$

If α_j is constant for all j, then $F_{1\alpha} = 0$. On the other hand, if we assume that $\alpha_j = \alpha(z_j)$ and $\alpha'(z_j) > 0$, then the sign of F_{1a} term is determined in the following way:

$$F_{1\alpha} \leq (>) 0, \quad if \quad L_{hj} \geq (<) L_{\ell j}.$$

Proposition 11 shows that the $F1$ term—redistribution across workers at different firms—can be further decomposed into two channels. The first channel F_{1z} captures the direct effect of increasing z_j on the markdown schedule, which is exactly the same mechanism as in Proposition 6 with only single-type workers. We focus on the case where $\Delta_s > 0$ for all s. Then the sign of F_{1z} is negative. The planner attempts to close the wage gap across firms within a skill type by compressing the wage distribution across firms, which provides a force for wider markdown of wage—lower μ_j—for large firms with high marginal productivities.

On the other hand, the second term $F_{1\alpha}$ captures the effect of increasing z_j on the markdown schedule *through the change in skill intensity* $\alpha(z_j)$ (holding the ratio of employment $\frac{L_{hj}}{L_{\ell j}}$ constant). Different from F_{1z}, $F_{1\alpha}$ can be either positive or negative depending on the relative size of $\frac{L_{hj}}{L_{\ell j}}$ at firm j. The planner still wants to provide redistribution against the dispersion of marginal revenue products across firms, but the skill intensity schedule which is increasing in firm's productivity ($\alpha'(z_j) > 0$) can either increase or decrease the dispersion of average marginal revenue product depending on the employment share. When skill intensity increases in firm's productivity, the increase in skill intensity reduces the gap of average marginal product of labor between firms with different productivities in the region where low-skilled workers have a larger employment share $L_{hj} < L_{\ell j}$, while it increases the gap in the region with $L_{hj} > L_{\ell j}$. Smaller (larger) dispersion of average marginal product across firms of different productivities calls for less (more) need for redistribution, which provides forces for increase (decrease) in subsidy rate in firm size.

In sum, the redistribution motive across workers at different firms can make the slope of $\mu(z_j)$-schedule either positive or negative depending on the employment share of different skills and the dominance between F_{1z} and $F_{1\alpha}$. We need to carry out numerical investigation to determine the sign. We next move to the investigation of the second term $F2$.

Proposition 12. *The F2-term in the slope of the optimal markdown $\mu(z_j)$ is given by*

$$F2 = \Psi \times \left(\frac{\Delta_h}{MR_{hj}} - \frac{\Delta_\ell}{MR_{\ell j}} \right) \times \left[\frac{d\left(\frac{L_{hj}}{L_{\ell j}}\right)}{d\alpha_j} \times \alpha'(z_j) + \frac{d\left(\frac{L_{hj}}{L_{\ell j}}\right)}{d\left(\frac{g_{hj}}{g_{\ell j}}\right)} \times \frac{d\left(\frac{g_{hj}}{g_{\ell j}}\right)}{dz_j} \right], \tag{22}$$

$$\text{where} \qquad \Psi = \alpha_j(1-\alpha_j) \times \frac{L_{hj}^{\frac{\sigma-1}{\sigma}} L_{\ell j}^{\frac{\sigma-1}{\sigma}}}{\left(\alpha_j L_{hj}^{\frac{\sigma-1}{\sigma}} + (1-\alpha_j) L_{\ell j}^{\frac{\sigma-1}{\sigma}}\right)^2} \times \left(\frac{L_{hj}}{L_{\ell j}}\right)^{-1},$$

$$\frac{\Delta_h}{MR_{hj}} - \frac{\Delta_\ell}{MR_{\ell j}} = -\frac{\zeta_j}{\lambda(1+\beta)} \left(1 + \frac{\beta}{\sigma}\right) \times \frac{1}{y_j} \times \left(\frac{L_{hj}^{\frac{1-\sigma}{\sigma}}}{\alpha_j} + \frac{L_{\ell j}^{\frac{1-\sigma}{\sigma}}}{1-\alpha_j} \right) \times \mu_j.$$

If the no-discrimination constraint is not binding with $\zeta_j = 0$, then $F_2 = 0$. On the other hand, if the no-discrimination constraint is binding ($\zeta_j > 0$), then $\frac{\Delta_h}{MR_{hj}} - \frac{\Delta_\ell}{MR_{\ell j}} < 0$. Then the sign of $F2$ is determined by:

$$sign(F2) = -sign\left(\frac{d\left(\frac{L_{hj}}{L_{\ell j}}\right)}{d\alpha_j} \times \alpha'(z_j) + \frac{d\left(\frac{L_{hj}}{L_{\ell j}}\right)}{d\left(\frac{g_{hj}}{g_{\ell j}}\right)} \times \frac{d\left(\frac{g_{hj}}{g_{\ell j}}\right)}{dz_j} \right).$$

Proposition 12 shows that the F2-term (20) captures the planner's motive to redistribute between different types of workers. This can be best understood by observing that without NDC, $F2$-term does not show up as the planner has its separate tool for redistribution between skill types.

If $\zeta_j > 0$, the planner wants to redistribute from high- to low-skilled workers but cannot directly achieve this due to the NDC. The planner provides this redistribution indirectly by setting the subsidy according to the firm's labor share of each type. If the labor share of high-skilled workers is higher at productive firms, then at such firms a larger proportion of subsidies is allocated to the high-skilled workers. Thus the planner who wants to redistribute from the high-skilled to the low-skilled assigns lower subsidy (or wider markdown of wage with lower μ_j) for such high productivity firms, generating forces for a negative slope of the $\mu(z_j)$—negative $F2$. If the labor share of each type of workers is constant across firms, however, $F2 = 0$ because despite the redistributive concerns between different types of workers, it is impossible to close the wage gap by subsidizing a particular firm.

In sum, negative F_{1z} and $F2$ terms make the $\mu(z_j)$-schedule decreasing in firm's productivity, which provide redistribution by compressing the wage schedule. On the other hand, $F_{1\alpha}$

makes the $\mu(z_j)$-schedule non-monotone if this complementarity channel is strong enough. We numerically investigate the $\mu(z_j)$-schedule in Section 5.

4.4 Sources of Sorting and the Optimal Regulation

One of the key findings of Lamadon, Mogstad, and Setzler (2022) is that the dominant source in explaining why better workers sort into larger (and more productive) firms is the production complementarity not the heterogeneous tastes for amenities.[12] What is the implication of this feature on the optimal firm regulation? In this section, we discuss how different sources of sorting affect the shape of optimal regulation schedule across firms.

Our model permits two potential sources for sorting of high-skilled workers at high-productivity firms : (i) under complementarity between the firm's productivity and worker's skill ($\alpha'(z_j) > 0$), more productive firms have higher demand for skilled workforce thus employ such workers by a larger proportion relative to low-productivity firms do, and (ii) if workers with higher skill have stronger preferences for amenities and more productive firms provide more amenities to workers $\left(\frac{d\left(\frac{g_{hj}}{g_{\ell j}}\right)}{dz_j} > 0\right)$, then high-skilled workers value working at productive firms more than low-skilled workers do.

The source of sorting matters for determining the optimal allocation and regulation schedule across firms. Depending on the source of sorting, an increase in the degree of positive sorting between firms and workers can affect the optimal $\mu(z_j)$-schedule in a different manner. The key reason is because the size of $F_{1\alpha}$ term depends on the source. The demand-side sorting affects the marginal revenue product dispersion across firms while the supply-side sorting does not. When the sorting is mostly coming from the complementarity between firm's productivity and worker's skill, the $F_{1\alpha}$-channel becomes stronger, which makes the optimal schedule non-monotone. In the hypothetical economy where supply-side sorting through amenity channel dominates, positive sorting between worker skills and firm produc-

[12]The importance of worker-firm sorting in explaining the firm size-wage gap(Brown and Medoff (1989), Oi and Idson (1999)) or the firm-side effects on wage inequality(Song, Price, Guvenen, Bloom, and von Wachter (2019)) has been widely recognized.

tivities makes the slope of optimal subsidy schedule more negative (only through the $F2$ channel). We investigate this quantitatively in Section 5.

On the other hand, the effects of sorting on the $F2$ term in the slope of markdown—the term to redistribute between differently skilled workers—does not depend on its sources. Both sources of sorting contribute to larger employment ratio of high-skilled workers at high-productivity firms. This exerts a downward pressure on subsidy rates on larger firms by strengthening the $F2$ term regardless of the sources.

Important lesson from this section and Section 3.4 is that the determinant of the shape of the optimal firm subsidy schedule is not just the distribution of the firm premium but the detailed sources of wage inequality and sorting mechanisms. This implies that simple policy recommendation based on sufficient statistics such as variance of wages between firms and correlation between firm size and skilled worker's employment share can be misleading.

5 Quantitative Analysis

5.1 Quantitative Model and Parametric Forms

We extend the baseline model in Section 2 to include more than two types of workers for the quantitative investigation. There are M types of differently skilled workers with the mass of the each type N_m and the total mass $N = \sum_{m=1}^{M} N_m$.

There are J firms in our economy. Firm j is characterized by its productivity z_j and level of amenities g_j. We assume that (z_j, g_j) follows a bivariate lognormal distribution:

$$\begin{pmatrix} \log z_j \\ \log g_j \end{pmatrix} \sim N(0, \Sigma), \quad \text{where} \quad \Sigma = \begin{pmatrix} \sigma_z^2 & \rho_{zg}\sigma_z\sigma_g \\ \rho_{zg}\sigma_z\sigma_g & \sigma_g^2 \end{pmatrix},$$

where σ_z and σ_g determine the variance of productivity and amenities across firms, and ρ_{zg} controls the correlation between the two.

Firm j with productivity z_j produces final output y_j using a production technology

$$y_j = z_j \tilde{y}_j^{\eta}, \quad \tilde{y}_j = \left(\sum_m \alpha_{mj} L_{mj}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}},$$

where $\tilde{y}_j$ is the total efficiency unit of labor, η governs the returns to scale, α_{mj} is the efficiency of a type m worker working at firm j, L_{mj} is the number of type-m workers employed at firm j, and σ is the elasticity of substitution between differently skilled workers.

The efficiency of a worker with type m at firm j, α_{mj}, takes the following parametric form:

$$\alpha_{mj} = X_m^{\theta_j}, \quad \text{where} \quad \theta_j = 1 + \theta \cdot (\log z_j - \kappa).$$

A type-m worker has a labor productivity X_m which is common across all firms. We assume that X_m follows a log-normal distribution $\log X_m \sim N(0, \sigma_x^2)$. Each firm has a firm-specific skill intensity θ_j. Importantly, the parameter θ governs the degree of complementarity between worker's skill and firm's productivity. If $\theta = 0$, a given type worker has identical labor efficiency in all firms ($\theta_j = 1$ and $\alpha_{mj} = X_m, \forall j, m$). If $\theta > 0$, then high-skilled workers are relatively more productive than low-skilled workers at more productive firms, and thus skilled workers sort into high productivity firms. Finally, κ controls the level of overall wage variance. Note that there is no supply-side sorting in this baseline analysis as we assume that worker's valuation of amenities does not depend on type ($g_{mj} = g_j$). We relax this in Section 5.4.

5.2 Calibration

In our simulation, there are $M = 10$ types of differently skilled workers with total mass $N = 1,000,000$. The number of firms is $J = 100,000$. We externally calibrate the parameters β, η, σ, and κ. We set the labor supply elasticity $\beta = 6$ and the returns to scale parameter $\eta = 0.75$.[13,14] We assume perfect substitution between differently skilled workers, i.e. $\sigma = \infty$.[15] Finally, we set $\kappa = 1.5$.[16]

[13] Studies estimating the labor supply elasticity using the responses of employment and wages to exogenous shocks report values ranging from 4 to 6 (**?**).

[14] Typically used values for η range from 0.7 to 0.9. We set relatively low value of η to attain a reasonable degree of firm size distribution.

[15] Under the assumption of $\sigma = \infty$ we are able to analytically characterize the equilibrium worker wage as a function of three components—worker skill x_m, firm pay premium ψ_j, and complementarity in production θ_j. We use the empirical moments of these components as calibration targets.

[16] We set κ to approximate the overall wage variance. Specifically, we aimed to match the wage ratio of 90th to 10th quantile workers working at a median productivity firm.

We then jointly calibrate the standard deviation of worker skill σ_x, firm productivity σ_z, firm amenity σ_g, the correlation between productivity and amenity ρ_{zg}, and the complementarity parameter θ to match the variance decomposition of worker's wage. In equilibrium, log wage can be decomposed as follows (see Online Appendix B for the derivation):

$$\log w_{mj} = \underbrace{\log X_m}_{x_m} \times \underbrace{(1+\theta(\log z_j - \kappa))}_{\theta_j} + \underbrace{\underbrace{\frac{\log z_j}{1+\beta(1-\eta)}}_{\psi_{zj}} + \underbrace{\log \frac{\beta\eta}{1+\beta} - (1-\eta)\log H_j}_{\psi_{gj}}}_{\psi_j}. \quad (23)$$

A worker with higher skill x_m earns higher wages, and even more so if the worker works at high productive firm (with higher θ_j) when $\theta > 0$. In addition, workers working at a firm paying higher premium ψ_j earn higher wages. Firms pay high premium ψ_j either because they have higher productivity (ψ_{zj}) or provide lower level of amenities (ψ_{gj}).[17]

We use the estimates of the wage decomposition equation (23) from Lamadon, Mogstad, and Setzler (2022) for our calibration. We set the standard deviation of log of workers' skills $\sigma_x = 0.55$ to match the component of log wage variance due to the type-specific labor productivity $Var(x_m) = 0.31$, and we set the complementarity parameter $\theta = 0.36$ to match the degree of sorting of high-skilled workers to higher paying firms $Corr(x_m, \psi_j) = 0.37$. The parameters of the joint distribution of firms' productivities and amenities are set as $\sigma_z = 0.90$, $\sigma_g = 0.43$, and $\rho_{zg} = 0.90$ to match the variance decomposition of firm pay premium ($Var(\psi_z) = 0.17$, $Var(\psi_g) = 0.13$, $2Cov(\psi_z, \psi_g) = -0.27$).

Table 1 summarizes the calibrated parameters and their targeted moments. Figure 3 plots the predicted log earnings for each skill type working at different firms. It shows that workers at high-productivity firms with higher premiums ψ_j earn higher wages, and the wage gains from more productive firms are greater for skilled workers.[18]

[17] We define $\psi_{gj} = \log \frac{\beta\eta}{1+\beta} - (1-\eta)\log H_j$, where H_j is a firm-specific parameter which positively depends on the firm's amenity g_j, but not on productivity z_j. In Online Appendix B, we show that H_j is solution to the following fixed point equation: $H_j = \left(\sum_m \left(N_m \frac{\alpha_{mj}^{1+\beta} g_{mj}^{\beta}}{\sum_k (z_k^{\frac{\beta}{1+\beta(1-\eta)}} H_k^{-\beta(1-\eta)} \alpha_{mk}^{\beta} g_{mk}^{\beta})} \right) \right)^{\frac{1}{1+\beta(1-\eta)}}$.

[18] The variance of overall wages in our model ($Var(\log w) = 0.75$) is higher than the level documented in the data (0.40). However, in quantifying the optimal markdown and its decomposition, capturing the firm-induced component of wages and employment composition is of first-order importance, a feature that our calibration effectively captures.

Parameters	Targeted Moments	Values
σ_x	$Var(x)$	0.5567
θ	$Corr(x, \psi)$	0.5408
σ_z	$Var(\psi_z)$	0.7497
σ_g	$Var(\psi_g)$	0.4542
ρ_{zg}	$2Cov(\psi_z, \psi_g)$	0.9199

Table 1: Internally Calibrated Parameters

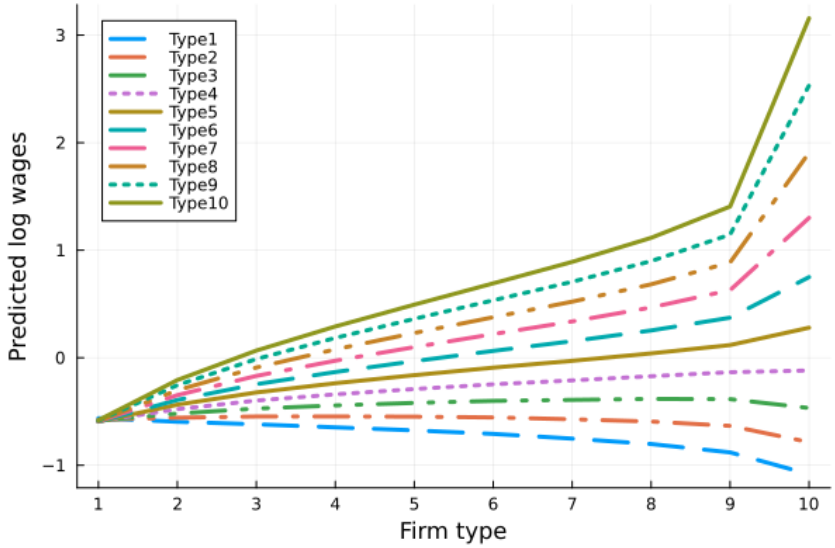


Figure 3: Predicted log earnings

Note: Firm types are assigned by ranking and grouping firms according to their productivities. Predicted log earnings for each type of worker are calculated by taking conditional expectations of those workers' wages working at each group of firms.

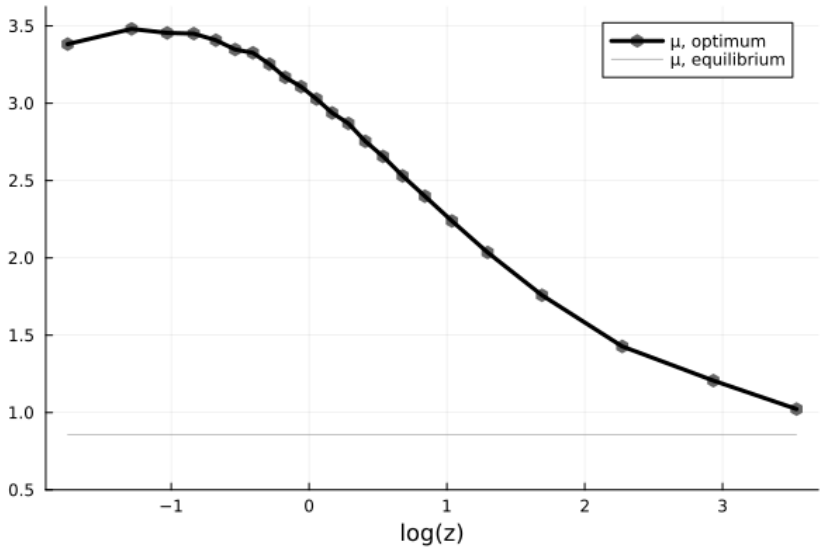


Figure 4: Optimal Markdown Schedule

Note: We plot the conditional averages of optimal markdowns for firms with productivity z_j. We assign each firms to 23 groups according to their productivities and take the conditional medians of μ_j for firms in each group.

5.3 Optimal Markdown Schedule

In this quantitative model, the optimal markdown—the ratio of wage to the marginal product of labor—of firm j is given by

$$\mu_j \quad = \quad \frac{\beta}{1+\beta} \; + \; \sum_m \left(\frac{\alpha_{mj} L_{mj}}{\sum_k \alpha_{kj} L_{kj}} \; \times \; \frac{\Delta_m}{MR_{mj}} \right), \tag{24}$$

where $MR_{mj} = \eta z_j L_j^{\eta-1} \alpha_{mj}$ is the marginal product of a type-m worker working at firm j (see Online Appendix B for the derivation).

Figure 4 plots the optimal markdown schedule across the firm's productivity distribution. Throughout the quantitative analysis, we impose zero Pareto weight on the firm owners ($\psi_f = 0.0$). The average of the optimal μ_j (weighted by employment) is 1.86, while the average μ_j in equilibrium is 0.86. Thus on average, the optimal wages are much less marked down from the marginal products of labor compared to the equilibrium wages. The marginal subsidy rates implied by the optimal markdown schedule are positive for all firms.[19] By

[19]Subsidy rate is computed by $1 + s_j = \frac{1+\beta}{\beta}\mu_j$. The marginal subsidy rates are 1.17 on average.

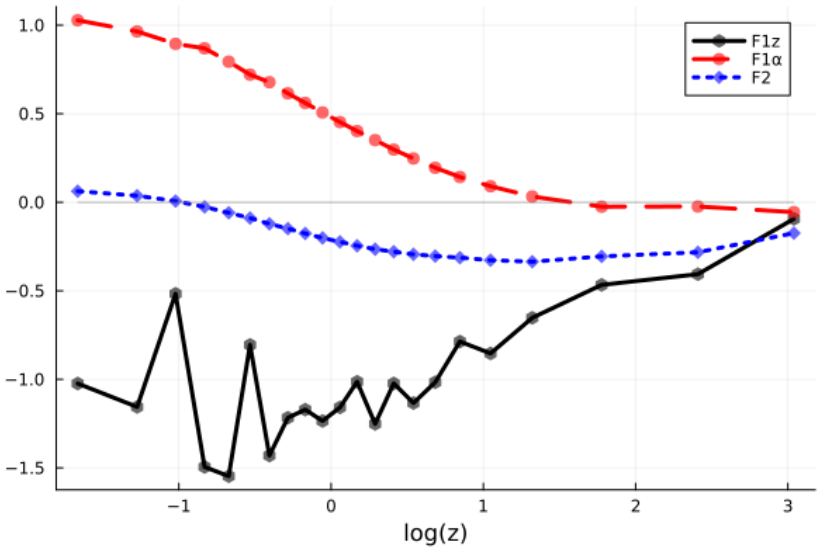


Figure 5: Decomposition of $\frac{d\mu}{d\log z}$

Note: We assign each firms to 23 groups according to their productivities and plot the decomposition of the slope of $E(\mu_j|z_j)$. See Online Appendix B for the details of the numerical approximation procedure.

subsidizing firms, the planner stimulates the overall demand for workers, thereby raising wages.

More important, Figure 4 also shows that the optimal markdown schedule exhibits an inverted U-shape. Figure 5 shows the forces of shaping this schedule with the numerical decomposition of the slope of the markdown $\frac{d\mu}{d\log z}$. Recall that the slope is determined largely by two forces: (i) $F1$—to redistribute across workers at different firms—combining the direct effect of z (F_{1z}) and indirect effect through α ($F_{1\alpha}$), and (ii) $F2$—to redistribute between differently skilled workers. The downward sloping schedule of $\mu(z_j)$ in the middle and top region of the firm's productivity distribution is driven by negative F_{1z} and $F2$.[20] Both terms make the planner set wider markdown of wage—lower μ—for more productive firms. With negative F_{1z}, the planner reduces the wage inequality across firms against the dispersion of firm productivity z_j. With negative $F2$, we can indirectly close the wage gap between high- and low-skilled workers as the planner subsidize skill-intensive firms less.

[20] $F2$ is slightly positive at the bottom because our calibration produces a counterfactual feature that the wages of the low-skilled are higher than those of the high-skilled at the bottom of the firm productivity distribution as we match the highly positive sorting ($Corr(x_m, \psi_j)$) with restrictive parametric form of α_{mj}.

The inverted U-shape of the $\mu(z_j)$-schedule in the regions where firms' productivities are below the median ($\log z_j = 0$) is driven by the $F_{1\alpha}$. The sign of $F_{1\alpha}$ is *positive* in most regions, and only becomes negative in the region with high firm's productivity. At the left tail of the firm's productivity distribution where the majority of employment consists of low-skilled workers, $F_{1\alpha}$ is largely positive thereby *attenuating* (or even dominating) the effect of F_{1z}. In this region, higher skill intensity of a more productive firm leads to lower average marginal revenue product within a firm, which calls for more subsidies to more productive firms—wider markdown of wage to less productive firms. At the right tail where skilled workers are predominantly employed, however, $F_{1\alpha}$ is slightly negative, augmenting the F_{1z}.

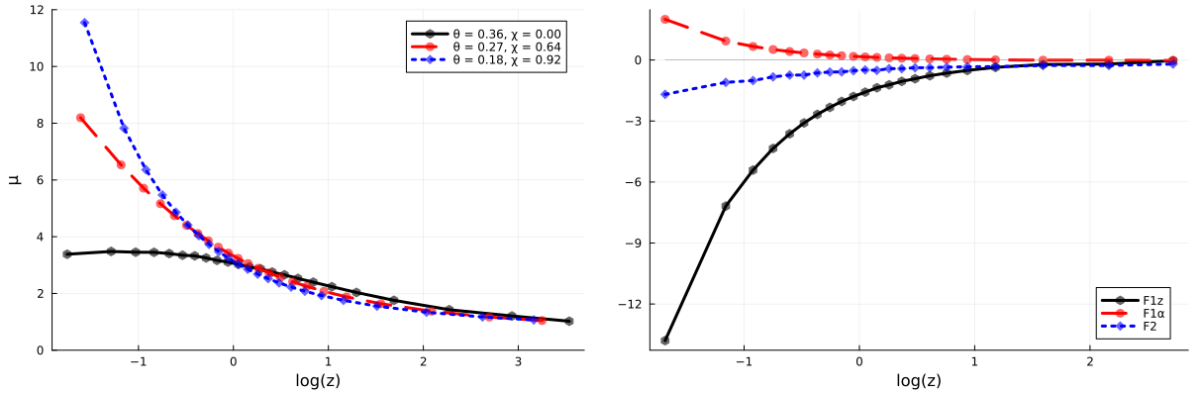
In Online Appendix B, we present robustness results for alternative parameter values of β, η, κ and σ. For lower values of β and σ, the degree of non-monotonicity of the markdown schedule becomes stronger with more dominant $F_{1\alpha}$-effects because the value of complementarity parameter θ to match the observed pattern of worker-firm sorting should be larger.[21] Under lower η, the dispersion of marginal revenue productivity across firms at the optimum allocation becomes larger, which makes the negative slope of the $\mu(z_j)$ steeper. When κ increases so that there is less overall wage inequality between workers, efficiency concerns dominate redistributive motives resulting less steeper downward sloping $\mu(z_j)$ schedule.

5.4 The Effects of Sources of Sorting on Optimal μ-Schedule

In Section 4.4, we have discussed that the sorting mechanism (demand-side vs supply-side) matters for the optimal markdown schedule as it determines the relative size of $F_{1\alpha}$. We now investigate this quantitatively to show that the source of sorting is crucial in determining the shape of optimal markdown, especially at the left tail of the firm's productivity distribution.

In the benchmark quantitative analysis, we have assumed that the positive sorting between firm and worker is purely generated by the demand side—the complementarity between firm's

[21] When β is lower, workers are less sensitive to wage differences across firms, requiring a greater degree of skill-intensity heterogeneity. Similarly, when σ is lower, different types of workers are more complementary in production, resulting in lower value of calibrated θ.



(a) Optimal markdown

(b) Decomposition, $\theta = 0.18, \chi = 0.92$

Figure 6: Sources of sorting

productivity and worker's skill.[22] We now introduce the supply-side sorting by allowing that workers with different skills value the amenities differently. Denote the value that a type-m worker places on the amenity of firm j by g_{mj}. We assume that g_{mj} takes the following form:

$$g_{mj} = g_j^{\chi_m}, \quad \text{where} \quad \chi_m = 1 + \chi \times \log X_m, \tag{25}$$

and g_j is the level of amenity provided by firm j. If $\chi = 0$, then $g_{mj} = g_j$ which goes back to the benchmark case where there is only demand-side sorting. If $\chi > 0$, then skilled workers value amenities provided by the firm relatively more than low-skilled. In this case, when more productive firms provide higher level of amenities ($\rho_{zg} > 0$), skilled workers are more likely to work at more productive firms.

We carry out the following experiment. Starting from the value of complemetarity parameter $\theta = 0.36$ in the baseline calibration ($\chi = 0$), we lower θ. For each value of θ, we recalibrate χ to match the overall worker-firm sorting $Corr(x_m, \psi_j) = 0.37$ observed in the data.[23] For each combination of θ and χ, we compute the optimal markdown and decompose its slope.

[22] Recently several studies have attempted to quantify the importance of each source of sorting, but the literature remains inconclusive. Lamadon, Mogstad, and Setzler (2022) insist that sorting occurs due to demand-side reasons; Lamadon, Lise, Meghir, and Robin (2024) argue that differences of workers' valuation of amenities are important sources of sorting.

[23] We also re-calibrate σ_z, σ_g, and ρ_{zg} to match the variance decomposition of firm pay premium.

Figure 6a plots the optimal markdown schedule for $\theta = 0.36$ (baseline), 0.27 and 0.18. Note that the required value of χ to match the degree of worker-firm sorting $Corr(x_m, \psi_j)$ becomes larger as the demand-side sorting gets weaker (i.e., lower θ). The differences of optimal markdown schedule between these cases are most apparent at the left tail of the productivity distribution. As the supply-side sorting gets stronger, the wage compression force becomes more dominant even in this region. When θ is low enough, the slope of the optimal $\mu(z_j)$-schedule becomes uniformly negative.

To understand what causes these differences, we plot the decomposition of the slope of the $\mu(z_j)$-schedule for $\theta = 0.18, \chi = 0.92$ case in Figure 6b. As we introduce supply-side sorting, we need less complementarity between worker skill and firm productivity to generate worker-firm sorting in the data. Then the magnitude of $F_{1\alpha}$ term, the indirect effect of increasing z through α, becomes much weaker relative to F_{1z} term, the direct effect of z. As a result, the optimal $\mu(z_j)$-schedule is monotonically decreasing in firm's productivity.

6 Firm Regulation in the presence of Labor Income Tax

So far, we have analyzed the optimal firm regulation for workers' redistribution in the absence of more direct policy for redistribution—income tax-transfer. Are the main results robust to the extension with direct redistributive policy? In this section, we show that the qualitative properties of the optimal firm regulation do not change even in the presence of labor income tax, and we also investigate the optimal joint policy.

6.1 Indeterminacy and Role of Labor Income Tax

Analysis with No Restrictions on the Policy

When the planner can access to the skill-dependent policies, there is no role of additional income tax. Intuitively, as the planner is not constrained by the policy tools at all, optimal

allocation does not depend on which tool is used for the decentralization, and the same allocation can be implemented by many different combinations of policies.[24]

For simplicity, we focus on the single skill-type case, which is essentially the same analysis with skill-dependent policy. Next proposition shows that infinitely many combinations of subsidy to the firm and income tax to the workers can implement the same optimal allocation. More important, the property of the subsidy schedule does not change in all cases—the subsidy rate decreases in firm's productivity if the income tax is not too progressive.[25]

Proposition 13. *Consider an economy with single skill type. An optimal allocation* $c_j = \frac{\beta}{1+\beta} MP_{Lj} + \Delta$ *can be implemented by any of the following policies:*

1. ***Revenue dependent subsidy to the firm only:*** *a sum-to-zero subsidy schedule* $\{S_j(y_j)\}$ *to the firm which satisfies* $S_j'(y_j) = \frac{1+\beta}{\beta} \cdot \frac{\Delta}{MP_{Lj}}$,

2. ***Subsidy to the firm and lumpsum transfer to the workers:*** *given some lumpsum transfer* $0 < Tr \leq \Delta$ *to the worker, subsidy schedule* $\{S_j(y_j)\}$ *to the firm which satisfies*[26,27]

$$(1+S_j'(y_j)) \times \frac{\beta \cdot \frac{w_j}{w_j+Tr}}{1+\beta \cdot \frac{w_j}{w_j+Tr}} + \frac{Tr}{MP_{Lj}} = \frac{\beta}{1+\beta} + \frac{\Delta}{MP_{Lj}}, \text{ and } \sum_j S_j(y_j) = -N \times Tr, \text{ and}$$

3. ***Subsidy to the firm and progressive income tax to the workers:*** *given some progressive labor income tax to the worker which has the form of* $T(w_j) = w_j - \lambda \cdot w_j^{1-\tau}$ *and satisfies* $\sum_j T(w_j) N_j = 0$*, a sum-to-zero subsidy schedule* $\{S_j(y_j)\}$ *to the firm which satisfies*

$$\lambda \cdot \left(1 + S_j'(y_j)\right)^{1-\tau} \times \left(\frac{(1-\tau)\beta}{1+(1-\tau)\beta}\right) \times MP_{Lj}^{-\tau} = \frac{\beta}{1+\beta} + \frac{\Delta}{MP_{Lj}}.$$

[24]Berger, Herkenhoff, Mongey, and Mousavi (2024) study the role of progressive taxes in the monopsonistic labor market and show that the progressivity amplifies the inefficiencies associated with labor market power as the lowered elasticity of labor supply leads to wider markdowns and further distortion of the employment allocation. This is the result when the government can only use income tax. When the planner can use other policies, the distortions caused by the progressivity can be eliminated by an appropriate choice of markdown.

[25]Proposition 13 also characterizes the aggregate level of subsidies. When the labor income tax is used for the pure redistribution across workers, the subsidies to the firms sum to zero. On the other hand, with lumpsum transfers to the workers, the aggregate firm's subsidy is negative to fund the transfers.

[26]We also remark that the elasticity of labor supply depends on the lumpsum transfer and who owns the firm's profit in this Logit model. On the other hand, since we assume that the firm's profit is owned by the entrepreneur (not by the workers), the elasticity of labor supply does not depend on the profit.

[27]Even when the lumpsum transfer is exactly set to Δ of the optimal consumption formula, implementation of optimal allocation requires additional subsidy because the transfer reduces the elasticity of labor supply.

Assume that $\frac{\tau}{1-\tau} < \frac{1+\beta}{\beta} \cdot \frac{\Delta}{MP_{Lj}}$, *for all* j. *In all the implementations above, the marginal subsidy rate* $S_j'(y_j)$ *decreases in firm's productivity* z_j.

Proof See Online Appendix C for the proof. ■

We also remark that despite the indeterminacy, as long as the income taxes do not depend on worker's employer, the income tax alone cannot implement the optimal allocation, and the firm's regulation has its own role in worker's redistribution.

Analysis without Skill-dependent Policy

Now suppose that the constrained planner cannot use any skill-dependent policy. Then the optimal allocation can be affected by the income tax as the income tax can relax the NDC. The introduction of the labor income tax has impact on the planner's problem only through its effects on the NDC. This is because in the planner's problem who can directly choose consumption allocation, the labor supply elasticity in terms of consumption does not depend on the labor income tax (we discuss this in more detail in Online Appendix C).

As we briefly discussed in Section 4, the form of the NDC does depend on the set of labor income taxes the planner can access to. Recall that in the baseline with no income tax, the NDC was: $\frac{c_{hj}}{MP_{hj}} = \frac{c_{\ell j}}{MP_{\ell j}} \quad \forall j$. This specific form of the NDC is consistent with the linear proportional tax/transfers (including the zero tax case of the baseline) to the workers. Thus when only the proportional labor income tax/transfer can be used, the planner's problem and optimal allocation do not change from those with no labor income tax.

When we introduce the progressive labor income tax, however, the planner's NDC should be changed accordingly. We focus on the so called HSV functional form for the labor income tax: $T(y) = y - \lambda \cdot y^{1-\tau}$, where τ controls the progressivity and λ controls the level of the tax schedule. Without skill-dependent policies, the markdown within the same firm should be the same for both skill types, which leads to the following NDC in terms of consumption (see Online Appendix C for the derivation):

$$\frac{c_{hj}}{MP_{hj}^{1-\tau}} = \frac{c_{\ell j}}{MP_{\ell j}^{1-\tau}}.$$

As the planner wants to redistribute from the high skilled to the low skilled, the NDC is essentially the following inequality: $\frac{c_{hj}}{MP_{hj}^{1-\tau}} \geq \frac{c_{\ell j}}{MP_{\ell j}^{1-\tau}}$ (if the progressive income tax does not provide enough redistribution), and increasing the progressivity τ has the benefit of relaxing the NDC. To analyze the role of the income tax progressivity in a meaningful way, however, we also need to consider the cost of the progressivity, which is not included in this model with only extensive labor supply. We proceed in this direction in the next section.

6.2 Analysis with Intensive Labor Supply and Optimal Joint Policy

We now investigate the joint policy with both firm's regulation and progressive labor income tax. For the labor income tax, we focus on the HSV function defined above. For the nontrivial optimal progressivity, we need to extend the model to include the cost of progressive tax. We do this by introducing intensive margin of labor supply.

6.2.1 Planner's Problem with Intensive Labor Supply

We first derive the labor supply curve. Workers have preferences represented by $u(c,h) + \log(g)$, where c is consumption, h is working hours and g is amenity. Without loss of generality, we simply assume that firm's provision of amenity is constant with $g_j = 1$ for all j, and all worker's have identical valuation of amenities. A worker with type s who chooses to work at firm j and decides to work h_{sj} hours earns before-tax income $w_{sj}h_{sj}$ and consumes after-tax income $\lambda \cdot (w_{sj}h_{sj})^{1-\tau}$. That is, the worker solves the following problem after choosing firm j:

$$\begin{aligned} V_s(w_{sj}) &= \max_{c_{sj},h_{sj}} u(c_{sj},h_{sj}) \\ s.t. &\quad c_{sj} = \lambda \cdot (w_{sj}h_{sj})^{1-\tau}. \end{aligned} \tag{26}$$

We denote the optimal working hours which solves (26) by $h_{sj}(w_{sj})$. Then the labor supply curve of the firm j for s-skilled workers is derived as follows (see Online Appendix C):

$$L_{sj}(w_{sj}) = N_{sj}(w_{sj}) \times h_{sj}(w_{sj}) = P_{sj} \cdot N_s \times h_{sj}(w_{sj}) = \frac{e^{\beta \cdot V_s(w_{sj})}}{\sum_k e^{\beta \cdot V_s(w_{sk})}} \cdot N_s \times h_{sj}(w_{sj})$$

and both the planner and firms face this labor supply cure.

Then we can derive the no-differentiation constraint by solving the firm's problem given type-independent subsidy (see Online Appendix C for derivation):

$$\textbf{(NDC-}\tau\textbf{):} \quad \frac{w_{hj}}{MP_{hj}} \times \frac{1+\epsilon_h(w_{hj})}{\epsilon_h(w_{hj})} = \frac{w_{\ell j}}{MP_{\ell j}} \times \frac{1+\epsilon_\ell(w_{\ell j})}{\epsilon_h(w_{\ell j})},$$

where $\epsilon_s(w_{sj})$ is the elasticity of labor supply for a type s. Note that the labor supply elasticity is the sum of intensive and extensive elasticities:

$$\epsilon_s(w_{sj}) \equiv \frac{\partial \log L_{sj}}{\partial \log w_{sj}} = \frac{\partial \log N_{sj}}{\partial \log w_{sj}} + \frac{\partial \log h_{sj}}{\partial \log w_{sj}} = \epsilon_s^P(w_{sj}) + \epsilon_s^h(w_{sj}).$$

The extensive labor supply elasticity is derived as follows:

$$\epsilon_s^P(w_{sj}) = \frac{\partial \log N_{sj}}{\partial \log w_{sj}} = \frac{\partial \log P_{sj}}{\partial \log w_{sj}} = \beta \cdot V'(w_{sj}) \cdot w_{sj} = \beta \times (1-\tau) \times u_c \cdot c_{sj},$$

where the last equality is from the envelope condition of worker's problem (26).[28] The intensive labor supply elasticity is $\epsilon_s^h(w_{sj}) = \frac{\partial \log h_{sj}}{\partial \log w_{sj}} = \frac{h'_{sj}(w_{sj}) \cdot w_{sj}}{h_{sj}}$. Thus under the general preferences, NDC-τ depends on the type-dependent elasticities as both the extensive and intensive elasticities can depend on the type-specific wage. Since this constraint makes further analysis complicated, we proceed further with special preferences.

6.2.2 Example with Special Preferences

We assume the following preferences so that elasticities do not depend on the wealth effects.

Assumption 14. $u(c,h) = \log c - \frac{1}{1+\frac{1}{\epsilon_h}} h^{1+\frac{1}{\epsilon_h}}$

Then worker's problem (26) has a closed form solution for the working hours:

$$h_{sj}(w_j) = (1-\tau)^{\frac{\epsilon_h}{1+\epsilon_h}} := h(\tau), \quad s = h, \ell.$$

Under Assumption 14, both extensive and intensive elasticities of labor supply do not depend on s: $\epsilon_s^P(w_{sj}) = \beta(1-\tau)$, and $\epsilon_s^h(w_{s,j}) = 0$, for all s, j. Then the NDC-τ constraint boils down to $\frac{w_{hj}}{MP_{hj}} = \frac{w_{\ell j}}{MP_{\ell j}}$, and in terms of consumption, it is written as $\frac{c_{hj}}{MP_{hj}^{1-\tau}} = \frac{c_{\ell j}}{MP_{\ell j}^{1-\tau}}$, which is the same form of NDC as in the economy with only extensive labor supply. Then it is

[28] The extensive elasticity depends on three factors: (i) dispersion of the idiosyncratic preference β, (ii) tax progressivity τ, and (iii) wealth effects $u_c \cdot c_{sj}$.

straightforward to solve the planner's problem (see Online Appendix C). Note that given progressivity τ, $\lambda(\tau)$ is determined to satisfy the budget constraint of the government.

In this example, the progressivity of the income tax affects welfare only through the three channels—(i) workers' disutility of working hours, (ii) inefficient production due to the distorted working hours, and (iii) relaxation of the NDC. This is because under the special preferences, the extensive elasticity with respect to consumption is simply β, and does not depend on the progressivity. In Online Appendix C, we show that with general preferences, the progressivity can affect welfare through an additional channel—change in extensive elasticities.

We define the optimal markdown and the optimal pseudo markdown by

$$\mu_j = \frac{w_{sj}}{MP_{sj}}, \quad \text{and} \quad \tilde{\mu}_j = \frac{c_{s_j}}{MP_{sj}^{1-\tau}},$$

respectively. Note that these markdowns have the following relation: $\tilde{\mu}_j = \lambda \times \mu_j^{1-\tau}$. The optimal pseudo markdown should satisfy (see Online Appendix C for the derivation):

$$\tilde{\mu}_j = A_j + \tilde{\mu}_j^{net} = A_j + s(\alpha_j) \cdot \frac{\mathbf{\Delta}_h}{MP_{hj}^{1-\tau}} + (1 - s(\alpha_j)) \cdot \frac{\mathbf{\Delta}_\ell}{MP_{\ell j}^{1-\tau}},$$

where $A_j = h(\tau) \cdot \frac{\beta}{1+\beta} \cdot \frac{y_j}{MP_{hj}^{1-\tau} \cdot L_{hj} + MP_{\ell j}^{1-\tau} \cdot L_{\ell j}}$ and $s(\alpha_j) = \frac{MP_{hj}^{1-\tau} \cdot L_{hj}}{MP_{hj}^{1-\tau} \cdot L_{hj} + MP_{\ell j}^{1-\tau} \cdot L_{\ell j}}$. Then we can decompose the slope of the optimal pseudo markdown $\tilde{\mu}_j = \tilde{\mu}(z_j)$ as in the baseline:

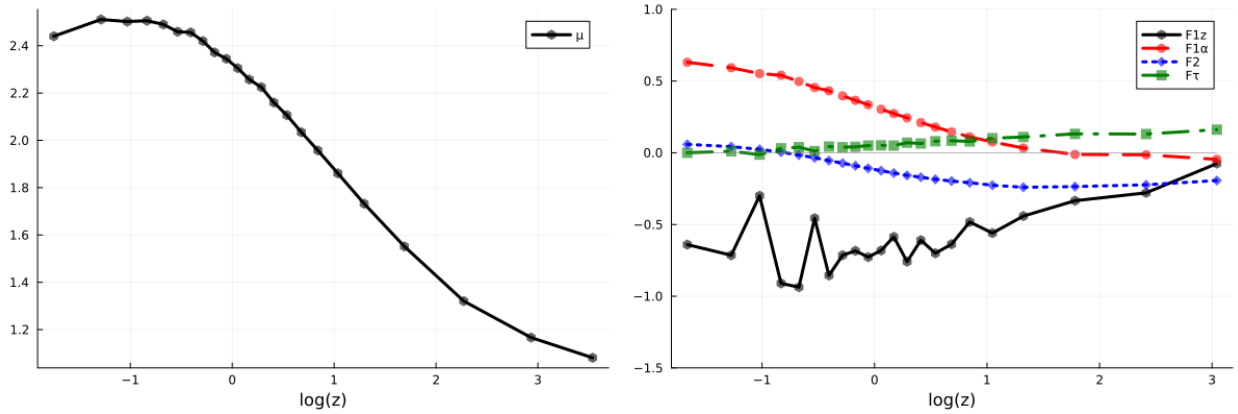
$$\frac{d\tilde{\mu}_j}{dz_j} = F_\tau + \underbrace{F_{1z} + F_{1_\alpha}}_{=F1} + F2,$$

with a new term $F_\tau = \frac{dA_j}{dz_j}$, which represents the direct substitution effect—progressivity of income tax τ replaces the wage compression of the markdown schedule.

As in the baseline with no labor income tax, the signs of F_{1z} and $F2$ are negative (under the positive sorting), and the signs of $F_{1\alpha}$ depends on the labor income share (see Online Appendix C). The key difference from the benchmark is the additional term F_τ, which is likely to be positive for most of the firms (except the firms with $L_{\ell j} << L_{hj}$), and this positive F_τ offsets the wage compression done by F_{1z} and $F2$. Since both theprogressive income tax and decreasing $\tilde{\mu}(z_j)$-schedule can reduce the consumption inequality, in the

presence of progressive income tax, the amount of redistribution provided by the markdown schedule is crowded out, which is the pure substitution effect.

Figures 7a and 7b plot the optimal pseudo markdown schedule and the decomposition of its slope under the current US labor income tax progressivity level $\tau = 0.18$. Although the planner has access to labor income tax, the planner still sets a moderately decreasing $\tilde{\mu}(z_j)$-schedule in most regions of the firm's productivity distribution to achieve redistribution among workers.[29] Figures 7a and 7b also confirm that the substitution effect F_τ makes the negative slope of the $\tilde{\mu}(z_j)$-schedule less steeper relative to that without labor income tax. At the left tail of the productivity distribution, the optimal $\tilde{\mu}(z_j)$-schedule still exhibits an inverted U-shape due to $F_{1\alpha}$ term.



(a) Markdown (b) Decomposition

Figure 7: $\tau = 0.18$

Finally, the first order condition with respect to progressivity yields

$$(N_h + N_l)\frac{\epsilon_h}{1+\epsilon_h} + \sum_j \phi_j \left(\frac{\log MP_{hj}}{MP_{hj}^{1-\tau}} c_{hj} - \frac{\log MP_{\ell j}}{MP_{\ell j}^{1-\tau}} c_{\ell j} \right) = \chi \cdot \frac{\epsilon_h}{1+\epsilon_h} \cdot \frac{1}{(1-\tau)} \sum_j y_j, \quad (27)$$

where the left-hand side captures the marginal benefit of increasing progressivity—reduction in disutility of working hours and relaxation of the NDC—and the right-hand side shows the marginal cost of increasing progressivity—reduction in production due to distorted working

[29] The planner sets 61% higher markdown to firms in the 10th productivity percentile compared to firms in the 90th productivity percentile.

hours. By solving for τ, we obtain the optimal progressivity of labor income tax:

$$1-\tau^* = \frac{\chi \cdot \frac{1}{N_h+N_\ell}\sum_j y_j}{1+\frac{1+\epsilon_h}{\epsilon_h}\cdot\frac{1}{N_h+N_\ell}\sum_j \phi_j\cdot\tilde{\mu}_j \log\left(\frac{MP_{hj}}{MP_{\ell j}}\right)}.$$

In our numerical example (with Frisch elasticity ϵ_h=0.5), the optimal progressivity is $\tau = 0.57$. In Online Appendix C, we provide further investigation of the joint optimal policy.

7 Extensions

7.1 Economy with Nonzero Market Share

So far, we have maintained the no-strategic interaction assumption that firms are small and labor market share of each firm is near zero, which makes the labor supply elasticity independent of market share. In this section, we relax this assumption and investigate how the optimal firm regulation is affected by this relaxation.

To focus on the role of market share in the equity-efficiency trade-off, we restrict our attention to the case with only one type of worker as in Section 3 and also assume constant amenity level $g_j = 1$ across firms. With these simplifying assumptions, the labor supply curve at firm j is given by $L_j(w_j) = \frac{w_j^\beta}{I} \times N$, where wage index I is $I = \sum_k w_k^\beta$. When firms are not small, the elasticity of labor supply is then calculated as follows:

$$\frac{d\log L_j}{d\log w_j} = \beta - \frac{d\log I}{d\log w_j} = \beta - \frac{d\log\left(\sum_k w_k^\beta\right)}{dw_j}\times w_j = \beta - \beta\times\frac{w_j^\beta}{I} = \beta\times(1-s_j),$$

where $s_j = \frac{L_j}{N}$ is labor market share of firm j. In the benchmark, we have assumed that firms are small with zero market share ($s_j = 0$ for all j), then the elasticity of labor supply is constant β for all firms because the effect of a firm's wage change on the total wage index I is ignored. When firms are not small, however, the effect of change in large firm's wage on the wage index I is not negligible, and the elasticity of labor supply $\beta(1-s_j)$ depends on the labor market share s_j. Higher market share implies lower elasticity. In an equilibrium without regulation, firms with non-zero market share set the wage and markdown as follows:

$$w_j^E = \frac{\beta(1-s_j)}{1+\beta(1-s_j)}\times MPL_j, \quad \text{and} \quad \mu_j^E = \frac{\beta(1-s_j)}{1+\beta(1-s_j)}.$$

What is the implication of the nonzero market share on the optimal allocation? One might conjecture that lower elasticity of labor supply of larger firms imply more redistribution with lower efficiency cost. We show that the nonzero market share leads to the opposite implication—more subsidy to larger firms.

Consider the constrained planner's problem when firms are not small. Note that in the planning problem in Definition 3 in Section 3, the planner already chooses the distribution of wage $\{w_k\}$ taking into account that increasing one firm's wage w_j changes not only firm j's labor supply L_j but also the whole distribution of $\{\frac{w_k^\beta}{I}\}$ (and thus the distribution of $\{L_k\}$) by changing the wage index $I = \sum_k w_k^\beta$. Thus even if we relax the small firm assumption, the planning problem in Definition 3 does not change.

This does *not* mean that the relaxation of the small firm assumption has no effect on the optimal allocation. Optimal wage and markdown schedule do depend on the size of market share. To see the role of market share more clearly, in the next proposition, we rewrite the optimal wage equation.[30]

Proposition 15. *The optimal wage (equivalently optimal consumption allocation) of workers at firm j is given by*

$$\begin{aligned} w_j^{SP} &= \frac{\tilde{\beta}_j}{1+\tilde{\beta}_j} \times MPL_j + \Delta_j \qquad (28) \\ where \quad & \tilde{\beta}_j = \beta \cdot (1 - s_j), \quad and \\ & \Delta_j = \frac{\frac{1}{\lambda} - \beta \cdot \sum_{k \neq j} (MPL_k - w_k) \cdot s_k}{1 + \beta(1 - s_j)}. \end{aligned}$$

Proposition 15 explicitly shows the role of the nonzero market share in optimal wage schedule. (See Appendix A for the proof) Equation (28) shows that there are two channels through which a market share s_j affects optimal wage schedule: (i) elasticity channel ($\tilde{\beta}_j$), and (ii) redistribution channel (Δ_j). First, higher market share s_j implies lower elasticity of labor supply, which lowers the efficiency cost of wage compression, and thus provides force for decreasing w_j for a firm with high market share (stronger wage compression). On the other hand, planner assigns a higher transfer Δ_j for a firm with higher market share s_j,

[30]The optimal consumption formula (4) of the benchmark also applies here. This formula can be rewritten as (28) in Proposition 15.

providing the opposite force on optimal w_j—increasing wage w_j for a firm with higher share s_j (weaker wage compression). Depending on which channel dominates, the nonzero market share can imply either more or less wage compression relative to the economy with only small firms.[31]

More important, the nonzero market share has a clear implication on the optimal subsidy schedule in the decentralization. We show that regardless of the dominance between the two channels just mentioned, higher market share of firm j unambiguously implies more subsidy to that firm. With the revenue-dependent subsidy policy $S_j(y_j)$, the owner of the firm j decides w_j according to

$$MPL_j \times \beta(1-s_j) \times (1+S_j'(y_j)) = w_j \times (1+\beta(1-s_j)).$$

Thus the optimal subsidy rate which implements the optimal allocation should satisfy

$$S_j'(y_j) = \frac{\frac{1}{\lambda} - \beta \sum_{k \neq j}(MPL_k - w_k)s_k}{\beta(1-s_j)} \times \frac{1}{MPL_j},$$

which shows that higher market share s_j increases the subsidy rate $S_j'(y_j)$. This is because both the firms in an equilibrium and the planner face the same labor supply elasticities. That is, a higher market share implies lower elasticity of labor supply for both firm and planner. Thus higher elasticity due to the non-zero market share itself does not create additional wedges. As the subsidy corrects the wedge between the firm and the planner, the weaker wage compression motive of the planner in the economy with nonzero market share implies more subsidy rate for the larger firms.

[31]When the second redistribution channel dominates, the multiplier ν associated with the labor market clearing constraint is positive. From the first order condition of the planner, the optimal markdown is given by

$$\mu_j^{SP} = \frac{\beta}{1+\beta} + \frac{1}{\lambda(1+\beta)} \cdot \frac{1}{MPL_j} - \frac{\beta\nu\frac{I}{N}}{\lambda(1+\beta)} \cdot \frac{1}{MPL_j},$$

and the last term associated with positive multiplier ν introduces a force for higher markdown for a firm with higher market share (and higher MPL_j) relative to a firm with smaller market share. As firms with higher market share sets lower wage with lower elasticity, the economy exhibits reduced wage inequality, which makes the planner's wage compression motives weaker.

7.2 Economy with both Product and Labor Market Power

We extend our results to a model where firms have product market power. We first consider the economy where there is only one type of worker as in Section 3. To analyze the impact of introducing product market power we modify the environment as follows. Entrepreneurs produce intermediate goods ($\{y_j\}$) which are then demanded by final good producers. The final good producers use the following technology to aggregate intermediate goods

$$y = \left(\sum_j y_j^{\frac{\theta-1}{\theta}} \right)^{\frac{\theta}{\theta-1}},$$

where $\theta < \infty$ is elasticity of substitution between different intermediate goods. Final good producers are competitive thus take prices of intermediate goods $\{p_j\}$ as given. The product demand curve faced by entrepreneur j is derived from the profit maximization decisions of final good producers

$$p_j = p(y_j) = \left(\frac{y_j}{y} \right)^{-\frac{1}{\theta}},$$

where we've normalized the price of final good to 1.

We first present the equilibrium outcome without regulation. Entrepreneurs facing downward sloping demand curves no longer take the prices of their products p_j as given. The problem of the entrepreneur at firm j is given by

$$\begin{aligned} max_{y_j, L_j, w_j} \quad & p_j y_j - w_j L_j \\ s.t. \quad & y_j = z_j L_j, \quad p_j = p(y_j) = \left(\frac{y_j}{y} \right)^{-\frac{1}{\theta}}, \quad \text{and} \quad L_j = L(w_j; g_j) = N \times \left(\frac{w_j g_j}{I} \right)^{\beta}, \end{aligned}$$

where we assume that the number of firms J is large enough such that entrepreneurs take y and I as given. Solving the entrepreneur problem yields

$$\underbrace{\underbrace{w_j}_{\text{W}} \times \left(1 + \frac{1}{\beta}\right)}_{\text{MCL}} = \underbrace{\underbrace{p_j z_j}_{\text{P} \times \text{MPL}} \times \left(1 - \frac{1}{\theta}\right)}_{\text{MRPL}}. \tag{29}$$

When making hiring decisions, the entrepreneur employs workers up to the point where marginal costs and marginal revenue products of labor are equalized. As in the baseline

model, there is a wedge between the wage and MCL, which is captured by $1 + \frac{1}{\beta}$ since entrepreneurs face the upward sloping labor supply curve. In contrast to the case without product market power, however, P×MPL is no longer equal to marginal revenue products of labor. Entrepreneurs should charge lower prices to attract sufficient demand for their products, thus MRPL is scaled down relative to P×MPL by $1 - \frac{1}{\theta}$.

By rearranging the first order condition (29), we obtain the double-markdown expression for wages as in Kroft, Luo, Mogstad, Setzler (2023):

$$\mu_j^E = \frac{w_j}{p_j z_j} = \frac{1 - \frac{1}{\theta}}{1 + \frac{1}{\beta}}.$$

The entrepreneur's markdown (μ_j^E) of wages relative to P×MPL is equal to $\frac{1-\frac{1}{\theta}}{1+\frac{1}{\beta}}$, where $1-\frac{1}{\theta}$ is inverse markup and $1 + \frac{1}{\beta}$ is the standard markdown.

We now investigate the planner's problem with both downward-sloping product demand and upward-sloping labor supply. As in the baseline case, the planner does not have a direct control for the worker's decision regarding where to work, thus she faces the same upward-sloping labor supply as the firm does. On the other hand, in assigning each firm's production, the planner is *not* constrained by the downward-sloping product demand. The planner who has an access to revenue-dependent subsidies can directly regulate production decision of firms by introducing a wedge between MRPL and MCL. The only change in the planner's problem relative to the case without product market power is then the final good production technology, which makes the firm's intermediate goods imperfectly substitutable ($\theta < \infty$) in producing the final consumption good.

The planner's problem is given by

$$\begin{aligned} max_{\{c_j\},\{c_{fj}\},\{L_j\}} \quad & \frac{N}{\beta} \log \left(\sum_j c_j^\beta g_j^\beta \right) + \psi_f \sum_j \log c_{fj} \\ s.t. \quad (\lambda) \quad & \sum_j c_j L_j + \sum_j c_{fj} = \left(\sum_j (z_j L_j)^{\frac{\theta-1}{\theta}} \right)^{\frac{\theta}{\theta-1}}, \\ (\nu) \quad & I = \left(\sum_j (c_j g_j)^\beta \right)^{\frac{1}{\beta}}, \\ & L_j = N \left(\frac{c_j g_j}{I} \right)^\beta. \end{aligned} \tag{30}$$

From the first order conditions of the planner's problem, we obtain the optimality condition for $w_j = c_j$ (see Online Appendix D):

$$\underbrace{w_j \times \left(1 + \frac{1}{\beta}\right)}_{\text{MCL}} = \underbrace{\left(\frac{y_j}{y}\right)^{-\frac{1}{\theta}} z_j + \frac{1}{\beta} \times \frac{1 - \nu \frac{I}{N}}{\lambda}}_{\text{MBL}}. \tag{31}$$

The planner equates marginal costs and marginal social benefits of labor. Since the planner is unable to directly control the worker's decision of where to work, the planner faces the same upward sloping labor supply curve, thus MCL is given by wage w_j times $1 + \frac{1}{\beta}$. Importantly, the planner calculates the margianl social benefits of labor by taking into account its value of redistribution. It is given by the sum of two terms: The first term, $\left(\frac{y_j}{y}\right)^{-\frac{1}{\theta}} z_j$, captures how an increase in labor at firm j directly raises the production of final good.[32] The second term, $\frac{1}{\beta} \times \frac{1-\nu\frac{I}{N}}{\lambda}$, reflects how the planner with redistributive concerns values the associated welfare gains.

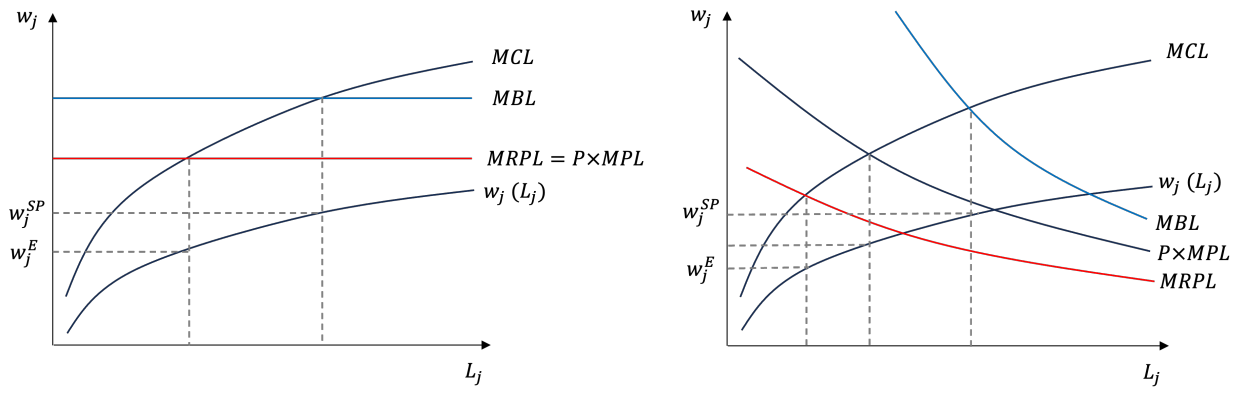
By comparing optimality conditions of the entrepreneur (29) and planner (31), we can show the wedge between the equilibrium markdown μ_j^E and the optimal markdown μ_j^{SP}:

$$\mu_j^{SP} = \mu_j^E \times \frac{1 \; + \; \frac{1}{\left(\frac{y_j}{y}\right)^{-\frac{1}{\theta}} z_j} \times \frac{1}{\beta} \times \left(1 - \nu \frac{I}{N}\right)}{1 - \frac{1}{\theta}}. \tag{32}$$

Relative to the equilibrium markdown μ_j^E, the planner's optimal markdown is adjusted by two factors. First, the planner corrects the inverse markup $1 - \frac{1}{\theta}$, which appears in the denominator of the right-hand side of (32). In the presence of product market power, the entrepreneurs facing the downward sloping demand curve exploit the effect of quantity on prices when making production decisions while the planner does not. The second factor is the redistributive subsidization $\frac{1}{\left(\frac{y_j}{y}\right)^{-\frac{1}{\theta}} z_j} \times \frac{1}{\beta} \times \left(1 - \nu \frac{I}{N}\right)$ appearing in the numerator of (32), which is essentially the same factor as in the benchmark economy. It corrects the wedge between P×MPL and the planner's MBL. The planner provides less subsidization for firms with higher productivity, which could be seen by noting that $\left(\frac{y_j}{y}\right)^{-\frac{1}{\theta}} z_j$ is decreasing in productivity z_j. The planner with redistributive concerns achieves wage compression relative to the equilibrium wage distribution.[33]

[32] Note that in equilibrium, $p_j = \left(\frac{y_j}{y}\right)^{-\frac{1}{\theta}}$, so the term $\left(\frac{y_j}{y}\right)^{-\frac{1}{\theta}} z_j$ corresponds to P×MPL.

[33] In Online Appendix D, we also discuss how the degree of wage compression respond to the introduction



(a) Without Product Market Power

(b) With Product Market Power

Figure 8: Equilibrium and Optimal Allocation

We graphically illustrate the planner's optimal allocation relative to the equilibrium in Figure 8. Figure 8a depicts the case where entrepreneurs do not have any product market power ($\theta = \infty$). Both the planner and entrepreneur face the same upward-sloping labor supply curve $w_j(L_j)$ and the corresponding MCL curve. The planner's MBL curve is above the P×MPL curve since the planner additionally takes into account the value of redistribution. The MRPL curve is equal to P×MPL because entrepreneurs take prices of their products as given. The planner provides redistributive subsidization up to the point where MBL equals MCL. On the other hand, Figure 8b shows that when firms have product market power ($\theta < \infty$), MRPL curve is scaled down relative to P×MPL. Thus the planner should first correct the inverse markup to raise the employment level up to the point where P×MPL is equal to MCL, then she provides redistributive subsidization by equating MCL to MBL.

The optimal allocation can be implemented with the revenue-dependent subsidies to firms. Given the subsidy schedule $S_j(py)$, the entrepreneur with productivity z_j solves the problem

$$max_{y_j, L_j, w_j} \quad p_j y_j + S_j(p_j y_j) - w_j L_j$$
$$s.t. \quad y_j = z_j L_j, \quad p_j = p(y_j) = \left(\frac{y_j}{y}\right)^{-\frac{1}{\theta}}, \quad L_j = L(w_j; g_j) = N \times \left(\frac{w_j g_j}{I}\right)^{\beta}.$$

of the product market power.

The subsidy schedule which implements the planner's allocation should satisfy

$$1 + S_j'(p_j y_j) = \frac{\mu_j^{SP}}{\mu_j^E} = \frac{1 \; + \; \frac{1}{\left(\frac{y_j}{y}\right)^{-\frac{1}{\theta}} z_j} \times \frac{1}{\beta} \times \left(1 - \nu \frac{I}{N}\right)}{1 - \frac{1}{\theta}}. \tag{33}$$

We can see that the two adjustment factors in the planner's markdown discussed above exactly determine the optimal subsidy schedule. Introducing the product market power amplifies the subsidy rate without changing the direction of the optimal regulation without product market.

We briefly discuss how the above results should be modified when we allow multiple types of workers. In the equilibrium without regulation entrepreneurs equate MRPL and MCL for each skill type of workers, yielding the following condition $\mu_j^E = \mu_{sj}^E = \frac{w_{sj}}{p_j \times MPL_{sj}} = \frac{1-\frac{1}{\theta}}{1+\frac{1}{\beta}}, \quad s = h, \ell$. The planner who cannot use the skill-dependent policy faces the same no-differentiation constraints (see Online Appendix D): $\frac{w_{hj}}{MPL_{hj}} = \frac{w_{\ell j}}{MPL_{\ell j}}$, and by solving the planner's problem and comparing it to the equilibrium, we obtain

$$\frac{\mu_j^{SP}}{\mu_j^E} = \frac{1}{1 - \frac{1}{\theta}} \times \left(1 + s_j(\alpha_j) \cdot \frac{\Delta_h \cdot \frac{1+\beta}{\beta}}{p_j \cdot MPL_{hj}} + (1 - s_j(\alpha_j)) \cdot \frac{\Delta_\ell \cdot \frac{1+\beta}{\beta}}{p_j \cdot MPL_{\ell j}}\right). \tag{34}$$

This again confirms that the introduction of product market power amplifies the subsidy rate without changing the direction of the original intervention driven by redistribution across firms and across skill types.

8 Conclusion

We have analyzed how to regulate firms in an economy where firm has a monopsonistic wage setting power. In this economy, part of the wage inequality is caused by the firm's effects, which generates motive of redistribution of workers. We have the following key findings. First, the optimal firm subsidy schedule features wage compression across firms, but the degree of redistribution depends on the underlying mechanism of the wage inequality. When the positive correlation between firm's productivity and amenity is the main reason for small firm's effects in the wage variation, the optimal regulation policy can provide a big redistribution at the low efficiency cost. Second, when the planner cannot use skill

dependent policy, optimal markdown schedule across firms should reflect the redistribution of both (i) across firms and (ii) between skill types. Third, the shape of optimal markdown schedule depends on the sources of sorting between firm and workers. When the positive sorting is mainly driven by the complementarity between firm's productivity and worker's skill, the optimal markdown schedule could be inversely U-shaped at the low region of firm's productivity distribution.

References

Abowd, J. A., and T. Lemieux (1993): "The Effects of Product Market Competition on Collective Bargaining Agreements: The Case of Foreign Competition in Canada*," *The Quarterly Journal of Economics*, 108(4), 983–1014.

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999): "High Wage Workers and High Wage Firms," *Econometrica*, 67(2), 251–333.

Ahlfeldt, G. M., D. Roth, and T. Seidel (2022): "Optimal minimum wages," .

Ales, L., and C. Sleet (2022): "Optimal Taxation of Income-Generating Choice," *Econometrica*, 90(5), 2397–2436.

Berger, D., K. Herkenhoff, and S. Mongey (2022a): "Labor Market Power," *American Economic Review*, 112(4), 1147–1193.

Berger, D., K. Herkenhoff, S. Mongey, and N. Mousavi (2024): "Monopsony Amplifies Distortions from Progressive Taxes," *AEA Papers and Proceedings*, 114, 555–560.

Berger, D., K. Herkenhoff, and A. Mongey, Simon Oppenheimer (2024): "Labor Market Power, Tax Progressivity and Inequality," *Working Paper*.

Berger, D. W., K. F. Herkenhoff, and S. Mongey (2022b): "Minimum wages, efficiency and welfare," Discussion paper.

Boar, C., and V. Midrigan (2019): "Markups and inequality," Discussion paper.

Bonhomme, S., K. Holzheu, T. Lamadon, E. Manresa, M. Mogstad, and B. Setzler (2022): "How Much Should We Trust Estimates of Firm Effects and Worker Sorting?," *Journal of Labor Economics*, 41(2), 291–322.

Bonhomme, S., T. Lamadon, and E. Manresa (2019): "A Distributional Framework for Matched Employer Employee Data," *Econometrica*, 87(3), 699–739.

Brown, C., and J. Medoff (1989): "The Employer Size-Wage Effect," *Journal of Political Economy*, 97(5), 1027–1059.

Cahuc, P., and G. Laroque (2014): "Optimal Taxation and Monopsonistic Labor Market: Does Monopsony Justify the Minimum Wage?," *Journal of Public Economic Theory*, 16(2), 259–273.

Card, D., A. R. Cardoso, J. Heining, and P. Kline (2017): "Firms and Labor Market Inequality: Evidence and Some Theory," *Journal of Labor Economics*, 36(S1), S13–S70.

Card, D., J. Heining, and P. Kline (2013): "Workplace Heterogeneity and the Rise of West German Wage Inequality*," *The Quarterly Journal of Economics*, 128(3), 967–1015.

da Costa, C. E., and L. J. Maestri (2019): "Optimal Mirrleesian taxation in non-competitive labor markets," *Economic Theory*, 68(4), 845–886.

Edmond, C., V. Midrigan, and D. Y. Xu (2022): "How Costly Are Markups?," *Journal of Political Economy*, 131(7), 1619–1675.

Heathcote, J., K. Storesletten, and G. L. Violante (2017): "Optimal Tax Progressivity: An Analytical Framework*," *The Quarterly Journal of Economics*, 132(4), 1693–1754.

Hurst, E., P. J. Kehoe, E. Pastorino, and T. Winberry (2022): "The distributional impact of the minimum wage in the short and long run," Discussion paper.

Kline, P., R. Saggio, and M. Sølvsten (2020): "Leave-Out Estimation of Variance Components," *Econometrica*, 88(5), 1859–1898.

Lamadon, T., J. Lise, C. Meghir, and J.-M. Robin (2024): "Labor market matching, wages, and amenities," .

Lamadon, T., M. Mogstad, and B. Setzler (2022): "Imperfect Competition, Compensating Differentials, and Rent Sharing in the US Labor Market," *American Economic Review*, 112(1), 169–212.

Manning, A. (2013): *Monopsony in motion: Imperfect competition in labor markets*. Princeton University Press.

Mongey, S., and M. E. Waugh (2024): "Discrete choice, complete markets, and equilibrium," .

Oi, W. Y., and T. L. B. T. H. o. L. E. Idson (1999): "Chapter 33 Firm size and wages," vol. 3, pp. 2165–2214. Elsevier.

Song, J., D. J. Price, F. Guvenen, N. Bloom, and T. von Wachter (2019): "Firming Up Inequality*," *The Quarterly Journal of Economics*, 134(1), 1–50.

Van Reenen, J. (1996): "The Creation and Capture of Rents: Wages and Innovation in a Panel of U. K. Companies*," *The Quarterly Journal of Economics*, 111(1), 195–226.

Appendix

A Proofs of Propositions in the Main Text

Proof for Lemma 2

From the firm's optimality condition (4), we obtain $\frac{d\log w_{hj}}{dz_j} = \frac{d\log MR_{hj}}{dz_j}$, $\frac{d\log w_{\ell j}}{dz_j} = \frac{d\log MR_{\ell j}}{dz_j}$, and the derivatives of marginal revenue products are

$$\frac{d\log MR_{hj}}{dz_j} = \frac{d\log z_j}{dz_j} + \frac{d\log \alpha_j}{dz_j} - \frac{1}{\sigma}\frac{d\log \frac{L_{hj}}{\bar{y}_j}}{dz_j},$$
$$\frac{d\log MR_{\ell j}}{dz_j} = \frac{d\log z_j}{dz_j} + \frac{d\log (1-\alpha_j)}{dz_j} - \frac{1}{\sigma}\frac{d\log \frac{L_{\ell j}}{\bar{y}_j}}{dz_j}.$$

Combining these yields

$$\frac{d\log \frac{w_{hj}}{w_{lj}}}{dj} = \frac{d\log MR_{hj}}{dz_j} - \frac{d\log MR_{\ell j}}{dz_j} = \frac{d\log\left(\frac{\alpha_j}{1-\alpha_j}\right)}{dz_j} - \frac{1}{\sigma}\cdot\frac{d\log\left(\frac{L_{hj}}{L_{\ell j}}\right)}{dz_j}. \tag{35}$$

From the labor supply equation (**??**), $\frac{d\log L_{hj}}{dz_j} = \beta\frac{d\log w_{hj}}{dz_j} + \beta\frac{d\log g_{hj}}{dz_j}$, $\frac{d\log L_{\ell j}}{dz_j} = \beta\frac{d\log w_{\ell j}}{dz_j} + \beta\frac{d\log g_{\ell j}}{dz_j}$, and thus

$$\frac{d\log \frac{L_{hj}}{L_{lj}}}{dj} = \beta\cdot\frac{d\log\left(\frac{w_{hj}}{w_{\ell j}}\right)}{dz_j} + \beta\cdot\frac{d\log\left(\frac{g_{hj}}{g_{\ell j}}\right)}{dz_j}. \tag{36}$$

Combining (35) and (36) yields

$$\left(1+\frac{\beta}{\sigma}\right)\cdot\frac{d\log\left(\frac{L_{hj}}{L_{\ell j}}\right)}{dz_j} = \beta\cdot\frac{d\log\frac{\alpha_j}{1-\alpha_j}}{dz_j} + \beta\cdot\frac{\beta}{\sigma}\frac{d\log\frac{g_{hj}}{g_{\ell j}}}{dz_j}, \tag{37}$$

The conclusion of Lemma 2 follows from (37).

Proof for Proposition 4

We solve the the planner's problem in Definition 3 to obtain the optimal consumption allocation. The Lagrangian of this planner's problem is

$$\mathscr{L} = \frac{N}{\beta}\left[\log\left(\sum_j (c_j g_j)^\beta\right)\right] + \psi_f\sum_j \log c_{fj} + \lambda(\sum_j z_j L_j - \sum_j (c_j L_j + c_{fj})) + \nu(I - \sum_j (c_j g_j)^\beta).$$

First order condition with respect to c_j is then

$$\frac{L_j}{c_j} + \lambda((z_j - c_j)\beta\frac{L_j}{c_j} - L_j) - \nu\beta\frac{I}{N}\frac{L_j}{c_j} = 0 \tag{38}$$

Multiplying (38) by $\frac{c_j}{L_j}$ yields

$$1 + \lambda(\beta(z_j - c_j) - c_j) - \nu\beta\frac{I}{N} = 0,$$

which can be rewritten as the optimal consumption formula:

$$c_j = \frac{\beta}{1+\beta}z_j + \Delta, \quad \Delta = \frac{1}{\lambda(1+\beta)}\left(1 - \beta\nu\frac{I}{N}\right). \tag{39}$$

We next derive the expressions for the Lagrange multipliers. Multiplying (39) by L_j and summing up over j, we have

$$\sum_j c_j L_j = \frac{\beta}{1+\beta}\sum_j z_j L_j + \frac{N}{\lambda(1+\beta)}\left(1 - \beta\nu\frac{I}{N}\right). \tag{40}$$

First order condition of the planner's problem with respect to I is

$$\nu = \lambda\frac{\sum_j (z_j - c_j)L_j}{I} \tag{41}$$

Substituting (41) into (40) yields

$$\sum_j c_j L_j = \frac{\beta}{1+\beta}\sum_j z_j L_j - \frac{\beta}{1+\beta}\sum_j (z_j - c_j)L_j + \frac{N}{\lambda(1+\beta)},$$

and by rearranging the terms, we obtain

$$\frac{1}{\lambda} = \frac{\sum_j c_j L_j}{N}.$$

First order condition of the planner's problem with respect to c_{fj} is

$$\frac{1}{\lambda} = \frac{c_f}{\psi_f}. \tag{42}$$

Combining (42) with the resource constraint $\sum_j c_j L_j + \sum_j c_{fj} = \sum_j z_j L_j$ yields

$$\frac{1}{\lambda} = \frac{1}{N + \psi_f J}\sum_j z_j L_j. \tag{43}$$

In sum,

$$\frac{1}{\lambda} = \frac{c_f}{\psi_f} = \frac{\sum_j c_j L_j}{N} = \frac{1}{N + \psi_f J}\sum_j z_j L_j$$

Thus, $\sum_j z_j L_j - \sum_j c_j L_j = \frac{\psi_f J}{N + \psi_f J}\sum_j z_j L_j$, which is substituted into (41), and also by substituting (43) into (41), we obtain

$$\nu = \frac{N + \psi_f J}{\sum_j z_j L_j} \cdot \frac{\frac{\psi_f J}{N + \psi_f J}\sum_j z_j L_j}{I} = \frac{\psi_f J}{I},$$

and thus $\nu\frac{I}{N} = \frac{\psi_f J}{N}$.

Proof for Lemma 5

At constrained optimum with $J = 2$, the following equations hold:

$$(z_2 - z_1)\frac{\beta}{1+\beta} = c_2 - c_1 \tag{44}$$

$$\frac{N}{N+\psi_f J}(z_2 L_2 + z_1 L_1) = c_2 L_2 + c_1 L_1 \tag{45}$$

$$L_2 = N\frac{c_2^\beta}{c_2^\beta + c_1^\beta}, \quad L_1 = N\frac{c_1^\beta}{c_2^\beta + c_1^\beta}. \tag{46}$$

Define $\gamma_c = \frac{c_2}{c_1}$, $\gamma_z = \frac{z_2}{z_1}$. Then (46) can be rewritten as

$$L_2 = N\frac{\gamma_c^\beta}{\gamma_c^\beta + 1}, \quad L_1 = N\frac{1}{\gamma_c^\beta + 1}. \tag{47}$$

Dividing the LHS and RHS of (44) with the LHS and RHS of (45) yields

$$\left(1 + \frac{\psi_f J}{N}\right)\frac{(z_2 - z_1)\frac{\beta}{1+\beta}}{z_2 L_2 + z_1 L_1} = \frac{c_2 - c_1}{c_2 L_2 + c_1 L_1}. \tag{48}$$

From (48), dividing numerator and denominator of the LHS by z_1 and dividing numerator and denominator of the RHS by c_1 yield

$$\left(1 + \frac{\psi_f J}{N}\right)\frac{\beta}{1+\beta}\frac{\gamma_z - 1}{\frac{\gamma_z \gamma_c^\beta}{\gamma_c^\beta + 1} + \frac{1}{\gamma_c^\beta + 1}} = \frac{\gamma_c - 1}{\frac{\gamma_c^{\beta+1}}{\gamma_c^\beta + 1} + \frac{1}{\gamma_c^\beta + 1}}. \tag{49}$$

Rearranging (49)

$$\left(1 + \frac{\psi_f J}{N}\right)\frac{\beta}{1+\beta}\frac{\gamma_z - 1}{\gamma_z \gamma_w^\beta + 1} = \frac{\gamma_w - 1}{\gamma_w^{\beta+1} + 1}. \tag{50}$$

From (50), solving for γ_z yields

$$\gamma_z = \frac{(\beta N + \beta J\psi_f)\gamma_w^{\beta+1} + (1+\beta)N\gamma_w + (\beta J\psi_f - N)}{(\beta J\psi_f - N)\gamma_w^{\beta+1} + (1+\beta)N\gamma_w^\beta + (\beta N + \beta J\psi_f)} \tag{51}$$

At $\gamma_c = 1, \gamma_z = 1$, (51) is satisfied. The slope of the RHS of (51) at $\gamma_c = 1$ is $\frac{1+\beta}{\beta}\frac{N}{N+J\psi_f}$. Using approximation around $(1, 1)$

$$(\gamma_z - 1) \approx \frac{1+\beta}{\beta}\left(1 + \frac{J\psi_f}{N}\right)(\gamma_c - 1)$$

$$\Leftrightarrow \gamma_c = \frac{\beta}{1+\beta}\frac{N + J\psi_f}{N}(\gamma_z - 1) + 1$$

For given $\gamma_z > 1$, $\gamma_c > 1$ is increasing wrt β. also increasing wrt ψ_f. From (47), it is straightforward to see that L_l decreases with β,ψ_f, and L_h increases with β,ψ_f.

Finally, since $\gamma_c = \frac{\frac{\beta}{1+\beta}z_2+\Delta}{\frac{\beta}{1+\beta}z_1+\Delta}$, we have

$$\frac{d\gamma_c}{d\Delta} = \frac{\frac{\beta}{1+\beta}(z_1 - z_2)}{\left(\frac{\beta}{1+\beta}z_1 + \Delta\right)^2} < 0$$

which implies that Δ decreases in β and ψ_f.

Proof for Proposition 6

From Proposition 4 we have

$$c_j = \frac{\beta}{1+\beta}z_j + \Delta \tag{52}$$

Dividing by z_j yields

$$\mu_j = \frac{c_j}{z_j} = \frac{\beta}{1+\beta} + \frac{\Delta}{z_j} \tag{53}$$

Differentiating by z_j yields the result.

Proof for Proposition 7

Using $L_j = N\frac{(c_j g_j)^\beta}{I}$, and $c_j = \frac{\beta}{1+\beta}z_j + \Delta$, it is straightforward to obtain

$$\frac{\partial L_j}{\partial z_j}\frac{z_j}{L_j} = \beta\left(\frac{\beta}{1+\beta}\cdot\frac{z_j}{\frac{\beta}{1+\beta}z_j + \Delta} + \frac{g'(z_j)z_j}{g_j}\right) \tag{54}$$

Also, note that by multiplying L_j to the consumption formula (39) and summing over j, we obtain $\sum_j c_j L_j = \frac{\beta}{1+\beta}\sum_j z_j L_j + \Delta \cdot N$, which implies that

$$\Delta = \frac{1}{N}\left[\sum_j c_j L_j - \frac{\beta}{1+\beta}\sum_j z_j L_j\right] = \frac{1}{1+\beta}\frac{\sum_j z_j L_j}{N}, \tag{55}$$

where he second equality uses the resource constraint with $c_{fj} = 0$ for all j.

We prove that if $\frac{g'(z_j)z_j}{g_j}$ increases then Δ increases. Suppose not. Then Δ decreases and (55) implies that $\frac{\partial L_j}{\partial z_j}\frac{z_j}{L_j}$ should also decrease for some firms. However, if $\frac{g'(z_j)z_j}{g_j}$ increases and Δ decreases then from (54) we observe that $\frac{\partial L_j}{\partial z_j}\frac{z_j}{L_j}$ always increases, resulting in contradiction.

Proof for Proposition 9 and Characterization of Lagrange multipliers.

Proposition 9 characterizes the optimal allocation which solves the planners problem in Definition 8. Below, we provide additional characterization of Lagrange multipliers and prove them together.

Proposition A.1. *The Lagrange multipliers defined in Proposition 9 is characterized as folllows. The Lagrange multiplier λ satisfies*

$$\frac{1}{\lambda} = \frac{c_f}{\psi_f} = \frac{\sum_j (c_{hj} L_{hj} + c_{\ell j} L_{\ell j})}{N_h + N_\ell} = \frac{1}{N_h + N_\ell + \psi_f J} \sum_j y_j, \tag{56}$$

where c_f is (common) consumption level of firm owners.
The Lagrange multiplier ν_h, ν_ℓ satisfies

$$\frac{\nu_h}{\lambda} I_h = \sum_j (MR_{hj} - c_{hj}) L_{hj} + \frac{1}{\sigma} \sum_j \frac{\zeta_j}{\lambda} \mu_j, \quad \frac{\nu_\ell}{\lambda} I_\ell = \sum_j (MR_{\ell j} - c_{\ell j}) L_{\ell j} - \frac{1}{\sigma} \sum_j \frac{\zeta_j}{\lambda} \mu_j \tag{57}$$

The Lagrange multiplier ζ_j satisfies

$$\frac{\zeta_j}{\lambda} \mu_j = \frac{\sigma(1+\beta)}{\sigma + \beta} \left(\frac{1}{MR_{hj} L_{hj}} + \frac{1}{MR_{\ell j} L_{\ell j}} \right)^{-1} \left(\frac{\Delta_\ell}{MR_{\ell j}} - \frac{\Delta_h}{MR_{hj}} \right) \tag{58}$$

Proof The Lagrangian of the planners problem in Definition 8 is given by

$$\begin{aligned} \mathscr{L} = & \frac{N_h}{\beta} \left[\log \left(\sum_j (c_{hj} g_{hj})^\beta \right) \right] + \frac{N_\ell}{\beta} \left[\log \left(\sum_j (c_{\ell j} g_{\ell j})^\beta \right) \right] + \psi_f \sum_j \log c_{fj} \\ & + \lambda (\sum_j z_j (\alpha_j L_{hj}^{\frac{\sigma-1}{\sigma}} + (1-\alpha_j) L_{\ell j}^{\frac{\sigma-1}{\sigma}})^{\frac{\sigma}{\sigma-1}} - \sum_j (c_{hj} L_{hj} + c_{\ell j} L_{\ell j} + c_{fj})) \\ & + \nu_h (I_h - \sum_j (c_{hj} g_{hj})^\beta) + \nu_{ell} (I_\ell - \sum_j (c_{\ell j} g_{\ell j})^\beta) + \sum_j \zeta_j (\frac{c_{hj}}{MR_{hj}} - \frac{c_{\ell j}}{MR_{\ell j}}), \end{aligned} \tag{59}$$

subject to the remaining labor supply constraints $L_{hj} = N_h \frac{(c_{hj} g_{hj})^\beta}{I_h}$ and $L_{\ell j} = N_\ell \frac{(c_{\ell j} g_{\ell j})^\beta}{I_\ell}$.

We take the first order conditions of the Lagrangian (59) with respect to c_{hj} and $c_{\ell j}$ subject to the labor supply constraints. Taking the first order condition with respecto c_{hj} and multiplying it by $\frac{c_{hj}}{L_{hj}}$ yields

$$1 - \lambda c_{hj} + \beta \left[\lambda MR_{hj} - \lambda c_{hj} - \zeta_j \frac{MR_{hhj}}{MR_{hj}^2} c_{hj} + \zeta_j \frac{MR_{\ell hj}}{MR_{\ell j}} \frac{c_{hj}}{MR_{hj}} \right] - \nu_h \beta \frac{I_h}{N_h} + \frac{\zeta_j}{MR_{hj}} \cdot \frac{c_{hj}}{L_{hj}} = 0.$$

Rearranging the equation for the consumption c_{hj} yields

$$\begin{aligned} c_{hj} &= \frac{\lambda \beta MR_{hj} + \left(1 - \nu_h \beta \frac{I_h}{N_h}\right)}{\lambda(1+\beta) + \zeta_j \left[\frac{1}{MR_{hj}} \left(\beta \frac{MR_{hhj}}{MR_h j} - \frac{1}{L_{hj}} - \beta \frac{MR_{\ell hj}}{MR_{\ell j}} \right) \right]} \\ &= \frac{\frac{\beta}{1+\beta} MR_{hj} + \frac{1 - \nu_h \beta \frac{I_h}{N_h}}{\lambda(1+\beta)}}{1 + \frac{\zeta_j}{\lambda(1+\beta)} \cdot \frac{1}{MR_{hj}} \cdot \left(\beta \frac{MR_{hhj}}{MR_h j} - \frac{1}{L_{hj}} - \beta \frac{MR_{\ell hj}}{MR_{\ell j}} \right)} \end{aligned}$$

Then we use $\frac{MR_{hhj}}{MR_{hj}} = \frac{1}{\sigma L_{hj}}\left(\tilde{y}_j^{\frac{1-\sigma}{\sigma}}\alpha_j L_{hj}^{-\frac{1-\sigma}{\sigma}} - 1\right)$ and $\frac{MR_{\ell hj}}{MR_{\ell j}} = \frac{1}{\sigma}\tilde{y}_j^{\frac{1-\sigma}{\sigma}}\alpha_j L_{hj}^{-\frac{1}{\sigma}}$ to obtain the consumption formula for c_{hj}. The consumption formula for $c_{\ell j}$ can be also derived by following the similar steps. In sum, we obtain

$$c_{hj} = \frac{\frac{\beta}{1+\beta}MR_{hj} + \frac{1}{\lambda(1+\beta)}\left(1-\nu_h\frac{I_h}{N_h}\right)}{1-\frac{\zeta_j}{\lambda(1+\beta)}\frac{1}{MR_{hj}L_{hj}}\left(1+\frac{\beta}{\sigma}\right)}, \quad \text{and} \quad c_{\ell j} = \frac{\frac{\beta}{1+\beta}MR_{\ell j} + \frac{1}{\lambda(1+\beta)}\left(1-\nu_\ell\frac{I_\ell}{N_\ell}\right)}{1+\frac{\zeta_j}{\lambda(1+\beta)}\frac{1}{MR_{\ell j}L_{\ell j}}\left(1+\frac{\beta}{\sigma}\right)}.$$

The formulas are then rewritten as

$$c_{hj} = \frac{1}{\Omega_{hj}}\cdot\left(\frac{\beta}{1+\beta}MR_{hj} + \Delta_h\right), \quad \text{and} \quad c_{\ell j} = \frac{1}{\Omega_{\ell j}}\left(\frac{\beta}{1+\beta}MR_{\ell j} + \Delta_\ell\right), \tag{60}$$

where

$$\Omega_{hj} = 1 - \frac{\zeta_j}{\lambda(1+\beta)}\frac{1}{MR_{hj}L_{hj}}\left(1+\frac{\beta}{\sigma}\right), \quad \Delta_h = \frac{1}{\lambda(1+\beta)}\left(1-\nu_h\frac{I_h}{N_h}\right),$$
$$\Omega_{\ell j} = 1 + \frac{\zeta_j}{\lambda(1+\beta)}\frac{1}{MR_{\ell j}L_{\ell j}}\left(1+\frac{\beta}{\sigma}\right), \quad \Delta_\ell = \frac{1}{\lambda(1+\beta)}\left(1-\nu_\ell\frac{I_\ell}{N_\ell}\right).$$

We multiply each formula in (60) by $\frac{\Omega_{hj}}{MR_{hj}}$ and $\frac{\Omega_{\ell j}}{MR_{\ell j}}$ respectively to obtain

$$\Omega_{hj}\frac{c_{hj}}{MR_{hj}} = \frac{\beta}{1+\beta} + \frac{\Delta_h}{MR_{hj}}, \quad \text{and} \quad \Omega_{\ell j}\frac{c_{\ell j}}{MR_{\ell j}} = \frac{\beta}{1+\beta} + \frac{\Delta_\ell}{MR_{\ell j}},$$

and using that $\mu_j = \frac{c_{hj}}{MR_{hj}} = \frac{c_{\ell j}}{MR_{\ell j}}$, we obtain

$$(\Omega_{hj} - \Omega_{\ell j})\mu_j = \frac{\Delta_h}{MR_{hj}} - \frac{\Delta_\ell}{MR_{\ell j}}.$$

Combining this with $\Omega_{hj} - \Omega_{\ell j} = -\frac{\zeta_j}{\lambda}\mu_j \cdot \frac{\sigma+\beta}{\sigma(1+\beta)}\cdot\left(\frac{1}{MR_{hj}L_{hj}} + \frac{1}{MR_{\ell j}L_{\ell j}}\right)$ yields the equation for ζ_j (58).

The first order conditions with respect to I_h, I_l are

$$\nu_h\frac{I_h}{N_h} = \lambda\frac{\sum_j(MR_{hj}-c_{hj})L_{hj}}{N_h} + \frac{1}{\sigma}\frac{\sum_j\zeta_j\mu_j}{N_h}, \quad \text{and} \quad \nu_l\frac{I_l}{N_l} = \lambda\frac{\sum_j(MR_{lj}-c_{lj})L_{lj}}{N_l} - \frac{1}{\sigma}\frac{\sum_j\zeta_j\mu_j}{N_l},$$

which yields (57).

The first order conditions with respect to c_{fj} is

$$\frac{1}{\lambda} = \frac{c_{fj}}{\psi_f}.$$

The remaining expressions for the Lagrange multiplier λ are proved below in the proof of Proposition 10.

■

Proof for Proposition 10

We can rewrite the optimal consumption formulas (60) using the definition of $\mu_j = \frac{c_{hj}}{MR_{hj}} = \frac{c_{\ell j}}{MR_{\ell j}}$

$$\Omega_{hj} \cdot \mu_j = \frac{\beta}{1+\beta} + \frac{\Delta_h}{MR_{hj}} \tag{61}$$

$$\Omega_{\ell j} \cdot \mu_j = \frac{\beta}{1+\beta} + \frac{\Delta_h}{MR_{\ell j}}. \tag{62}$$

Summing up $\alpha_j \cdot L_{hj}^{-\frac{1-\sigma}{\sigma}} \times$(61) and $(1-\alpha_j) \cdot L_{\ell j}^{-\frac{1-\sigma}{\sigma}} \times$(62) yields

$$\left[\alpha_j L_{hj}^{\frac{\sigma-1}{\sigma}} + (1-\alpha_j) L_{\ell j}^{\frac{\sigma-1}{\sigma}}\right] \cdot \mu_j = \left[\alpha_j L_{hj}^{\frac{\sigma-1}{\sigma}} + (1-\alpha_j) L_{\ell j}^{\frac{\sigma-1}{\sigma}}\right] \cdot \frac{\beta}{1+\beta} + \alpha_j L_{hj}^{\frac{\sigma-1}{\sigma}} \cdot \frac{\Delta_h}{MR_{hj}} + (1-\alpha_j) L_{\ell j}^{\frac{\sigma-1}{\sigma}} \cdot \frac{\Delta_\ell}{MR_{\ell j}}. \tag{63}$$

Then dividing the equation (63) by $\alpha_j L_{hj}^{\frac{\sigma-1}{\sigma}} + (1-\alpha_j) L_{\ell j}^{\frac{\sigma-1}{\sigma}}$ yields the markdown formula

$$\mu_j = \frac{1}{1+\beta} + S_{hj} \frac{\Delta_h}{MR_{hj}} + S_{\ell j} \frac{\Delta_\ell}{MR_{\ell j}}, \tag{64}$$

where $S_{hj} = \frac{\alpha_j L_{hj}^{\frac{\sigma-1}{\sigma}}}{\alpha_j L_{hj}^{\frac{\sigma-1}{\sigma}} + (1-\alpha_j) L_{\ell j}^{\frac{\sigma-1}{\sigma}}} \equiv S_j$ and $S_{\ell j} = 1 - S_j$, which proves the optimal markdown formula in Proposition 10.

We now complete the proof of Proposition A.1 by deriving the remaining expression of the Lagrange multiplier. Combining $\mu_j = \frac{c_{hj}}{MR_{hj}} = \frac{c_{\ell j}}{MR_{\ell j}}$ and (64) and multiplying MR_{hj} and $MR_{\ell j}$ respectively, c_{hj} and $c_{\ell j}$ can be rewritten as follows:

$$c_{hj} = \frac{\beta}{1+\beta} MR_{hj} + S_j \cdot \Delta_h + (1-S_j) \cdot \Delta_\ell \cdot \frac{\alpha_j}{1-\alpha_j} \cdot \left(\frac{L_{hj}}{L_{\ell j}}\right)^{-\frac{1}{\sigma}}, \tag{65}$$

$$c_{\ell j} = \frac{\beta}{1+\beta} MR_{\ell j} + S_j \cdot \Delta_h \cdot \frac{1-\alpha_j}{\alpha_j} \cdot \left(\frac{L_{hj}}{L_{\ell j}}\right)^{\frac{1}{\sigma}} + (1-S_j) \cdot \Delta_\ell. \tag{66}$$

Then summing up (65)$\times L_{hj}+$ (66)$\times L_{\ell j}$ over j yields

$$\sum_j c_{hj} L_{hj} + \sum_j c_{\ell j} L_{\ell j} = \frac{\beta}{1+\beta} \sum_j y_j + \Delta_h \sum_j L_{hj} + \Delta_\ell \sum_j L_{\ell j},$$

where we have used $MP_{hj} L_{hj} + MP_{\ell j} L_{\ell j} = y_j$, $S_j \cdot \left(L_{hj} + \frac{1-\alpha_j}{\alpha_j} \left(\frac{L_{hj}}{L_{\ell j}}\right)^{\frac{1}{\sigma}} L_{\ell j}\right) = L_{hj}$, and $(1-S_j) \cdot \left(\frac{\alpha_j}{1-\alpha_j} \left(\frac{L_{hj}}{L_{\ell j}}\right)^{-\frac{1}{\sigma}} L_{hj} + L_{\ell j}\right) = L_{\ell j}$. Then by rearranging the equation, we obtain

$$N_h \Delta_h + N_\ell \Delta_\ell = \sum_j c_{hj} L_{hj} + \sum_j c_{\ell j} L_{\ell j} - \frac{\beta}{1+\beta} \sum_j y_j. \tag{67}$$

On the other hand, in the consumption formulas (60), we have defined the common transfer terms by

$$\Delta_h - \frac{1}{\lambda(1+\beta)} \left(1 - \nu \frac{I_h}{N_h}\right), \quad \text{and} \quad \Delta_\ell - \frac{1}{\lambda(1+\beta)} \left(1 - \nu \frac{I_\ell}{N_\ell}\right).$$

Thus weighted sum of Δ_s's yields

$$N_h\Delta_h + N_\ell\Delta_\ell = \frac{1}{\lambda(1+\beta)} \times \left[N_h + N_\ell - \beta \cdot (\nu_h I_h + \nu_\ell I_\ell)\right]. \tag{68}$$

From (57), we obtain

$$\nu_h I_h + \nu_\ell I_\ell = \lambda \cdot \left(\sum_j y_j - \sum_j c_{hj} L_{hj} - \sum_j c_{\ell j} L_{\ell j}\right). \tag{69}$$

Combining (68) and (69) yields

$$N_h\Delta_h + N_\ell\Delta_\ell = \frac{N_h + N_\ell}{\lambda(1+\beta)} - \frac{\beta}{1+\beta}\sum_j y_j + \frac{\beta}{1+\beta} \cdot \left(\sum_j c_{hj} L_{hj} + \sum_j c_{\ell j} L_{\ell j}\right). \tag{70}$$

Then Combining (68) and (70) yields

$$\frac{1}{\lambda} = \frac{\sum_j (c_{hj} L_{hj} + c_{lj} L_{lj})}{N_h + N_l}$$

In addition, using the resource constraint, we obtain

$$\frac{1}{\lambda} = \frac{1}{N_h + N_l + \psi_f J}\sum_j y_j, \tag{71}$$

which completes the proof of Proposition A.1.

Proof for Proposition 11 and Proposition 12

We now characterize the slope of markdown $\mu_j = \mu(z_j)$ with respect to firm's productivity z_j. As we have described in Proposition 10, the slope can be decomposed into two terms $\frac{d\mu_j}{dz_j} = F1 + F2$, where

$$F1 = \left.\frac{d\mu_j}{dz_j}\right|_{\frac{L_{hj}}{L_{\ell j}}} \quad \text{and} \quad F2 = \frac{\partial \mu_j}{\partial\left(\frac{L_{hj}}{L_{\ell j}}\right)} \times \frac{d\left(\frac{L_{hj}}{L_{\ell j}}\right)}{dz_j}.$$

We first express $F1$—the derivative of $\mu(z_j)$ holding the employment share $\frac{L_{hj}}{L_{\ell j}}$ constant. The derivative of the markdown (64) with respect to z_j given $\frac{L_{hj}}{L_{\ell j}}$ can be written as the sum of the direct effect of change in z_j and the indirect effect through the change in α_j:

$$\begin{aligned}
F1 &= \left.\frac{d\mu_j}{dz_j}\right|_{\frac{L_{hj}}{L_{\ell j}}} = F_{1z} + F_{1\alpha}, \\
\text{where} \quad F_{1z} &= -\left[S_j \cdot \frac{\Delta_h}{MR_{hj}^2} \cdot \frac{\partial MR_{hj}}{\partial z_j} + (1 - S_j) \cdot \frac{\Delta_\ell}{MR_{\ell j}^2} \cdot \frac{\partial MR_{\ell j}}{\partial z_j}\right], \text{ and} \\
F_{1\alpha} &= \left[\begin{array}{l} \frac{\partial S_j}{\partial \alpha_j}\Big|_{\frac{L_{hj}}{L_{\ell j}}} \times \left(\frac{\Delta_h}{MR_{hj}} - \frac{\Delta_\ell}{MR_{\ell j}}\right) \\ -\left(S_j \cdot \frac{\Delta_h}{MR_{hj}^2} \left.\frac{\partial MR_{hj}}{\partial \alpha_j}\right|_{\frac{L_{hj}}{L_{\ell j}}} + (1 - S_j) \cdot \frac{\Delta_\ell}{MR_{\ell j}^2} \left.\frac{\partial MR_{\ell j}}{\partial \alpha_j}\right|_{\frac{L_{hj}}{L_{\ell j}}}\right) \end{array}\right] \times \frac{d\alpha_j}{dz_j}
\end{aligned} \tag{72}$$

The sign of F_{1z} is negative because $\frac{\partial MR_{hj}}{\partial z_j} = \alpha_j \left(\alpha_j + (1-\alpha_j)\left(\frac{L_{hj}}{L_{\ell j}}\right)^{-\frac{\sigma-1}{\sigma}}\right)^{\frac{1}{\sigma-1}} > 0$ and $\frac{\partial MR_{\ell j}}{\partial z_j} = (1-\alpha_j)\left(\alpha_j\left(\frac{L_{hj}}{L_{\ell j}}\right)^{\frac{\sigma-1}{\sigma}} + (1-\alpha_j)\right)^{\frac{1}{\sigma-1}} > 0$.

To derive the expression for $F_{1\alpha}$, we first take a derivative of weight S_j with respect to α_j:

$$\left.\frac{\partial S_j}{\partial \alpha_j}\right|_{\frac{L_{hj}}{L_{\ell j}}} = \frac{L_{hj}^{\frac{\sigma-1}{\sigma}}}{\left(\alpha_j L_{hj}^{\frac{\sigma-1}{\sigma}} + (1-\alpha_j)L_{\ell j}^{\frac{\sigma-1}{\sigma}}\right)} - \frac{\alpha_j L_{hj}^{\frac{\sigma-1}{\sigma}} \cdot \left(L_{hj}^{\frac{\sigma-1}{\sigma}} - L_{\ell j}^{\frac{\sigma-1}{\sigma}}\right)}{\left(\alpha_j L_{hj}^{\frac{\sigma-1}{\sigma}} + (1-\alpha_j)L_{\ell j}^{\frac{\sigma-1}{\sigma}}\right)^2}. \tag{73}$$

Next, we take derivatives of the marginal revenue products with respect to α_j. Using that $MR_{hj} = \alpha_j z_j \left(\alpha_j + (1-\alpha_j)\left(\frac{L_{hj}}{L_{\ell j}}\right)^{-\frac{\sigma-1}{\sigma}}\right)^{\frac{1}{\sigma-1}}$ and $MR_{\ell j} = (1-\alpha_j) z_j \left(\alpha_j\left(\frac{L_{hj}}{L_{\ell j}}\right)^{\frac{\sigma-1}{\sigma}} + 1 - \alpha_j\right)^{\frac{1}{\sigma-1}}$, we obtain

$$\left.\frac{\partial MR_{hj}}{\partial \alpha_j}\right|_{\frac{L_{hj}}{L_{\ell j}}} = MR_{hj} \cdot \left[\frac{1}{\alpha_j} + \frac{1}{\sigma-1} \cdot \frac{L_{hj}^{\frac{\sigma-1}{\sigma}}\left(1 - \left(\frac{L_{hj}}{L_{\ell j}}\right)^{-\frac{\sigma-1}{\sigma}}\right)}{\alpha_j L_{hj}^{\frac{\sigma-1}{\sigma}} + (1-\alpha_j)L_{\ell j}^{\frac{\sigma-1}{\sigma}}}\right], \tag{74}$$

$$\left.\frac{\partial MR_{\ell j}}{\partial \alpha_j}\right|_{\frac{L_{hj}}{L_{\ell j}}} = MR_{\ell j} \cdot \left[-\frac{1}{1-\alpha_j} + \frac{1}{\sigma-1} \cdot \frac{L_{\ell j}^{\frac{\sigma-1}{\sigma}}\left(\left(\frac{L_{hj}}{L_{\ell j}}\right)^{\frac{\sigma-1}{\sigma}} - 1\right)}{\alpha_j L_{hj}^{\frac{\sigma-1}{\sigma}} + (1-\alpha_j)L_{\ell j}^{\frac{\sigma-1}{\sigma}}}\right]. \tag{75}$$

Using (74), (75), and $S_j = \frac{\alpha_j L_{hj}^{\frac{\sigma-1}{\sigma}}}{\alpha_j L_{hj}^{\frac{\sigma-1}{\sigma}} + (1-\alpha_j)L_{\ell j}^{\frac{\sigma-1}{\sigma}}}$, we obtain

$$\begin{aligned}
S_j \frac{\Delta_h}{MR_{hj}^2} \cdot \frac{\partial MR_{hj}}{\partial \alpha_j} &= \frac{\Delta_h}{MR_{hj}} \cdot \left[\frac{L_{hj}^{\frac{\sigma-1}{\sigma}}}{\left(\alpha_j L_{hj}^{\frac{\sigma-1}{\sigma}} + (1-\alpha_j)L_{\ell j}^{\frac{\sigma-1}{\sigma}}\right)} + \frac{1}{\sigma-1} \cdot \frac{\alpha_j L_{hj}^{\frac{\sigma-1}{\sigma}}\left(L_{hj}^{\frac{\sigma-1}{\sigma}} - L_{\ell j}^{\frac{\sigma-1}{\sigma}}\right)}{\left(\alpha_j L_{hj}^{\frac{\sigma-1}{\sigma}} + (1-\alpha_j)L_{\ell j}^{\frac{\sigma-1}{\sigma}}\right)^2}\right] \\
&= \frac{\Delta_h}{MR_{hj}} \cdot \left[\left.\frac{\partial S_j}{\partial \alpha_j}\right|_{\frac{L_{hj}}{L_{\ell j}}} + \frac{\sigma}{\sigma-1} \cdot \frac{\alpha_j L_{hj}^{\frac{\sigma-1}{\sigma}}\left(L_{hj}^{\frac{\sigma-1}{\sigma}} - L_{\ell j}^{\frac{\sigma-1}{\sigma}}\right)}{\left(\alpha_j L_{hj}^{\frac{\sigma-1}{\sigma}} + (1-\alpha_j)L_{\ell j}^{\frac{\sigma-1}{\sigma}}\right)^2}\right] \\
(1-S_j) \frac{\Delta_\ell}{MR_{\ell j}^2} \cdot \frac{\partial MR_{\ell j}}{\partial \alpha_j} &= \frac{\Delta_\ell}{MR_{\ell j}} \cdot \left[\frac{-L_{\ell j}^{\frac{\sigma-1}{\sigma}}}{\left(\alpha_j L_{hj}^{\frac{\sigma-1}{\sigma}} + (1-\alpha_j)L_{\ell j}^{\frac{\sigma-1}{\sigma}}\right)} + \frac{1}{\sigma-1} \cdot \frac{(1-\alpha_j)L_{\ell j}^{\frac{\sigma-1}{\sigma}}\left(L_{hj}^{\frac{\sigma-1}{\sigma}} - L_{\ell j}^{\frac{\sigma-1}{\sigma}}\right)}{\left(\alpha_j L_{hj}^{\frac{\sigma-1}{\sigma}} + (1-\alpha_j)L_{\ell j}^{\frac{\sigma-1}{\sigma}}\right)^2}\right] \\
&= \frac{\Delta_\ell}{MR_{\ell j}} \cdot \left[-\left.\frac{\partial S_j}{\partial \alpha_j}\right|_{\frac{L_{hj}}{L_{\ell j}}} + \frac{\sigma}{\sigma-1} \cdot \frac{(1-\alpha_j)L_{\ell j}^{\frac{\sigma-1}{\sigma}}\left(L_{hj}^{\frac{\sigma-1}{\sigma}} - L_{\ell j}^{\frac{\sigma-1}{\sigma}}\right)}{\left(\alpha_j L_{hj}^{\frac{\sigma-1}{\sigma}} + (1-\alpha_j)L_{\ell j}^{\frac{\sigma-1}{\sigma}}\right)^2}\right].
\end{aligned}$$

Thus

$$
\begin{aligned}
& -\left[S_j \frac{\Delta_h}{MR_{hj}^2} \cdot \frac{\partial MR_{hj}}{\partial \alpha_j} + 1 - S_j) \frac{\Delta_\ell}{MR_{\ell j}^2} \cdot \frac{\partial MR_{\ell j}}{\partial \alpha_j}\right] \\
= & -\left.\frac{\partial S_j}{\partial \alpha_j}\right|_{\frac{L_{hj}}{L_{\ell j}}} \cdot \left(\frac{\Delta_h}{MR_{hj}} - \frac{\Delta_\ell}{MR_{\ell j}}\right) - \frac{\sigma}{\sigma - 1} \cdot \frac{\left(L_{hj}^{\frac{\sigma-1}{\sigma}} - L_{\ell j}^{\frac{\sigma-1}{\sigma}}\right)}{\left(\alpha_j L_{hj}^{\frac{\sigma-1}{\sigma}} + (1-\alpha_j) L_{\ell j}^{\frac{\sigma-1}{\sigma}}\right)} \cdot \left(S_j \frac{\Delta_h}{MR_{hj}} + (1 - S_j) \frac{\Delta_\ell}{MR_{\ell j}}\right),
\end{aligned}
$$

which is substituted into (72) to yield

$$
F_{1\alpha} = -\frac{\sigma}{\sigma - 1} \cdot \frac{\left(L_{hj}^{\frac{\sigma-1}{\sigma}} - L_{\ell j}^{\frac{\sigma-1}{\sigma}}\right)}{\left(\alpha_j L_{hj}^{\frac{\sigma-1}{\sigma}} + (1-\alpha_j) L_{\ell j}^{\frac{\sigma-1}{\sigma}}\right)} \cdot \left(S_j \frac{\Delta_h}{MR_{hj}} + (1 - S_j) \frac{\Delta_\ell}{MR_{\ell j}}\right).
$$

We then express $F2$—change in $\mu(z_j)$ through the change in employment share $\frac{L_{hj}}{L_{\ell j}}$. The derivative of the markdown (64) with respect to $\frac{L_{hj}}{L_{\ell j}}$ is

$$
\frac{\partial \mu_j}{\partial \left(\frac{L_{hj}}{L_{\ell j}}\right)} = \frac{\partial S_j}{\partial \left(\frac{L_{hj}}{L_{\ell j}}\right)} \cdot \left(\frac{\Delta_h}{MR_{hj}} - \frac{\Delta_\ell}{MR_{\ell j}}\right) - \left(S_j \frac{\Delta_h}{MR_{hj}^2} \cdot \frac{\partial MR_{hj}}{\partial \left(\frac{L_{hj}}{L_{\ell j}}\right)} + (1 - S_j) \frac{\Delta_\ell}{MR_{\ell j}^2} \cdot \frac{\partial MR_{\ell j}}{\partial \left(\frac{L_{hj}}{L_{\ell j}}\right)}\right), \quad (76)
$$

which includes the derivatives of weight S_j and marginal revenue products of labor MR_{sj}, $s = h, \ell$. Since $S_j = \frac{\alpha_j}{\alpha_j + (1-\alpha_j)\left(\frac{L_{hj}}{L_{\ell j}}\right)^{-\frac{\sigma-1}{\sigma}-1}}$, we obtain

$$
\begin{aligned}
\frac{\partial S_j}{\partial \left(\frac{L_{hj}}{L_{\ell j}}\right)} &= \frac{\sigma - 1}{\sigma} \cdot \frac{\alpha_j \cdot (1 - \alpha_j)}{\left(\alpha_j + (1 - \alpha_j \left(\frac{L_{hj}}{L_{\ell j}}\right)^{-\frac{\sigma-1}{\sigma}-1})\right)^2} \cdot \left(\frac{L_{hj}}{L_{\ell j}}\right)^{-\frac{\sigma-1}{\sigma}-1} \\
&= (\sigma - 1) \cdot \frac{\alpha_j (1 - \alpha_j)}{\sigma} \cdot \frac{L_{hj}^{\frac{\sigma-1}{\sigma}} L_{\ell j}^{\frac{\sigma-1}{\sigma}}}{\left(\alpha_j L_{hj}^{\frac{\sigma-1}{\sigma}} + (1-\alpha_j) L_{\ell j}^{\frac{\sigma-1}{\sigma}}\right)^2} \cdot \left(\frac{L_{hj}}{L_{\ell j}}\right)^{-1}. \quad (77)
\end{aligned}
$$

Once again, from the marginal revenue products $MR_{hj} = \alpha_j z_j \left(\alpha_j + (1 - \alpha_j)\left(\frac{L_{hj}}{L_{\ell j}}\right)^{-\frac{\sigma-1}{\sigma}}\right)^{\frac{1}{\sigma-1}}$ and $MR_{\ell j} = (1 - \alpha_j) z_j \left(\alpha_j \left(\frac{L_{hj}}{L_{\ell j}}\right)^{\frac{\sigma-1}{\sigma}} + 1 - \alpha_j\right)^{\frac{1}{\sigma-1}}$, we obtain

$$
\begin{aligned}
\frac{\partial MR_{hj}}{\partial \left(\frac{L_{hj}}{L_{\ell j}}\right)} &= -MR_{hj} \cdot \frac{\alpha_j}{\sigma} \left(\alpha_j L_{hj}^{\frac{\sigma-1}{\sigma}} + (1-\alpha_j) L_{\ell j}^{\frac{\sigma-1}{\sigma}}\right)^{-1} \cdot L_{\ell j}^{\frac{\sigma-1}{\sigma}} \cdot \left(\frac{L_{hj}}{L_{\ell j}}\right)^{-1} \\
\frac{\partial MR_{\ell j}}{\partial \left(\frac{L_{hj}}{L_{\ell j}}\right)} &= MR_{\ell j} \cdot \frac{(1 - \alpha_j)}{\sigma} \left(\alpha_j L_{hj}^{\frac{\sigma-1}{\sigma}} + (1-\alpha_j) L_{\ell j}^{\frac{\sigma-1}{\sigma}}\right)^{-1} \cdot L_{hj}^{\frac{\sigma-1}{\sigma}} \cdot \left(\frac{L_{hj}}{L_{\ell j}}\right)^{-1},
\end{aligned}
$$

and thus

$$S_j \frac{\Delta_h}{MR_{hj}^2} \cdot \frac{\partial MR_{hj}}{\partial\left(\frac{L_{hj}}{L_{\ell j}}\right)} = -\frac{\alpha_j(1-\alpha_j)}{\sigma} \cdot \frac{L_{hj}^{\frac{\sigma-1}{\sigma}} L_{\ell j}^{\frac{\sigma-1}{\sigma}}}{\left(\alpha_j L_{hj}^{\frac{\sigma-1}{\sigma}} + (1-\alpha_j) L_{\ell j}^{\frac{\sigma-1}{\sigma}}\right)^2} \cdot \left(\frac{L_{hj}}{L_{\ell j}}\right)^{-1} \cdot \frac{\Delta_h}{MR_{hj}} \quad (78)$$

$$(1-S_j) \frac{\Delta_\ell}{MR_{\ell j}^2} \cdot \frac{\partial MR_{\ell j}}{\partial\left(\frac{L_{hj}}{L_{\ell j}}\right)} = \frac{\alpha_j(1-\alpha_j)}{\sigma} \cdot \frac{L_{hj}^{\frac{\sigma-1}{\sigma}} L_{\ell j}^{\frac{\sigma-1}{\sigma}}}{\left(\alpha_j L_{hj}^{\frac{\sigma-1}{\sigma}} + (1-\alpha_j) L_{\ell j}^{\frac{\sigma-1}{\sigma}}\right)^2} \cdot \left(\frac{L_{hj}}{L_{\ell j}}\right)^{-1} \cdot \frac{\Delta_\ell}{MR_{\ell j}}. \quad (79)$$

Substituting (77), (78), and (79) into (76) yields

$$\frac{\partial \mu_j}{\partial\left(\frac{L_{hj}}{L_{\ell j}}\right)} = \alpha_j(1-\alpha_j) \cdot \frac{L_{hj}^{\frac{\sigma-1}{\sigma}} L_{\ell j}^{\frac{\sigma-1}{\sigma}}}{\left(\alpha_j L_{hj}^{\frac{\sigma-1}{\sigma}} + (1-\alpha_j) L_{\ell j}^{\frac{\sigma-1}{\sigma}}\right)^2} \cdot \left(\frac{L_{hj}}{L_{\ell j}}\right)^{-1} \cdot \left(\frac{\Delta_h}{MR_{hj}} - \frac{\Delta_\ell}{MR_{\ell j}}\right).$$

Since $\frac{d\left(\frac{L_{hj}}{L_{\ell j}}\right)}{dz_j} = \left[\frac{d\left(\frac{L_{hj}}{L_{\ell j}}\right)}{d\alpha_j} \cdot \alpha'(z_j) + \frac{d\left(\frac{L_{hj}}{L_{\ell j}}\right)}{d\left(\frac{g_{hj}}{g_{\ell j}}\right)} \cdot \frac{d\left(\frac{g_{hj}}{g_{\ell j}}\right)}{dz_j}\right]$, it is straightforward to obtain the expression (22) using $F2 = \frac{\partial \mu_j}{\partial\left(\frac{L_{hj}}{L_{\ell j}}\right)} \times \frac{d\left(\frac{L_{hj}}{L_{\ell j}}\right)}{dz_j}$.

Lastly, to determine the sign of $F2$, we determine the sign of $\frac{\Delta_h}{MR_{hj}} - \frac{\Delta_\ell}{MR_{\ell j}}$. From (58), we have

$$\frac{\Delta_h}{MR_{hj}} - \frac{\Delta_\ell}{MR_{\ell j}} = -\frac{\zeta_j}{\lambda} \frac{\sigma+\beta}{\sigma(1+\beta)} \left(\frac{1}{MR_{hj} L_{hj}} + \frac{1}{MR_{\ell j} L_{\ell j}}\right) \cdot \mu_j,$$

which shows that $sign(\frac{\Delta_h}{MR_{hj}} - \frac{\Delta_\ell}{MR_{\ell j}}) = -sign(\zeta_j) < 0$ if the no discrimination constraint is binding.

Proof of Proposition 15

We solve the planning problem in Definition 3 assuming $g_j = 1$ for all j. By combining the first order conditions with respect to $w_j (= c_j)$ and L_j and multiplying $\frac{w_j}{L_j}$, we obtain

$$1 - \lambda(1+\beta) \cdot w_j + \lambda\beta \times MPL_j - \nu\beta + \frac{I}{N} = 0. \quad (80)$$

We denote the multiplier associated with labor supply curve for firm j (7) by ξ_j. From the first order conditions with respect to wage index I, we obtain

$$\nu = \sum_k \xi_k \frac{w_k^\beta}{I^2} N = \frac{1}{I} \cdot \sum_k \xi_k L_k = \frac{\lambda}{I} \cdot \sum_k \left(MPL_k - w_k\right) L_k = \frac{\lambda N}{I} \cdot \sum_k \left(MPL_k - w_k\right) s_k, \quad (81)$$

where the second equality is from the labor supply curve, and the third equality is from the first order condition with respect to L_j. Then by combining (80) and (81), we obtain

$$1 - \lambda(1+\beta) w_j + \lambda\beta \times MPL_j - \lambda\beta \cdot \sum_k \left(MPL_k - w_k\right) s_k = 0.$$

We can rewrite this by collecting terms associated with w_j and MPL_j, which results in (28).

ONLINE APPENDIX

Optimal Firm Regulation with Labor Market Monopsony

Hanvit Kim
Seoul National University

Yena Park
Seoul National University

A Type-Specific Policy in the Economy with Multiple Types

In Section 3 of the main text, we have discussed that the analysis in an economy with single skill type is essentially identical to the analysis in an economy with multiple types in the presence of type-specific policy. In this section, we show this explicitly.

The extension to the multiple skill types is very straightforward as long as the planner can use a type-specific subsidy schedule. This is because the planner can separate the redistribution across workers at different firms within skill type from the redistribution between the skill types. Here we consider a baseline model with two skill types $s \in \{h, \ell\}$. A firm j produces consumption good using a CES production technology $y_j = z_j(\alpha_j L_{hj}^{\frac{\sigma-1}{\sigma}} + (1-\alpha_j)L_{\ell j}^{\frac{\sigma-1}{\sigma}})^{\frac{\sigma}{\sigma-1}}$, where Lhj and $L_{\ell j}$ are number of workers hired at the firm j. Next proposition characterizes the type-specific optimal markdown schedule.

Proposition AP.1. *In an economy with two types of workers and type-specific policy, the optimal consumption allocation is given by*

$$c_{hj} = \frac{\beta}{1+\beta} \times MR_{hj} + \Delta_h, \quad c_{\ell j} = \frac{\beta}{1+\beta} \times MR_{\ell j} + \Delta_\ell \tag{A.1}$$

where $\Delta_h = \frac{1}{\lambda(1+\beta)}\left(1 - \beta\nu_h \frac{I_h}{N_h}\right)$, $\Delta_\ell = \frac{1}{\lambda(1+\beta)}\left(1 - \beta\nu_\ell \frac{I_\ell}{N_\ell}\right)$, *and* $MR_{hj}, MR_{\ell j}$ *are marginal revenue products of each type of worker at firm* j.
The slope of the optimal markdown schedule is given by

$$\frac{d\mu_{hj}}{dz_j} = -\frac{\Delta_h}{MR_{hi}} \frac{d\log MR_{hj}}{dz_j}, \quad \frac{d\mu_{\ell j}}{dz_j} = -\frac{\Delta_\ell}{MR_{\ell i}} \frac{d\log MR_{\ell j}}{dz_j}. \tag{A.2}$$

Suppose that $\alpha_i = \alpha, g_{hi} = g_{\ell i} = g\ \forall i$. *There exists a cutoff Pareto weight* $\bar{\psi}_f$ *such that for* $\psi_f < \bar{\psi}_f$, $\Delta_h > 0$ *and* $\Delta_l > 0$. *Since* $\frac{d\log MR_{hj}}{dz_j}, \frac{d\log MR_{\ell j}}{dz_j} > 0$, *the slope of the optimal markdown schedule is negative:* $\frac{d\log \mu_{hj}}{dz_j}, \frac{d\log \mu_{\ell j}}{dz_j} < 0$.

Proof

The problem is

$$max_{\{c_{hj},c_{\ell j},c_{fj}\},\{L_{hj},L_{\ell j}\},I_h,I_\ell} \quad N_h \int_\varepsilon V_h(\varepsilon;\{c_{hj}\})dF(\varepsilon) + N_\ell \int_\varepsilon V_\ell(\varepsilon;\{c_{\ell j}\})dF(\varepsilon) + \psi_f \left(\sum_j \log c_{fj}\right) \tag{A.3}$$

$$s.t. \quad \sum_j (c_{hj}L_{hj} + c_{\ell j}L_{\ell j} + c_{fj}) \leq \sum_j z_j(\alpha_j L_{hj}^{\frac{\sigma-1}{\sigma}} + (1-\alpha_j)L_{\ell j}^{\frac{\sigma-1}{\sigma}})^{\frac{\sigma}{\sigma-1}} \tag{A.4}$$

$$I_h = \sum_j (c_{hj}g_{hj})^\beta, \quad I_\ell = \sum_j (c_{\ell j}g_{\ell j})^\beta \tag{A.5}$$

$$L_{hj} = N_h \frac{(c_{hj}g_{hj})^\beta}{I_h}, \quad L_{\ell j} = N_\ell \frac{(c_{\ell j}g_{\ell j})^\beta}{I_\ell}, \quad \forall j = 1,\ldots,J \tag{A.6}$$

Lagrangian is given by

$$\mathscr{L} = \frac{N_h}{\beta}\left[\log\left(\sum_j (c_{hj}g_{hj})^\beta\right)\right] + \frac{N_\ell}{\beta}\left[\log\left(\sum_j (c_{\ell j}g_{\ell j})^\beta\right)\right] + \psi_f \sum_j \log c_{fj} \tag{A.7}$$

$$+ \lambda(\sum_j z_j(\alpha_j L_{hj}^{\frac{\sigma-1}{\sigma}} + (1-\alpha_j)L_{\ell j}^{\frac{\sigma-1}{\sigma}})^{\frac{\sigma}{\sigma-1}} - \sum_j (c_{hj}L_{hj} + c_{\ell j}L_{\ell j} + c_{fj})) \tag{A.8}$$

$$+ \nu_h(I_h - \sum_j (c_{hj}g_{hj})^\beta) + \nu_\ell(I_\ell - \sum_j (c_{\ell j}g_{\ell j})^\beta) \tag{A.9}$$

When taking first order conditions, we take into account $L_{hj} = N_h \frac{(c_{hj}g_{hj})^\beta}{I_h}, L_{\ell j} = N_\ell \frac{(c_{\ell j}g_{\ell j})^\beta}{I_\ell}$ First order condition with respect to c_{hj}, c_{lj} yields

$$c_{hj} = \frac{\beta}{1+\beta} MR_{hj} + \Delta_h, \quad \Delta_h = \frac{1}{\lambda(1+\beta)}\left(1 - \beta\nu_h \frac{I_h}{N_h}\right) \tag{A.10}$$

$$c_{\ell j} = \frac{\beta}{1+\beta} MR_{\ell j} + \Delta_\ell, \quad \Delta_\ell = \frac{1}{\lambda(1+\beta)}\left(1 - \beta\nu_\ell \frac{I_\ell}{N_\ell}\right) \tag{A.11}$$

Multiplying by $L_{hj}, L_{\ell j}$ and summing up over j we have

$$\sum_j c_{hj}L_{hj} = \frac{\beta}{1+\beta}\sum_j MR_{hj}L_{hj} + N_h\Delta_h, \quad \sum_j c_{\ell j}L_{\ell j} = \frac{\beta}{1+\beta}\sum_j MR_{\ell j}L_{\ell j} + N_\ell\Delta_\ell \tag{A.12}$$

First order condition with respect to I_h, I_ℓ yields

$$\nu_h = \lambda\frac{\sum_j (MR_{hj} - c_{hj})L_{hj}}{I_h}, \quad \nu_\ell = \lambda\frac{\sum_j (MR_{\ell j} - c_{\ell j})L_{\ell j}}{I_\ell} \tag{A.13}$$

Substituting formula for ν_h, ν_ℓ to Δ_h, Δ_ℓ yields

$$\frac{1}{\lambda} = \frac{\sum_j c_{hj}L_{hj}}{N_h} = \frac{\sum_j c_{\ell j}L_{\ell j}}{N_\ell} \tag{A.14}$$

First order conditions with respect to c_{fj} yields

$$\frac{1}{\lambda} = \frac{c_{fj}}{\psi_f} \tag{A.15}$$

Thus $\forall j, c_{fj} = c_f$.
Using resource constraint we have

$$\frac{1}{\lambda} = \frac{1}{N_h + N_\ell + \psi_f J} \sum_j z_j \tilde{y}_j \tag{A.16}$$

Thus

$$\frac{1}{\lambda} = \frac{c_{fj}}{\psi_f} = \frac{\sum_j c_{hj} L_{hj}}{N_h} = \frac{\sum_j c_{\ell j} L_{\ell j}}{N_\ell} = \frac{1}{N_h + N_\ell + \psi_f J} \sum_j z_j \tilde{y}_j \tag{A.17}$$

Now suppose that $\alpha_j = \alpha, \forall j$, $g_{hj} = g_{\ell j} = 1, \forall j$.
Substituting $\frac{1}{\lambda} = \frac{1}{N_h + N_\ell + \psi_f J} \sum_j z_j \tilde{y}_j$ to ν_h we have

$$\nu_h \frac{I_h}{N_h} = \frac{N_h + N_\ell + \psi_f J}{N_h} \left(\frac{\sum_j MR_{hj} L_{hj}}{\sum_j z_j \tilde{y}_j} - \frac{\sum_j c_{hj} L_{hj}}{\sum_j z_j \tilde{y}_j} \right) \tag{A.18}$$

Using $\frac{\sum_j MR_{hj} L_{hj}}{\sum_j z_j \tilde{y}_j} = \alpha$ and $\frac{\sum_j c_{hj} L_{hj}}{\sum_j z_j \tilde{y}_j} = \frac{N_h}{N_h + N_\ell + \psi_f J}$ we have

$$\nu_h \frac{I_h}{N_h} = \alpha \frac{N_h + N_\ell + \psi_f J}{N_h} - 1 \tag{A.19}$$

Similarly,

$$\nu_\ell \frac{I_\ell}{N_\ell} = (1 - \alpha) \frac{N_h + N_\ell + \psi_f J}{N_\ell} - 1 \tag{A.20}$$

Thus

$$1 - \beta \nu_h \frac{I_h}{N_h} > 0 \quad \Leftrightarrow \quad \frac{N_h}{N_h + N_\ell + \psi_f J} > \alpha \frac{\beta}{1 + \beta} \tag{A.21}$$

$$1 - \beta \nu_\ell \frac{I_\ell}{N_\ell} > 0 \quad \Leftrightarrow \quad \frac{N_\ell}{N_h + N_\ell + \psi_f J} > (1 - \alpha) \frac{\beta}{1 + \beta} \tag{A.22}$$

This completes the proof. ■

The planner with type-dependent subsidy is able to set desired level of average consumption of each skill. Thus, the planner only faces the exactly same equity-efficiency trade-off as in the one-type model. If the planner's desire to redistribute to workers is strong enough (i.e. $\Delta_h, \Delta_l > 0$), it is optimal to compress consumption allocation across firms within a skill type, which can be obtained by the downward sloping markdown schedule for each skill type.

B Appendix for Quantitative Analysis

B.1 Optimal Markdown Formula of the Quantitative Model in Section 5

We present and prove the optimal markdown formula for the model we have used in our quantitative exercise. In the model, there are $M \geq 2$ types of workers, and workers of different type are perfectly substitutable in production. In addition, we assume that there are decreasing returns to scale.

In Section 5 of the main text, the optimal markdown equation is given by

$$\mu_j = \frac{\beta}{1+\beta} + \sum_m \frac{\alpha_{mj} L_{mj}}{L_j} \frac{\Delta_m}{\eta z_j L_j^{\eta-1} \alpha_{mj}}, \quad \Delta_m = \frac{1}{\lambda(1+\beta)} \left(1 - \nu_m \frac{I_m}{N_m}\right), \tag{A.23}$$

where $I_m = \left(\sum_m c_{mj}^\beta g_{mj}^\beta\right)^{1/\beta}$ is the wage index for type m workers, and multipliers λ and ν_m satisfy

$$\frac{1}{\lambda} = \frac{\sum_j z_j L_j^\eta}{N + \psi_f J}, \quad \frac{\nu_m}{\lambda} = \beta \frac{\sum_j (\eta z_j L_j^{\eta-1} \alpha_{mj} - c_{mj}) L_{mj}}{I_m}. \tag{A.24}$$

To prove this formula, we start from stating the planner's problem

$$\max_{\{c_{mj}, c_{fj}\}} \quad \sum_m \frac{N_m}{\beta} \log\left(\sum_m c_{mj}^\beta g_{mj}^\beta\right) + \psi_f \sum_j \log c_{fj} \tag{A.25}$$

$$s.t. \quad \sum_j \sum_m c_{mj} L_{mj} + \sum_j c_{fj} \leq \sum_j z_j \left(\sum_m \alpha_{mj} L_{mj}\right)^\eta \tag{A.26}$$

$$I_m = \left(\sum_m c_{mj}^\beta g_{mj}^\beta\right)^{1/\beta}, \quad m = 1, \ldots, M \tag{A.27}$$

$$\frac{c_{m+1,j}}{\eta z_j L_j^{\eta-1} \alpha_{m+1,j}} = \frac{c_{mj}}{\eta z_j L_j^{\eta-1} \alpha_{mj}}, \quad j = 1 \ldots J, m = 1, \ldots, M-1 \tag{A.28}$$

and $L_{mj} = N_m \frac{c_{mj}^\beta g_{mj}^\beta}{I_m}$.

Note that the no-discrimination constraint can be rewritten as

$$\frac{c_{m+1,j}}{\eta z_j L_j^{\eta-1} \alpha_{m+1,j}} = \frac{c_{mj}}{\eta z_j L_j^{\eta-1} \alpha_{mj}} \quad \Leftrightarrow \frac{c_{m+1,j}}{\alpha_{m+1,j}} = \frac{c_{mj}}{\alpha_{mj}}. \tag{A.29}$$

From now, we will use the constraint $\frac{c_{m+1,j}}{\alpha_{m+1,j}} = \frac{c_{mj}}{\alpha_{mj}}$ instead and denote the associated lagrange multiplier $\tilde{\zeta}$.

Lagrangian is given by

$$\mathscr{L} = \sum_m \frac{N_m}{\beta} \log\left(\sum_m c_{mj}^\beta g_{mj}^\beta\right) + \psi_f \sum_j \log c_{fj} + \lambda \left(\sum_j z_j \left(\sum_m \alpha_{mj} L_{mj}\right)^\eta - \sum_j \sum_m c_{mj} L_{mj} + \sum_j c_{fj}\right)$$
$$+ \sum_m \nu_m \left(I_m - \left(\sum_m c_{mj}^\beta g_{mj}^\beta\right)^{1/\beta}\right) + \sum_j \sum_{m=1}^{M-1} \tilde{\zeta_{mj}} \left(\frac{c_{m+1,j}}{\alpha_{m+1,j}} - \frac{c_{mj}}{\alpha_{mj}}\right)$$

First order condition with respect to c_{mj} is

$$\frac{L_{mj}}{c_{mj}} + \lambda((\eta z_j L_j^{\eta-1}\alpha_{mj} - c_{mj})\beta\frac{L_{mj}}{c_{mj}} - L_{mj}) - \nu_m\frac{I_m}{N_m}\frac{L_{mj}}{c_{mj}} + \frac{1}{\alpha_{mj}}(\tilde{\zeta}_{m-1,j} - \tilde{\zeta}_{mj}) = 0. \tag{A.30}$$

Multiply by $\frac{L_{mj}}{c_{mj}}\frac{1}{\lambda(1+\beta)}$ and rearranging yields

$$c_{mj}\left(1 + \frac{1}{1+\beta}\frac{1}{\alpha_{mj}L_{mj}}\left(\frac{\tilde{\zeta}_{mj}}{\lambda} - \frac{\tilde{\zeta}_{m-1,j}}{\lambda}\right)\right) = \frac{\beta}{1+\beta}\eta z_j L_j^{\eta-1}\alpha_{mj} + \Delta_m, \tag{A.31}$$

where we define $\Delta_m = \frac{1}{\lambda(1+\beta)}\left(1 - \nu_m\frac{I_m}{N_m}\right)$.

To obtain the expression for $\mu_j = \frac{c_{mj}}{\eta z_j L_j^{\eta-1}\alpha_{mj}}$, divide by $\eta z_j L_j^{\eta-1}\alpha_{mj}$

$$\mu_j\left(1 + \frac{1}{1+\beta}\frac{1}{\alpha_{mj}L_{mj}}\left(\frac{\tilde{\zeta}_{mj}}{\lambda} - \frac{\tilde{\zeta}_{m-1,j}}{\lambda}\right)\right) = \frac{\beta}{1+\beta} + \frac{\Delta_m}{\eta z_j L_j^{\eta-1}\alpha_{mj}}. \tag{A.32}$$

To cancel out the multipliers for no-discrimination constraint $\tilde{\zeta}$, multiply by $\frac{\alpha_{mj}L_{mj}}{L_j}$ and sum over m

$$\mu_j = \frac{\beta}{1+\beta} + \sum_m \frac{\alpha_{mj}L_{mj}}{L_j}\frac{\Delta_m}{\eta z_j L_j^{\eta-1}\alpha_{mj}}, \tag{A.33}$$

which gives us the equation (A.23).

We proceed to obtain equations for the multipliers ν_m and λ, which are used to define Δ_m. First order condition with respect to I_m is given by

$$\nu_m - \lambda\sum_j \beta\frac{L_{mj}}{I_m}(\eta z_j L_j^{\eta-1}\alpha_{mj} - c_{mj}) = 0. \tag{A.34}$$

Dividing by λ and rearranging yields

$$\frac{\nu_m}{\lambda} = \beta\frac{\sum_j(\eta z_j L_j^{\eta-1}\alpha_{mj} - c_{mj})L_{mj}}{I_m}. \tag{A.35}$$

First order condition with respect to c_{fj} is

$$\frac{\psi_f}{c_{fj}} = \lambda, \tag{A.36}$$

or equivalently $c_{fj} = \frac{\psi_f}{\lambda}$. Recall that the formula for c_{mj} is given by (A.25). Multiply by L_{mj}

$$c_{mj}L_{mj} + \frac{1}{1+\beta}\frac{\mu_j}{\eta z_j L_j^{\eta-1}}\left(\frac{\tilde{\zeta}_{mj}}{\lambda} - \frac{\tilde{\zeta}_{m-1,j}}{\lambda}\right) \tag{A.37}$$

$$= \frac{\beta}{1+\beta}\eta z_j L_j^{\eta-1}\alpha_{mj}L_{mj} + \frac{L_{mj}}{\lambda(1+\beta)} - \frac{\beta}{1+\beta}\frac{\sum_j(\eta z_j L_j^{\eta-1}\alpha_{mj} - c_{mj})L_{mj}}{N_m}L_{mj}, \tag{A.38}$$

where we've used $\Delta_m = \frac{1}{\lambda(1+\beta)}\left(1-\nu_m\frac{I_m}{N_m}\right)$ and equation (A.35). Summing up over m, j yields

$$\sum_j \sum_m c_{mj} L_{mj} = \frac{N}{\lambda}. \tag{A.39}$$

Substituting $c_{fj} = \frac{\psi_f}{\lambda}$ and $\sum_j \sum_m c_{mj} L_{mj} = \frac{N}{\lambda}$ to the resource constraint yields $\frac{1}{\lambda} = \frac{\sum_j z_j L_j^\eta}{N + \psi_f J}$. This ends the proof.

B.2 $\mu'(z_j)$-Decomposition Formula in the Quantitative Model

We derive and present the decomposition of $\frac{d\mu}{d\log z}$ for the model we used in quantitative exercise. Recall that the optimal markdown formula is given by

$$\mu_j = \frac{\beta}{1+\beta} + \frac{1}{\eta z_j L_j^{1-\eta}} \frac{\sum_m L_{mj}\Delta_m}{L_j}. \tag{A.40}$$

Rearranging the terms, we have

$$\mu_j = \frac{\beta}{1+\beta} + \frac{1}{\eta z_j \left(\sum_m \alpha_{mj} \frac{L_{mj}}{\bar{L}_j}\right)^\eta \bar{L}_j^{\eta-1}} \sum_m \Delta_m \frac{L_{mj}}{\bar{L}_j}, \tag{A.41}$$

where $\bar{L}_j$ is the total number of employees working at firm j.
The formula for the decomposition of $\frac{d\mu}{d\log z}$ is given as

$$\frac{d\mu_j}{d\log z_j} = F1_z + F1_\alpha + F2, \tag{A.42}$$

$$F1_z = -\left(\mu_j - \frac{\beta}{1+\beta}\right) \times \left[1 - (1-\eta) \times \frac{d\log \bar{L}_j}{d\log z_j}\right], \tag{A.43}$$

$$F1_\alpha = -\left(\mu_j - \frac{\beta}{1+\beta}\right) \times \left[\eta \times \sum_m \left(\frac{\alpha_{mj} L_{mj}}{\sum_k \alpha_{kj} L_{kj}} \times \frac{d\log \alpha_{mj}}{d\log z_j}\right)\right], \tag{A.44}$$

$$F2 = -\left(\mu_j - \frac{\beta}{1+\beta}\right) \times \left[\eta \times \sum_m \left(\frac{\alpha_{mj} L_{mj}}{\sum_k \alpha_{kj} L_{kj}} \times \frac{d\log \frac{L_{mj}}{\bar{L}_j}}{d\log z_j}\right) - \sum_m \left(\frac{\Delta_m L_{mj}}{\sum_k \Delta_k L_{kj}} \times \frac{d\log \frac{L_{mj}}{\bar{L}_j}}{d\log z_j}\right)\right]. \tag{A.45}$$

We start from the derivation of $F1_z$, which captures the direct effect of increasing z on the optimal schedule. Taking derivatives, we have

$$\frac{\partial \mu_j}{\partial z_j} + \frac{\partial \mu_j}{\partial \bar{L}_j}\frac{d\bar{L}_j}{dz_j} = -\frac{\sum_k \Delta_k \frac{L_{kj}}{\bar{L}_j}}{\eta z_j \left(\sum_k \alpha_{kj} \frac{L_{kj}}{\bar{L}_j}\right)^\eta \bar{L}_j^{\eta-1}} \times \frac{1}{z_j} + \frac{\sum_k \Delta_k \frac{L_{kj}}{\bar{L}_j}}{\eta z_j \left(\sum_k \alpha_{kj} \frac{L_{kj}}{\bar{L}_j}\right)^\eta \bar{L}_j^{\eta-1}} \times \frac{1-\eta}{\bar{L}_j}\frac{d\bar{L}_j}{dz_j}. \tag{A.46}$$

Multiplying by z_j, replacing $d\bar{L}_j, dz_j$ with $\bar{L}_j d\log \bar{L}_j, z_j \log z_j$ and substituting $\frac{\sum_k \Delta_k \frac{L_{kj}}{\bar{L}_j}}{\eta z_j \left(\sum_k \alpha_{kj} \frac{L_{kj}}{\bar{L}_j}\right)^{\eta} \bar{L}_j^{\eta-1}} = \mu_j - \frac{\beta}{1+\beta}$ yields

$$\begin{aligned}\frac{\partial \mu_j}{\partial \log z_j} + \frac{\partial \mu_j}{\partial \log \bar{L}_j}\frac{d\log \bar{L}_j}{d\log z_j} &= -\left(\mu_j - \frac{\beta}{1+\beta}\right) + \left(\mu_j - \frac{\beta}{1+\beta}\right) \times (1-\eta) \times \frac{d\log \bar{L}_j}{d\log z_j} \\ &= -\left(\mu_j - \frac{\beta}{1+\beta}\right) \times \left[1 - (1-\eta) \times \frac{d\log \bar{L}_j}{d\log z_j}\right]. \end{aligned} \tag{A.47}$$

which gives our expression for $F1_z$.

$F1_\alpha$ captures the indirect effect of increasing z on the markdown schedule through α. Taking derivatives with respect to α_{mj}, we have

$$\frac{\partial \mu_j}{\partial \alpha_{mj}}\frac{d\alpha_{mj}}{dz_j} = -\frac{\sum_k \Delta_k \frac{L_{kj}}{\bar{L}_j}}{\eta z_j \left(\sum_k \alpha_{kj} \frac{L_{kj}}{\bar{L}_j}\right)^{\eta} \bar{L}_j^{\eta-1}} \times \eta \times \frac{\frac{L_{mj}}{\bar{L}_j}}{\sum_k \alpha_{kj} \frac{L_{kj}}{\bar{L}_j}} \times \frac{d\alpha_{mj}}{dz_j}. \tag{A.48}$$

Multiplying by z_j, replacing $d\alpha_{mj}, dz_j$ with $\alpha_{mj} d\log \alpha_{mj}, z_j \log z_j$ and substituting $\frac{\sum_k \Delta_k \frac{L_{kj}}{\bar{L}_j}}{\eta z_j \left(\sum_k \alpha_{kj} \frac{L_{kj}}{\bar{L}_j}\right)^{\eta} \bar{L}_j^{\eta-1}} = \mu_j - \frac{\beta}{1+\beta}$ yields

$$\frac{\partial \mu_j}{\partial \log \alpha_{mj}}\frac{d\log \alpha_{mj}}{d\log z_j} = -\left(\mu_j - \frac{\beta}{1+\beta}\right) \times \eta \times \frac{\alpha_{mj} L_{mj}}{\sum_k \alpha_{kj} L_{kj}} \times \frac{d\log \alpha_{mj}}{d\log z_j}. \tag{A.49}$$

Summing up over m gives

$$\sum_m \left(\frac{\partial \mu_j}{\partial \log \alpha_{mj}}\frac{d\log \alpha_{mj}}{d\log z_j}\right) = -\left(\mu_j - \frac{\beta}{1+\beta}\right) \times \eta \times \sum_m \left(\frac{\alpha_{mj} L_{mj}}{\sum_k \alpha_{kj} L_{kj}} \times \frac{d\log \alpha_{mj}}{d\log z_j}\right), \tag{A.50}$$

which gives our expression for $F1_\alpha$.

Finally, we derive the term $F2$. Taking derivatives with respect to $\frac{L_{mj}}{\bar{L}_j}$,

$$\frac{\partial \mu_j}{\partial \frac{L_{mj}}{\bar{L}_j}}\frac{d\frac{L_{mj}}{\bar{L}_j}}{dz_j} = -\frac{\sum_k \Delta_k \frac{L_{kj}}{\bar{L}_j}}{\eta z_j \left(\sum_k \alpha_{kj} \frac{L_{kj}}{\bar{L}_j}\right)^{\eta} \bar{L}_j^{\eta-1}} \times \eta \times \frac{\alpha_{mj}}{\sum_k \alpha_{kj} \frac{L_{kj}}{\bar{L}_j}} \times \frac{d\frac{L_{mj}}{\bar{L}_j}}{dz_j} \tag{A.51}$$

$$+ \frac{\sum_k \Delta_k \frac{L_{kj}}{\bar{L}_j}}{\eta z_j \left(\sum_k \alpha_{kj} \frac{L_{kj}}{\bar{L}_j}\right)^{\eta} \bar{L}_j^{\eta-1}} \times \frac{\Delta_m}{\sum_k \Delta_k \frac{L_{kj}}{\bar{L}_j}} \times \frac{d\frac{L_{mj}}{\bar{L}_j}}{dz_j}. \tag{A.52}$$

Multiplying by z_j, replacing $d\frac{L_{mj}}{\bar{L}_j}, dz_j$ with $\frac{L_{mj}}{\bar{L}_j} d\log \frac{L_{mj}}{\bar{L}_j}, z_j \log z_j$ and substituting $\frac{\sum_k \Delta_k \frac{L_{kj}}{\bar{L}_j}}{\eta z_j \left(\sum_k \alpha_{kj} \frac{L_{kj}}{\bar{L}_j}\right)^{\eta} \bar{L}_j^{\eta-1}} =$

$\mu_j - \frac{\beta}{1+\beta}$ yields

$$\frac{\partial \mu_j}{\partial \log \frac{L_{mj}}{\bar{L}_j}} \frac{d \log \frac{L_{mj}}{\bar{L}_j}}{d \log z_j} = -\left(\mu_j - \frac{\beta}{1+\beta}\right) \times \eta \times \frac{\alpha_{mj} L_{mj}}{\sum_k \alpha_{kj} L_{kj}} \times \frac{d \log \frac{L_{mj}}{\bar{L}_j}}{d \log z_j} \tag{A.53}$$

$$+\left(\mu_j - \frac{\beta}{1+\beta}\right) \times \frac{\Delta_m L_{mj}}{\sum_k \Delta_k L_{kj}} \times \frac{d \log \frac{L_{mj}}{\bar{L}_j}}{d \log z_j}. \tag{A.54}$$

Summing up over m and n gives

$$\sum_m \left(\frac{\partial \mu_j}{\partial \log \frac{L_{mj}}{\bar{L}_j}} \frac{d \log \frac{L_{mj}}{\bar{L}_j}}{d \log z_j} \right) = -\left(\mu_j - \frac{\beta}{1+\beta}\right) \times \tag{A.55}$$

$$\left[\eta \times \sum_m \left(\frac{\alpha_{mj} L_{mj}}{\sum_k \alpha_{kj} L_{kj}} \times \frac{d \log \frac{L_{mj}}{\bar{L}_j}}{d \log z_j} \right) - \sum_m \left(\frac{\Delta_m L_{mj}}{\sum_k \Delta_k L_{kj}} \times \frac{d \log \frac{L_{mj}}{\bar{L}_j}}{d \log z_j} \right) \right], \tag{A.56}$$

which is the $F2$ term.

B.3 Algorithm for Computing the Equilibrium

In equilibrium without any redistributive policy, by solving the firm's optimization, we know that wage of a type-m worker working at firm j is equal to the markdown of type-m worker's marginal revenue product at firm j:

$$w_{mj} = \frac{\beta}{1+\beta} \eta z_j L_j^{\eta-1} \alpha_{mj}. \tag{A.57}$$

Workers' optimal choice of workplace imply the following labor supply curve

$$L_{mj} = N_m \frac{w_{mj}^\beta g_{mj}^\beta}{\sum_k (w_{mk}^\beta g_{mk}^\beta)} = N_m \frac{z_j^\beta L_j^{-\beta(1-\eta)} \alpha_{mj}^\beta g_{mj}^\beta}{\sum_k (z_k^\beta L_k^{-\beta(1-\eta)} \alpha_{mk}^\beta g_{mk}^\beta)}. \tag{A.58}$$

Using the fact that $L_j = \sum_m \alpha_{mj} L_{mj}$

$$L_j = \sum_m (\alpha_{mj} L_{mj}) = \sum_m \left(N_m \frac{z_j^\beta L_j^{-\beta(1-\eta)} \alpha_{mj}^{1+\beta} g_{mj}^\beta}{\sum_k (z_k^\beta L_k^{-\beta(1-\eta)} \alpha_{mk}^\beta g_{mk}^\beta)} \right). \tag{A.59}$$

Define new variable $H_{j,equ} = L_j z_j^{-\frac{\beta}{1+\beta(1-\eta)}}$ and use $H_{j,equ}$ to substitute for L_j. Simple algebra shows that

$$H_{j,equ} = \left(\sum_m \left(N_m \frac{\alpha_{mj}^{1+\beta} g_{mj}^\beta}{\sum_k (z_k^{\frac{\beta}{1+\beta(1-\eta)}} H_{k,equ}^{-\beta(1-\eta)} \alpha_{mk}^\beta g_{mk}^\beta)} \right) \right)^{\frac{1}{1+\beta(1-\eta)}}. \tag{A.60}$$

Note that $H_{j,equ}$ depends on g_{mj}, and doesn't depend on z_j. Define $\tilde{H}_{j,equ} = H_{j,equ}^{1-\eta}$ and we have

$$\tilde{H}_{j,equ} = \left(\sum_m \left(N_m \frac{\alpha_{mj}^{1+\beta} g_{mj}^{\beta}}{\sum_k (z_k^{\frac{\beta}{1+\beta(1-\eta)}} \tilde{H}_{k,equ}^{-\beta} \alpha_{mk}^{\beta} g_{mk}^{\beta})} \right) \right)^{\frac{1-\eta}{1+\beta(1-\eta)}}. \tag{A.61}$$

We numerically solve the fixed point equation (A.61) to obtain the sequence of $\{H_{j,equ}\}$. [1] We then recover $\{L_j\}$ from $\{H_{j,equ}\}$ using $L_j = H_{j,equ} z_j^{\frac{1}{1+\beta(1-\eta)}}$. We use equation (A.57) and (A.58) to calculate $\{w_{mj}\}$ and $\{L_{mj}\}$.

B.4 Variance Decomposition of $\log w_{mj}$ in an Equilibrium

In our numerical exercise, we internally calibrate five parameters : σ_x, σ_z, σ_g, ρ_{zg} and θ. The moments which we target are from the variance decomposition of the equilibrium log wage, $\log w_{mj}$. Here, we demonstrate how we obtain the decomposition of $\log w_{mj}$.

From the definition of $H_{j,equ}$ we have

$$\log L_j = \log H_{j,equ} + \frac{1}{1+\beta(1-\eta)} \log z_j. \tag{A.62}$$

Substituting expression for L_j to $MR_{mj} = \eta z_j L_j^{\eta-1} \alpha_{mj}$ yields

$$\log MR_{mj} = \log \alpha_{mj} + \frac{1}{1+\beta(1-\eta)} \log z_j + \log \eta - (1-\eta) \log h_j. \tag{A.63}$$

At equilibrium $w_{mj} = \frac{\beta}{1+\beta} MR_{mj}$ thus

$$\log w_{mj} = \log \alpha_{mj} + \frac{1}{1+\beta(1-\eta)} \log z_j + \log \frac{\beta\eta}{1+\beta} - (1-\eta) \log h_j. \tag{A.64}$$

Substituting α_{mj} with $X_m^{(1+\theta(\log z_j - \kappa))}$,

$$\log w_{mj} = \underbrace{\log X_m}_{x_m} \times \underbrace{(1+\theta(\log z_j - \kappa))}_{\theta_j} + \underbrace{\underbrace{\frac{1}{1+\beta(1-\eta)} \log z_j}_{\psi_{zj}} + \underbrace{\log \frac{\beta\eta}{1+\beta} - (1-\eta) \log h_j}_{\psi_{gj}}}_{\psi_j}. \tag{A.65}$$

which is our desired decomposition of $\log w_{mj}$.

[1] Lamadon, Mogstad, and Setzler (2022) provides the proof for the convergence of the fixed point equation (A.61).

B.5 Algorithm for Computing the Optimal Markdown Schedule

To compute the optimal markdown, we use the following formula

$$\mu_j = \frac{\beta}{1+\beta} + \sum_m \frac{\alpha_{mj} L_{mj}}{L_j} \frac{\Delta_m}{\eta z_j L_j^{\eta-1} \alpha_{mj}}, \quad \Delta_m = \frac{1}{1+\beta} \frac{\sum_j z_j L_j^{\eta}}{N + \psi_f J} - \frac{\beta}{1+\beta} \frac{\sum_j (\eta z_j L_j^{\eta-1} \alpha_{mj} - c_{mj}) L_{mj}}{N_m}. \tag{A.66}$$

The algorithm is as follows : First, start with the guess for $\{\mu_j\}$. Second, obtain the values of $\{c_{mj}\}$ and $\{L_{mj}\}$ which are consistent with our guess of the (possibly non-optimal) markdown scheudle $\{\mu_j\}$. Third, using $\{c_{mj}\}$ and $\{L_{mj}\}$, update $\{\mu_j\}$ by calculating the right hand side of equation (A.66). Last, iterate until convergence.

We now explain how to obtain the values of $\{c_{mj}\}$ and $\{L_{mj}\}$ which are consistent with the (possibly non-optimal) markdown schedule $\{\mu_j\}$. In the procedure below, we only make use of the fact that workers optimally choose their preferred workplace after observing the wages and amenities of each firm. Since the markdown is defined as the ratio of wages(which in our case corresponsds to each worker's consumption) to worker's marginal product, we can express c_{mj} as

$$c_{mj} = \mu_j \eta z_j L_j^{\eta-1} \alpha_{mj}. \tag{A.67}$$

Worker's optimization decisions imply that given $\{c_{mj}\}$, type m workers' labor supply to firm j is equal to

$$L_{mj} = N_m \frac{c_{mj}^{\beta} g_{mj}^{\beta}}{\sum_k (c_{mk}^{\beta} g_{mk}^{\beta})} = N_m \frac{\mu_j^{\beta} z_j^{\beta} L_j^{-\beta(1-\eta)} \alpha_{mj}^{\beta} g_{mj}^{\beta}}{\sum_k (\mu_k^{\beta} z_k^{\beta} L_k^{-\beta(1-\eta)} \alpha_{mk}^{\beta} g_{mk}^{\beta})}. \tag{A.68}$$

Using the fact that $L_j = \sum_m \alpha_{mj} L_{mj}$, we have

$$L_j = \sum_m (\alpha_{mj} L_{mj}) = \sum_m \left(N_m \frac{\mu_j^{\beta} z_j^{\beta} L_j^{-\beta(1-\eta)} \alpha_{mj}^{1+\beta} g_{mj}^{\beta}}{\sum_k (\mu_k^{\beta} z_k^{\beta} L_k^{-\beta(1-\eta)} \alpha_{mk}^{\beta} g_{mk}^{\beta})} \right). \tag{A.69}$$

Define new variable $H_{j,opt} = L_j \mu_j^{-\frac{\beta}{1+\beta(1-\eta)}} z_j^{-\frac{\beta}{1+\beta(1-\eta)}}$ and use $H_{j,opt}$ to substitute for L_j. Simple algebra shows that

$$H_{j,opt} = \left(\sum_m \left(N_m \frac{\alpha_{mj}^{1+\beta} g_{mj}^{\beta}}{\sum_k (\mu_k^{\frac{\beta}{1+\beta(1-\eta)}} z_k^{\frac{\beta}{1+\beta(1-\eta)}} H_{k,opt}^{-\beta(1-\eta)} \alpha_{mk}^{\beta} g_{mk}^{\beta})} \right) \right)^{\frac{1}{1+\beta(1-\eta)}} \tag{A.70}$$

Define $\tilde{H}_{j,opt} = H_{j,opt}^{1-\eta}$ and we have

$$\tilde{H}_{j,opt} = \left(\sum_m \left(N_m \frac{\alpha_{mj}^{1+\beta} g_{mj}^{\beta}}{\sum_k (\mu_k^{\frac{\beta}{1+\beta(1-\eta)}} z_k^{\frac{\beta}{1+\beta(1-\eta)}} \tilde{H}_{k,opt}^{-\beta} \alpha_{mk}^{\beta} g_{mk}^{\beta})} \right) \right)^{\frac{1-\eta}{1+\beta(1-\eta)}}. \tag{A.71}$$

We numerically solve the fixed point equation (A.71) to obtain the sequence of $\{H_{j,opt}\}$. [2] We then recover $\{L_j\}$ from $\{H_{j,opt}\}$ (Note that $L_j = H_{j,opt}\mu_j^{\frac{\beta}{1+\beta(1-\eta)}} z_j^{\frac{\beta}{1+\beta(1-\eta)}}$). Finally, we caculate $\{c_{mj}\}$ and $\{L_{mj}\}$ using equation (A.67) and (A.68).

B.6 Numerical Approximation of the Markdown and Decomposition in Figure 4 and 5

We aim to plot the conditional averages of the optimal markdown across the firm productivity distribution :

$$E(\mu_j|z_j) = E\left(\frac{\beta}{1+\beta} + \frac{1}{\eta z_j L_j^{\eta-1}} \sum_m \Delta_m \frac{L_j}{L_j}\right), \tag{A.72}$$

and the decomposition of its slope:

$$\frac{dE(\mu_j|z_j)}{d\log z_j} = F1_z + F1_\alpha + F2, \quad \text{where} \tag{A.73}$$

$$F1_z = -E\left(\left(\mu_j - \frac{\beta}{1+\beta}\right) \times \left[1 - (1-\eta) \times \frac{d\log \bar{L}_j}{d\log z_j}\right]\right), \tag{A.74}$$

$$F1_\alpha = -E\left(\left(\mu_j - \frac{\beta}{1+\beta}\right) \times \left[\eta \times \sum_m \left(\frac{\alpha_{mj} L_{mj}}{\sum_k \alpha_{kj} L_{kj}} \times \frac{d\log \alpha_{mj}}{d\log z_j}\right)\right]\right), \tag{A.75}$$

$$F2 = -E\left(\left(\mu_j - \frac{\beta}{1+\beta}\right) \times \left[\eta \times \sum_m \left(\frac{\alpha_{mj} L_{mj}}{\sum_k \alpha_{kj} L_{kj}} \times \frac{d\log \frac{L_{mj}}{\bar{L}_j}}{d\log z_j}\right) - \sum_m \left(\frac{\Delta_m L_{mj}}{\sum_k \Delta_k L_{kj}} \times \frac{d\log \frac{L_{mj}}{\bar{L}_j}}{d\log z_j}\right)\right]\right). \tag{A.76}$$

We take numerical approximations to plot the conditional averages of the optimal markdown and the decomposition.

We first divide firms into 23 groups. The first group contains firms with productivity from 0th to 5th percentile. The 19th group contains firms with productivity from 90th to 95th percentile. The 20th group contains firms with productivity from 95th to 99th percentile, and the 21th group are firms with productivity from 99th to 99.9th percentile, and so forth. Denote each group of firms as $J_1, \ldots, J_{23}$. We split firm groups more densely at the right tail of the productivity distribution because these firms employ the vast majority of the workforce.

We plot the optimal markdown on 23 gridpoints. For the x-axis of the lth gridpoint, we use the value of $med(\log z_j | j \in J_l)$, where med stands for median. The corresponding value of the optimal markdown is $med(\mu_j | j \in J_l)$.

[2] Lamadon, Mogstad, and Setzler (2022) provides the proof for the convergence of the fixed point equation (A.71).

We plot $F1_z$, $F1_\alpha$ and $F2$ on 22 gridpoints. To determine $F1_z$, we first approximate $\frac{d\log \bar{L}_j}{d\log z_j}$. We approximate $\frac{d\log \bar{L}_j}{d\log z_j}$ at the lth gridpoint as

$$\frac{E(\log \bar{L}_j|j \in J_{l+1}) - E(\log \bar{L}_j|j \in J_l)}{E(\log z_j|j \in J_{l+1}) - E(\log z_j|j \in J_l)}. \tag{A.77}$$

The resulting approximation of $F1_z$ at the lth gridpoint is

$$F1_z^{\text{approx}} = -E\left(\left(\mu_j - \frac{\beta}{1+\beta}\right) \times \left[1 - (1-\eta) \times \frac{E(\log \bar{L}_j|j \in J_{l+1}) - E(\log \bar{L}_j|j \in J_l)}{E(\log z_j|j \in J_{l+1}) - E(\log z_j|j \in J_l)}\right] \Big| \; j \in J_l\right). \tag{A.78}$$

Plotting $F1_\alpha$ at the lth gridpoint is similar. We only present the resulting approximation value.

$$\begin{aligned} F1_\alpha^{\text{approx}} = -E\Bigg(&\left(\mu_j - \frac{\beta}{1+\beta}\right) \times \\ &\eta \times \sum_m \left(\frac{\alpha_{mj} L_{mj}}{L_j} \times \frac{E(\log \alpha_{mj}|j \in J_{l+1}) - E(\log \alpha_{mj}|j \in J_l)}{E(\log z_j|j \in J_{l+1}) - E(\log z_j|j \in J_l)}\right) \Big| \; j \in J_l\Bigg). \end{aligned} \tag{A.79}$$

The approximated value of $F2$ at the lth gridpoint is given by

$$\begin{aligned} F2^{\text{approx}} = -E\Bigg(&\left(\mu_j - \frac{\beta}{1+\beta}\right) \times \\ &\sum_m \left[\left(\eta \frac{\alpha_{mj} L_{mj}}{L_j} - \frac{\Delta_n L_{nj}}{\sum_k \Delta_k L_{kj}}\right) \times \frac{E(\log L_{mj}|j \in J_{l+1}) - E(\log \bar{L}_j|j \in J_l)}{E(\log z_j|j \in J_{l+1}) - E(\log z_j|j \in J_l)}\right] \Big| \; j \in J_l\Bigg). \end{aligned} \tag{A.80}$$

B.7 Sensitivity Analysis : β, η and κ

In this section, we do sensitivity analysis with respect to labor supply elasticity β, returns to scale η and parameter controlling the overall wage variance κ. For the labor supply elasticity, we present results for three different values $\beta = 6, 5.25, 4.5$. For each value of β we recalibrate the parameters $\sigma_z, \sigma_g, \rho_{zg}, \sigma_x$ and θ to match the calibration targets $Var(\psi_z) = 0.17, Var(\psi_g) = 0.13, 2Cov(\psi_z, \psi_g) = -0.27, Var(x) = 0.31, Corr(x, \psi) = 0.37$.

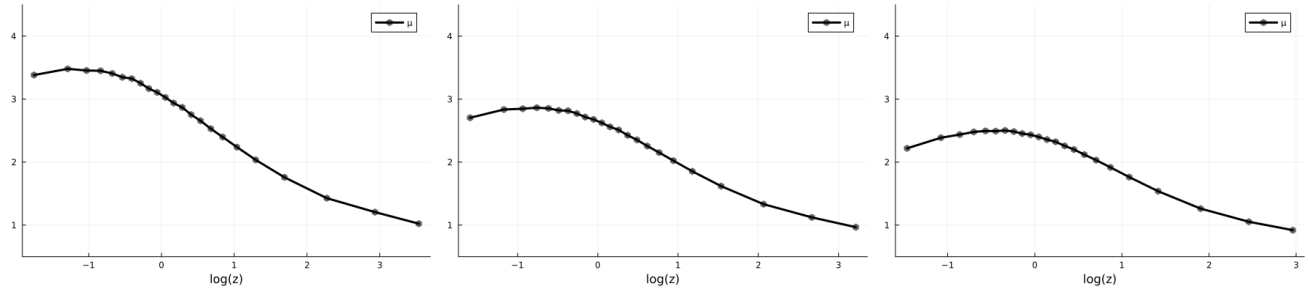
The calibrated parameter values are presented on Table 1. As labor supply becomes less elatic, workers become less sensitive to wage differences across firms. The variance of firm productivity should decrease in order to match the observed degree of dispersion of firm pay premiums. More importantly, we need a higher value of skill intensity heterogeneity θ to match the observed pattern of worker-firm sorting, $Corr(x, \psi) = 0.37$. Larger differences in within-firm relative wages(between differently skilled workers) across firms are required to induce high-skilled workers sort into higher paying firms.

Figure A.1a, A.1b, and A.1c plots the optimal markdown schedule for different values of β. As β decreases, the inverted U-shape of the optimal markdown schedule below the median productivity

Parameter	Moment	$\beta = 6$	$\beta = 5.25$	$\beta = 4.5$
σ_z	$Var(\psi_z)$	0.8952	0.8139	0.7497
σ_g	$Var(\psi_g)$	0.4346	0.4376	0.4542
ρ_{zg}	$2Cov(\psi_z, \psi_g)$	0.8995	0.9043	0.9199
σ_x	$Var(x)$	0.5567	0.5567	0.5567
θ	$Corr(x, \psi)$	0.3631	0.4368	0.5408

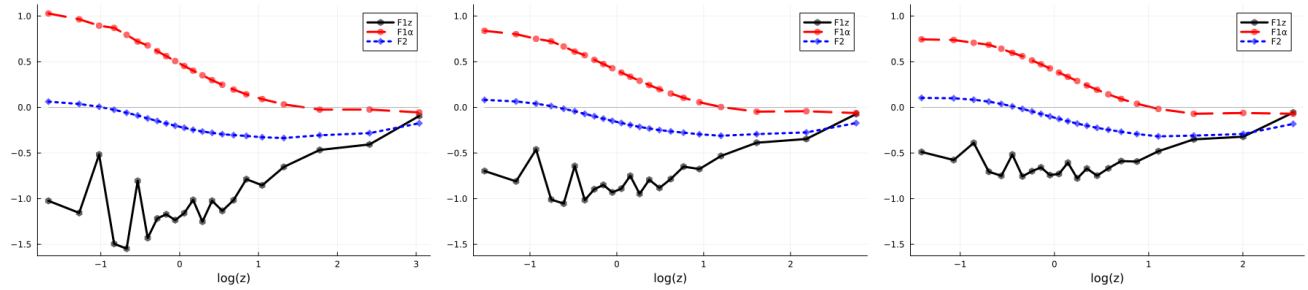
Table 1: Calibrated Parameters, β

firm becomes more pronounced. Figure A.2a, A.2b, and A.2c plots the decomposition of the slope of the optimal markdown schedule for values of $\beta = 6.00, 5.25, 4.50$. Observe that as labor supply elasticity becomes smaller, the magnitude of $F1_\alpha$ term relative to $F1_z$ becomes larger. With a higher calibrated value of θ, the degree of across-firm wage dispersion becomes much larger for high-skilled workers compared to low-skilled workers. As a result, we observe a stronger degree of non-monotonicity at the left tail of the productivity distribution.



(a) $\beta = 6.00$ (b) $\beta = 5.25$ (c) $\beta = 4.50$

Figure A.1: Optimal Markdown, varying β



(a) $\beta = 6.00$ (b) $\beta = 5.25$ (c) $\beta = 4.50$

Figure A.2: Decomposition, varying β

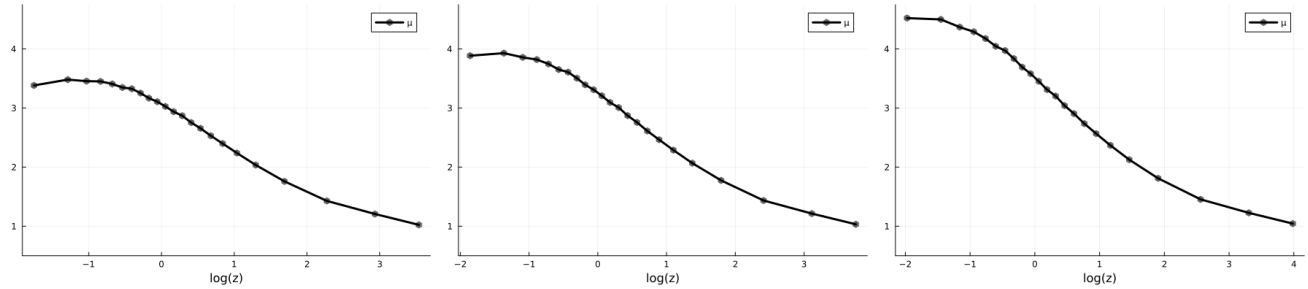
For returns to scale, we present results for three different values $\eta = 0.750, 0.725, 0.700$. For each value of η we recalibrate the parameters $\sigma_z, \sigma_g, \rho_{zg}, \sigma_x$ and θ to match $Var(\psi_z) = 0.17, Var(\psi_g) =$

$0.13, 2Cov(\psi_z, \psi_g) = -0.27, Var(x) = 0.31, Corr(x, \psi) = 0.37$. The calibrated parameter values are presented on Table 2. Note that under higher degree of decreasing returns to scale, larger productivity variance σ_z is needed to match the observed dispersion of firm pay premiums.

Parameter	Moments	$\eta = 0.750$	$\eta = 0.725$	$\eta = 0.700$
σ_z	$Var(\psi_z)$	0.8952	0.9499	1.0084
σ_g	$Var(\psi_g)$	0.4346	0.4102	0.3916
ρ_{zg}	$2Cov(\psi_z, \psi_g)$	0.8995	0.9002	0.9020
σ_x	$Var(x)$	0.5567	0.5567	0.5567
θ	$Corr(x, \psi)$	0.3654	0.3389	0.3173

Table 2: Calibrated Parameters, η

Figure A.3a, A.3b, and A.3c plots the optimal markdown schedule for values of $\eta = 0.750, 0.725, 0.700$. Observe that as η decreases, the optimal markdown schedule becomes more steeper. Figure A.4a, A.4b, and A.4c plots the decomposition of the slope of the optimal markdown schedule for values of $\eta = 0.750, 0.725, 0.700$. The steeper slope with a lower η is due to the larger magnitude of $F1_z$. With higher degree of decreasing returns to scale, the dispersion of marginal revenue productivity at the optimum allocation becomes larger, which contributes to a steeper optimal markdown schedule.

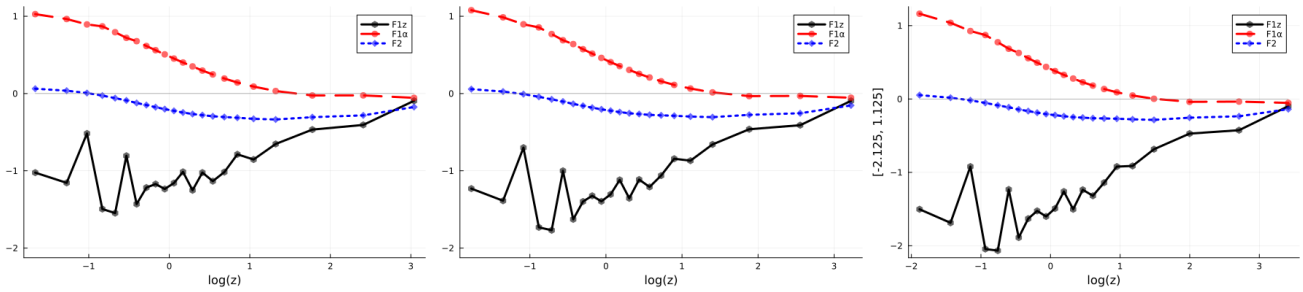


(a) $\eta = 0.750$ (b) $\eta = 0.725$ (c) $\eta = 0.700$

Figure A.3: Optimal Markdown, varying η

We proceed to the sensitivity analysis with respect to κ. κ is a parameter which determines the overall workers' wage variance. The closer the value of κ is to 0, the larger the variance of workers' wages. We use the difference of predicted log earnings between the 90th percentile productivity worker and the 10th percentile productivity worker at a median productivity firm as a measure of the overall wage dispersion. [3] We present the model-computed and data values of our measure of wage dispersion for each value of κ at Table 3. It shows that our choice of $\kappa = 1.5$ well approximates

[3]We divide firms into 10 groups according to their productivities. We denote median productivity firm as firms in the 5th and the 6th group.



(a) $\eta = 0.750$ (b) $\eta = 0.725$ (c) $\eta = 0.700$

Figure A.4: Decomposition, varying η

the degree of wage dispersion in the data. We present sensitivity analaysis results for three different values $\kappa = 1.0, 1.5, 2.0$.

Moment	Data	$\kappa = 1.0$	$\kappa = 1.5$	$\kappa = 2.0$
$\mathbb{E}(\log w_{90} \mid j = 5) - \mathbb{E}(\log w_{10} \mid j = 5)$	1.0838	1.3985	0.9719	0.5147
$\mathbb{E}(\log w_{90} \mid j = 6) - \mathbb{E}(\log w_{10} \mid j = 6)$	1.3794	1.5805	1.1586	0.7094

Table 3: Measure of wage dispersion, κ

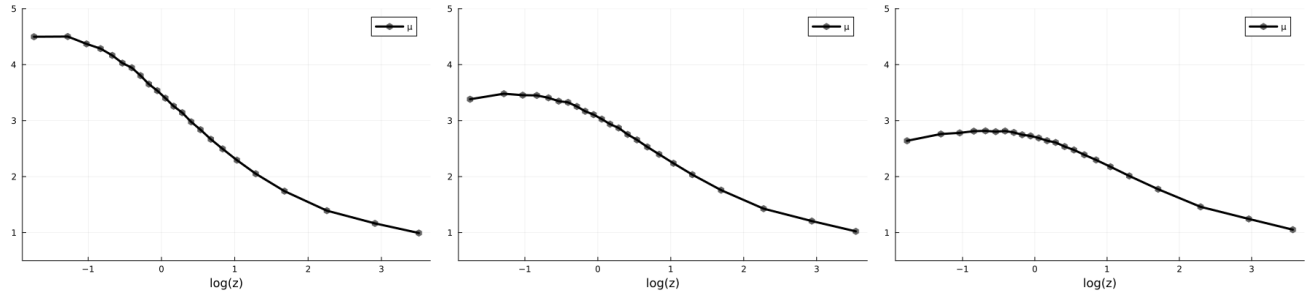
For each value of κ we recalibrate the parameters $\sigma_z, \sigma_g, \rho_{zg}, \sigma_x$ and θ to match $Var(\psi_z) = 0.17, Var(\psi_g) = 0.13, 2Cov(\psi_z, \psi_g) = -0.27, Var(x) = 0.31, Corr(x, \psi) = 0.37$. The calibrated parameter values are presented on Table 4. Observe that the calibrated values are roughly similar regardless of our choice of κ. Since we target the variance of firm pay premium in calibration, the degree of across-firm wage inequality for identically skilled workers is roughly similar for different values of κ. Thus, as the value of κ decreases (indicating a larger overall wage variance), the proportion of across-worker type wage inequality in explaining the overall wage variance increases, and the proportion of across-firm wage inequality for given worker type decreases.

Parameter	Moments	$\kappa = 1.0$	$\kappa = 1.5$	$\kappa = 2.0$
σ_z	$Var(\psi_z)$	0.8887	0.8952	0.9036
σ_g	$Var(\psi_g)$	0.4204	0.4346	0.4505
ρ_{zg}	$2Cov(\psi_z, \psi_g)$	0.8934	0.8995	0.9068
σ_x	$Var(x)$	0.5567	0.5567	0.5567
θ	$Corr(x, \psi)$	0.3631	0.3654	0.3727

Table 4: Calibrated Parameters, κ

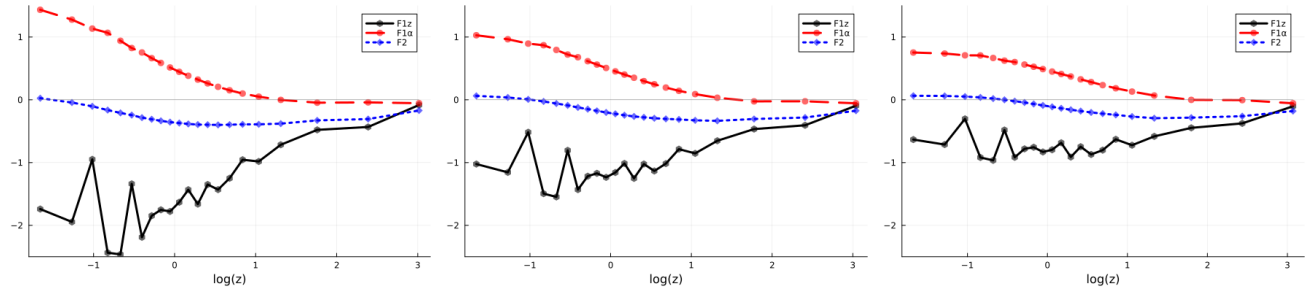
Figure A.5a, A.5b, and A.5c plots the optimal markdown schedule for values of $\kappa = 1.0, 1.5, 2.0$. Figure A.6a, A.6b, and A.6c plots the decomposition of the slope of the optimal markdown schedule

for values of $\kappa = 1.0, 1.5, 2.0$. As κ increases, the overall degree of wage inequality across workers becomes less severe, leading to less steeper markdown schedule across the firm productivity distribution.



(a) $\kappa = 1.0$ (b) $\kappa = 1.5$ (c) $\kappa = 2.0$

Figure A.5: Optimal Markdown, varying κ



(a) $\kappa = 1.0$ (b) $\kappa = 1.5$ (c) $\kappa = 2.0$

Figure A.6: Decomposition, varying κ

B.8 Sensitivity Analysis : Imperfect Substitution between Worker Type

In the model we use for numerical exercise, we assumed perfect substitution between worker type. In this section, we allow imperfect substitution between worker type and do the sensitivity analysis. Specifically, we assume that the production function of firm j takes the form

$$y_j = z_j L_j^{\eta}, \quad L_j = \left(\sum_m \alpha_{mj} L_{mj}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}, \tag{A.81}$$

where y_j is the output produced, z_j is productivity, L_j is efficiency unit of labor, α_{mj} is firm-specific labor productivity, L_{mj} is the number of type m workers employed at firm j, and σ is the elasticity of substitution. Note that $\sigma = \infty$ corresponds to the perfect substitution case.

We employ the following calibration strategy. We externally calibrate labor supply elasticity $\beta = 6.0$, returns to scale $\eta = 0.75$, and $\kappa = 1.50$. We report results for three parameter values of

elasticity of substitution σ, 5,10 and ∞. We make the simplifying assumption that all firms provide the same level of amenities, $g_j = g$, and thus firms only vary with respect to their productivities z_j. We set the standard deviation of log worker skill $\sigma_x = 0.56$ to match the variance of log worker skill $Var(\log X_m) = 0.31$. We internally calibrate the standard deviation of firm productivity σ_z and the parameter governing the degree of skill intensity heterogeneity θ. We target the variance of firm pay premium $Var(\psi_j) = 0.02$ to calibrate σ_z and the correlation between log worker skill and firm pay premium $Corr(x_m, \psi_j) = 0.37$ to calibrate θ.[4]

Parameter	$\sigma = \infty$	$\sigma = 10$	$\sigma = 5$
σ_z	0.4205	0.4244	0.3085
θ	0.2994	0.4751	0.9389

Table 5: Calibrated Parameters, Imperfect Substitution

The resulting parameter values are presented on Table 5. As σ gets smaller, the calibrated value of θ becomes larger. When workers become less substitutable, even for the firms with high labor productivity for skilled workers the productivity gains from employing low-skilled workforce becomes larger. Thus to generate the degree of sorting observed in the data $Corr(x, \psi) = 0.37$, higher degree of skill intensity heterogeneity θ is required.

We state the optimal markdown and its decomposition formula for imperfect substitution case below. The optimal markdown μ_j satisfies

$$\mu_j = \frac{\beta}{1+\beta} + \sum_m \frac{\alpha_{mj} L_{mj}^{\frac{\sigma-1}{\sigma}}}{L_j^{\frac{\sigma-1}{\sigma}}} \frac{\Delta_m}{MR_{mj}}, \tag{A.82}$$

where $\Delta_m = \frac{1}{\lambda(1+\beta)}\left(1 - \nu_m \frac{I_m}{N_m}\right) = \frac{1}{\lambda(1+\beta)}\left(1 + \beta\lambda\frac{\sigma-1}{\sigma}\frac{\sum_j c_{mj} L_{mj}}{N_m} - \beta\lambda\frac{\sum_j MR_{mj} L_{mj}}{N_m} + \frac{\beta}{\sigma}\right)$, $\lambda = \left(\frac{\sum_j y_j}{N + \psi_f J}\right)^{-1}$, and $MR_{mj} = \eta z_j L_j^{\eta-1} \alpha_{mj} \left(\frac{L_{mj}}{L_j}\right)^{-\frac{1}{\sigma}}$ is the marginal product of type m worker at

[4]The equilibrium log wage is $\log w_{mj} = \log \frac{\beta\eta}{1+\beta} + \log z_j - (1-\eta)\log L_j + \log \alpha_{mj} - \frac{1}{\sigma}\log \frac{L_{mj}}{L_j}$. We define firm pay premium ψ_j as $\log \frac{\beta\eta}{1+\beta} + \log z_j - (1-\eta)\log L_j$.

firm j. The decomposition of the slope of the optimal markdown, $\frac{d\mu_j}{d\log z_j}$ is given as

$$\frac{d\mu_j}{d\log z_j} = F1_z + F1_\alpha + F2, \quad \text{where} \tag{A.83}$$

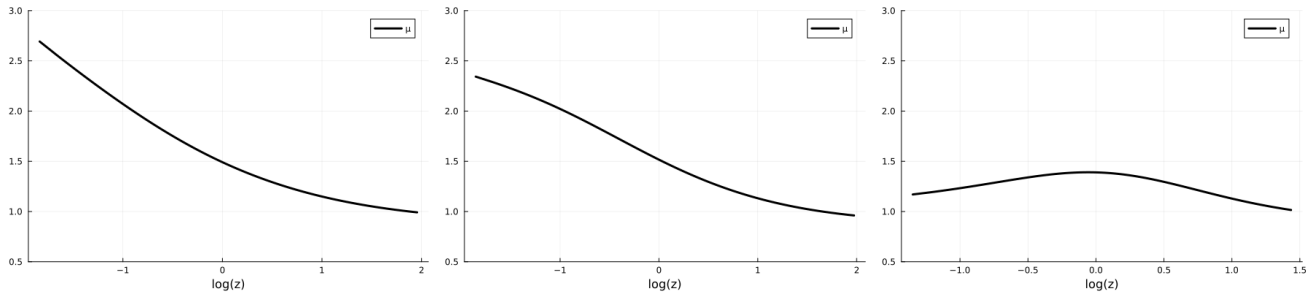
$$F1_z = -\left(\mu_j - \frac{\beta}{1+\beta}\right) \times \left(1 - (1-\eta) \times \frac{d\log \bar{L}_j}{d\log z_j}\right), \tag{A.84}$$

$$F1_\alpha = -\left(\mu_j - \frac{\beta}{1+\beta}\right) \times \frac{\eta\sigma}{\sigma-1} \times \sum_m \left(\frac{\alpha_{mj} L_{mj}^{\frac{\sigma-1}{\sigma}}}{L_j^{\frac{\sigma-1}{\sigma}}} \times \frac{d\log \alpha_{mj}}{d\log z_j}\right), \tag{A.85}$$

$$F2 = -\left(\mu_j - \frac{\beta}{1+\beta}\right) \times \sum_m \left(\eta \times \frac{\alpha_{mj} L_{mj}^{\frac{\sigma-1}{\sigma}}}{L_j^{\frac{\sigma-1}{\sigma}}} \times \frac{d\log \frac{L_{mj}}{\bar{L}_j}}{d\log z_j} - \frac{\Delta_m L_{mj}}{\sum_k \Delta_k L_{kj}} \times \frac{d\log \frac{L_{mj}}{\bar{L}_j}}{d\log z_j}\right). \tag{A.86}$$

The proof is essentially the same as in the case of perfect substitution.

The computational algorithm uses the equation (A.82) as an updating rule to find a fixed-point of the optimal markdown schedule. [5] Figure A.7a, A.7b, and A.7c plots the optimal markdown



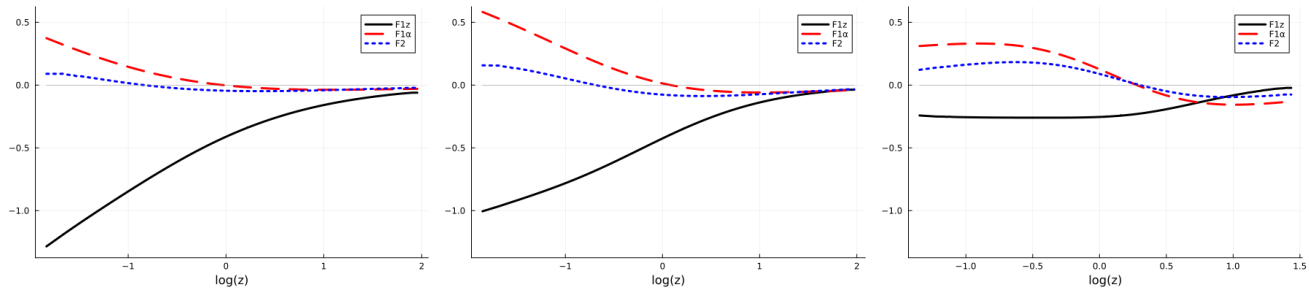
(a) $\sigma = \infty$ (b) $\sigma = 10$ (c) $\sigma = 5$

Figure A.7: Optimal Markdown, varying σ

schedule for values of $\sigma = \infty, 10, 5$. As workers becomes less substitutable, the optimal markdown schedule becomes less steeper, particularly at the left tail of the productivity distribution. For $\sigma = 5$ case, the optimal markdown is increasing at at the left tail of the productivity distribution, resulting in an inverted U shape.

Figure A.8a, A.8b, and A.8c plots the decomposition of the slope of the optimal markdown schedule for values of $\sigma = \infty, 10, 5$. The most noticeable difference is that the magnitude of the $F1_\alpha$ relative to $F1_z$ becomes larger under lower values of σ. It is the $F1_\alpha$ term which makes the optimal markdown schedule increasing at the left tail of the productivity distribution under $\sigma = 5$. Recall that under lower values of σ we need larger value of θ to match the observed sorting pattern

[5] Since we impose $g_j = g, \forall j$, the markdown schedule and the decomposition of its slope is smooth in $\log z_j$, so we don't need approximations to plot the graph.



(a) $\sigma = \infty$ (b) $\sigma = 10$ (c) $\sigma = 5$

Figure A.8: Decomposition, varying σ

of workers to firms $Corr(x, \psi) = 0.37$. As the degree of skill intensity hetergeneity gets larger, the magnitude of the $F1_\alpha$ term, which reflects the indirect effect of increasing z_j on μ_j through firm-specific labor productivity α_{mj}, becomes larger.

C Appendix for Section 6

C.1 Proofs of Proposition and Equations in Section 6

Proof of Proposition 13

As we have discussed in the main text, with single type only (or with skill-dependent policy), there is no restriction on the policy tool, thus regardless of which policy tool the government uses, the optimal consumption should satisfy the same formula: $c_j = \frac{\beta}{1+\beta} \cdot MP_{Lj} + \Delta$. The implementation with only revenue dependent subsidy to the firm was already proved in the main text of Section 3. Here we prove the other two cases in Proposition 13.

Consider an implementation combining lumpsum transfers to the workers and subsidies to the firm. Given some positive lumpsum transfers $0 < Tr < \Delta$ to the workers, consumption of a worker at firm j is $c_j = w_j + Tr$, which yields the labor supply curve: $L_j = \frac{c_j^\beta}{\sum_j c_j^\beta} \times N = \frac{(w_j+Tr)^\beta}{\sum_j (w_j+Tr)^\beta} \times N$. Thus the elasticity of labor supply given the lumpsum transfers Tr is

$$\frac{\partial \log L_j}{\partial \log w_j} = \beta \times \frac{w_j}{w_j + Tr}. \tag{A.87}$$

Then the firm's problem given subsidy schedule $S_j(y_j)$ subject to the elasticity of labor supply (A.87) should satisfy the following first order condition:

$$w_j = \left(1 + S_j'(y_j)\right) \times \frac{\beta \cdot \frac{w_j}{w_j+Tr}}{1 + \beta \cdot \frac{w_j}{w_j+Tr}} \times MP_{Lj}.$$

Since the optimal allocation satisfies $\frac{w_j+Tr}{MP_{Lj}} = \frac{c_j}{MP_{Lj}} = \frac{\beta}{1+\beta} + \frac{\Delta}{MP_{Lj}}$, the subsidy schedule which implements the optimal allocation is set to satisfy

$$(1+S_j'(y_j)) \times \frac{\beta \cdot \frac{w_j}{w_j+Tr}}{1+\beta \cdot \frac{w_j}{w_j+Tr}} + \frac{Tr}{MP_{Lj}} = \frac{\beta}{1+\beta} + \frac{\Delta}{MP_{Lj}}.$$

Rewriting for the subsidy rate, we obtain $1+S_j'(y_j) = \frac{1}{\beta}\left(\frac{\beta}{1+\beta} + \frac{\Delta-Tr}{MP_{Lj}}\right) \times \left(\beta + \frac{\frac{\beta}{1+\beta}MP_{Lj}+\Delta}{\frac{\beta}{1+\beta}MP_{Lj}+\Delta-Tr}\right)$.

Then it is straightforward to show that the subsidy rate decreases in firm's productivity:

$$\begin{aligned}
\frac{\partial S_j'(y_j)}{\partial z_j} &= -\frac{1}{\beta} \cdot \frac{\Delta-Tr}{MP_{Lj}^2} \cdot \frac{\partial MP_{Lj}}{\partial z_j} \times \left(\beta + \frac{\frac{\beta}{1+\beta}MP_{Lj}+\Delta}{\frac{\beta}{1+\beta}MP_{Lj}+\Delta-Tr}\right) \\
&\quad + \frac{1}{\beta}\left(\frac{\beta}{1+\beta} + \frac{\Delta-Tr}{MP_{Lj}}\right) \times \frac{\beta}{1+\beta} \cdot \frac{\partial MP_{Lj}}{\partial z_j} \times \frac{-Tr}{\left(\frac{\beta}{1+\beta}MP_{Lj}+\Delta-Tr\right)^2} \\
&= -\frac{1}{\beta} \cdot \frac{\Delta-Tr}{MP_{Lj}^2} \cdot \frac{\partial MP_{Lj}}{\partial z_j}\left(\beta + \frac{\frac{\beta}{1+\beta}+\frac{\Delta}{MP_{Lj}}}{\frac{\beta}{1+\beta}+\frac{\Delta-Tr}{MP_{Lj}}}\right) - \frac{1}{1+\beta} \cdot \frac{1}{MP_{Lj}^2} \cdot \frac{\partial MP_{Lj}}{\partial z_j} \cdot \frac{Tr}{\frac{\beta}{1+\beta}+\frac{\Delta-Tr}{MP_{Lj}}} \\
&= -\frac{1}{MP_{Lj}^2} \cdot \frac{\partial MP_{Lj}}{\partial z_j} \times \frac{(\Delta-Tr)\cdot\left(1+\frac{\frac{1+\beta}{\beta}\delta-Tr}{MP_{Lj}}\right)+\frac{1}{1+\beta}Tr}{\frac{\beta}{1+\beta}+\frac{\Delta-Tr}{MP_{Lj}}} < 0.
\end{aligned}$$

In this implementation, the aggregate lumpsum transfers should be funded by taxing the firms, which implies that the sum of subsidies to the firms should be negative. Note that $\pi_j = y_j + S_j(y_j) - w_jL_j$ and $c_j = w_j - +Tr$. Thus,

$$\begin{aligned}
\sum_j S_j(y_j) &= \sum_j \pi_j - \sum_j y_j + \sum_j w_jL_j \\
&= \sum_j c_j^e - \sum_j y_j + \sum_j c_jL_j - Tr \cdot \sum_j L_j \\
&= -N \times Tr,
\end{aligned}$$

where the last equality uses the resource constraint.

Next, we consider an implementation of combining the progressive labor income tax and subsidies to the firms. Suppose that the income tax is given with the HSV functional form, $T(y) = y - \lambda \cdot y^{1-\tau}$ for some progressivity τ, where $\lambda(\tau)$ is set to satisfy the government budget constraint. Then the consumption of a worker at firm j is $c_j = \lambda \cdot w_j^{1-\tau}$, which yields the following labor supply curve: $L_j = \frac{c_j^\beta}{\sum_j c_j^\beta} \times N = \frac{w_j^{(1-\tau)\beta}}{\sum_j w_j^{(1-\tau)\beta}} \times N$. Then solving the firm's problem under the subsidy schedule $S_j(y_j)$ subject to the elasticity of labor supply $\frac{\partial \log L_j}{\partial \log w_j} = (1-\tau)\beta$ yields

$$w_j = \left(1+S_j'(y_j)\right) \times \frac{(1-\tau)\beta}{1+(1-\tau)\beta} \times MP_{Lj}.$$

Since the optimal allocation satisfies $\frac{\lambda \cdot w_j^{1-\tau}}{MP_{Lj}} = \frac{c_j}{MP_{Lj}} = \frac{\beta}{1+\beta} + \frac{\Delta}{MP_{Lj}}$, the subsidy schedule which implements the optimal allocation is set to satisfy

$$\lambda \cdot \left(1 + S_j'(y_j)\right)^{1-\tau} \times \left(\frac{(1-\tau)\beta}{1+(1-\tau)\beta}\right) \times MP_{Lj}^{-\tau} = \frac{\beta}{1+\beta} + \frac{\Delta}{MP_{Lj}}.$$

Rewriting for the subsidy rate, we obtain $1 + S_j'(y_j) = \left(\frac{1}{\lambda}\right)^{\frac{1}{1-\tau}} \times \frac{1+(1-\tau)\beta}{(1-\tau)\beta} \times \left(\frac{\beta MP_{Lj}^{\tau}}{1+\beta} + \frac{\Delta}{MP_{Lj}^{1-\tau}}\right)^{\frac{1}{1-\tau}}$.

Then the derivative of the subsidy rate with respect to firm's productivity is

$$\frac{\partial S_j'(y_j)}{\partial z_j} = \left(\frac{1}{\lambda}\right)^{\frac{1}{1-\tau}} \frac{1+(1-\tau)\beta}{(1-\tau)\beta} \cdot \frac{1}{1-\tau} \cdot \left(\frac{\beta MP_{Lj}^{\tau}}{1+\beta} + \frac{\Delta}{MP_{Lj}^{1-\tau}}\right)^{\frac{\tau}{1-\tau}} \cdot \frac{1}{MP_{Lj}^{1-\tau}} \left(\frac{\tau\beta}{1+\beta} - \frac{(1-\tau)\Delta}{MP_{Lj}}\right) \frac{\partial MP_{Lj}}{\partial z_j},$$

which is negative under the assumption of $\frac{\tau}{1-\tau} < \frac{1+\beta}{\beta} \cdot \frac{\Delta}{MP_{Lj}}$. That is, if the progressivity is not too big, the subsidy rate schedule still decreases in firm's productivity.

It is straightforward to show that the subsidies to the firm sum to zero:

$$\begin{aligned} \sum_j S_j(y_j) &= \sum_j \pi_j - \sum_j y_j + \sum_j w_j L_j \\ &= \sum_j c_j^e - \sum_j y_j + \sum_j w_j L_j = 0. \end{aligned}$$

Derivation of the Labor Supply and NDC under the HSV Income Tax

First, we consider an economy with extensive margin of labor supply only. Under the HSV income tax, s-skilled worker's utility working at firm j is $\log\left(\lambda \cdot w_{sj}^{1-\tau}\right) + \epsilon_{ij}$, from which it is straightforward to derive the labor supply curve of a skill type s at a firm j: $N_s = \frac{w_{sj}^{\beta(1-\tau)}}{\sum_k w_{sk}^{(1-\beta)\tau}}$. Then by solving the j-firm's problem given subsidy schedule, we obtain

$$\left(1 + S_j'(y_j)\right) \times \frac{(1-\tau)\beta}{1+(1-\tau)\beta} = \frac{w_{sj}}{MP_{sj}}, \quad s = h, \ell,$$

which shows that under the skill-independent subsidy the markdown within the same firm should be the same for both skill types. This leads to the NDC: $\frac{w_{hj}}{MP_{hj}} = \frac{w_{\ell j}}{MP_{\ell j}}$. We then rewrite the NDC in terms of consumption using that $c_{sj} = \lambda \cdot w_{sj}^{1-\tau}$:

$$\frac{c_{hj}}{MP_{hj}^{1-\tau}} = \frac{c_{\ell j}}{MP_{\ell j}^{1-\tau}}.$$

Next, we consider an economy with intensive labor supply under the general preferences. Worker's choice over the employers is the same as before. The workers draw idiosyncratic preferences $\{\epsilon_{i,j}\}$ over the firms which follows a Gumbel distribution with scale parameter $\frac{1}{\beta}$. Given the value function $V_s(w_j)$ defined above, worker's problem of choosing a firm is:

$$V_s = E\left[\max_j V_{sj}(w_{sj}) + \epsilon_{ij}\right].$$

Then the probability of s-type worker's choosing firm j is

$$P_{sj} = \frac{e^{\beta \cdot V_s(w_{sj})}}{\sum_k e^{\beta \cdot V_s(w_{sk})}}.$$

The labor supply curve of the firm j for s-skilled workers (see the Appendix for the derivation):

$$L_{sj}(w_{sj}) = N_{sj}(w_{sj}) \times h_{sj}(w_{sj}) = P_{sj} \cdot N_s \times h_{sj}(w_{sj}) = \frac{e^{\beta \cdot V_s(w_{sj})}}{\sum_k e^{\beta \cdot V_s(w_{sk})}} \cdot N_s \times h_{sj}(w_{sj}).$$

To derive the NDC, we solve the firm's problem under the labor supply curve obtained above. Given a skill-independent subsidy schedule $S_j(y_j)$, firm j solves

$$\begin{aligned} \pi_j \quad &= \quad \max_{L_{hj}, L_{\ell j}} F(L_{h_j}, L_{\ell j}) + S_j(F(L_{h_j}, L_{\ell j})) - w_{hj} L_{h_j} - w_{\ell j} L_{\ell j} \\ s.t. \qquad & L_{sj} = \frac{e^{\beta \cdot V_s(w_{sj})}}{\sum_k e^{\beta \cdot V_s(w_{sk})}} \cdot N_s \times h_{sj}(w_{sj}). \end{aligned}$$

The first order condition of the firm's problem yields $(1 + S_j'(y_j)) = \frac{w_{sj}}{MPsj} \times \frac{1+\epsilon_s(w_{sj})}{\epsilon_s(w_{sj})}$, which leads to the following no-differentiation constraint:

$$\textbf{(NDC-}\tau\textbf{):} \quad \frac{w_{hj}}{MP_{hj}} \times \frac{1 + \epsilon_h(w_{hj})}{\epsilon_h(w_{hj})} = \frac{w_{\ell j}}{MP_{\ell j}} \times \frac{1 + \epsilon_\ell(w_{\ell j})}{\epsilon_h(w_{\ell j})}.$$

Optimal Joint Policy under the General Preferences

In Section 6.2, we have started with the planner's problem under the general preferences, we have restricted our attention to the special preferences to have a sharp characterization of the optimal allocation and optimal joint policy. Here, we show that under the general preferences, there can be additional channels through which increase in progressivity affects optimal allocation. As in the main text, we take the HSV functional form of progressive labor income tax and analyze the effect of increase in progressivity parameter τ.

For expositional simplicity, we focus on the case with only single type of workers. Then the value function of the worker, worker's optimal working hours, and the probability of worker's choice of firm j defined in Section 6.2 do not depend on worker's skill type, and we denote them by $V(w_j)$, $h_j(w_j)$ and P_j respectively. Since we focus on the planner who uses the HSV functional form for the labor income tax, the planner chooses the distribution of wage $\{w_j\}$ and progressivity τ, which automatically determines consumption distribution $\left\{c_j = \lambda(\tau) \times (w_j \cdot h_j(w_j))^{1-\tau}\right\}$. We write down

the planner's problem as follows:

$$\begin{aligned}
\max_{\{w_j,N_j\},\tau} \quad & N\cdot\frac{1}{\beta}\log\left(\sum_j e^{\beta\cdot V(w_j)}\right)+\psi_f\sum_j u(c_j^e)\\
(\chi)\quad s.t.\quad & \sum_j \lambda(\tau)\cdot w_j^{1-\tau}h_j(w_j)^{1-\tau}N_j+\sum_j c_j^e=\sum_j z_j\left(N_j\cdot h_j(w_j)\right)^{\eta}\\
(\xi_j)\quad & N_j=\frac{e^{\beta\cdot V(w_j)}}{I}\times N,\quad \forall j\\
(\gamma)\quad & \sum_j e^{\beta\cdot V(w_j)}=I.
\end{aligned}$$

The first order condition with respect to w_j is

$$\begin{aligned}
&\frac{N}{\beta}\cdot\frac{e^{\beta\cdot V(w_j)}}{\sum_j e^{\beta\cdot V(w_j)}}\cdot\beta V'(w_j)+\chi\cdot\eta z_j N_j^{\eta}h_j^{\eta-1}h_j'(w_j)-\chi\cdot\lambda(1-\tau)w_j^{-\tau}h_j^{1-\tau}N_j\\
&-\chi\cdot\lambda(1-\tau)w_j^{1-\tau}h_j^{-\tau}h_j'(w_j)N_j+\xi_j\cdot e^{\beta\cdot V(w_j)}\beta V'(w_j)\frac{N}{I}-\gamma\cdot e^{\beta\cdot V(w_j)}\beta V'(w_j)=0. \quad (A.88)
\end{aligned}$$

And the first order condition with respect to N_j is

$$\chi\cdot\eta z_j N_j^{\eta-1}h_j^{\eta}-\chi\cdot\lambda w_j^{1-\tau}h_j^{1-\tau}-\xi_j=0. \quad (A.89)$$

We multiply the equation (A.88) by $\frac{w_j}{N_j}$ and substitute ξ_j using (A.89), which yields

$$\begin{aligned}
&V'(w_j)w_j+\chi\cdot MP_{Lj}h_j'(w_j)-\chi\cdot\lambda(1-\tau)w_j^{1-\tau}h_j^{1-\tau}-\chi\cdot\lambda(1-\tau)w_j^{1-\tau}h_j^{1-\tau}\frac{h_j'(w_j)w_j}{h_j}\\
&+\beta V'(w_j)w_j\times\left(\chi\cdot MP_{Lj}h_j-\chi\cdot\lambda w_j^{1-\tau}h_j^{1-\tau}\right)-\gamma\frac{I}{N}\beta V'(w_j)w_j=0.
\end{aligned}$$

Rewriting the equation for consumption, we obtain

$$c_j \;=\; \lambda\cdot w_j^{1-\tau}h_j^{1-\tau}=\frac{\frac{h_j'(w_j)w_j}{h_j}+\beta V'(w_j)w_j}{(1-\tau)\left(1+\frac{h_j'(w_j)w_j}{h_j}+\frac{\beta V'(w_j)}{1-\tau}\right)}\cdot MP_{Lj}\cdot h_j+\frac{\frac{1+\beta}{\beta}\cdot\beta V'(w_j)w_j}{(1-\tau)\left(1+\frac{h_j'(w_j)w_j}{h_j}+\frac{\beta V'(w_j)}{1-\tau}\right)}\Delta,$$

where $\Delta=\frac{1-\beta\gamma\frac{I}{N}}{\chi(1+\beta)}$. From the main text, we know that $\epsilon^P(w_j)=\beta V'(w_j)w_j$ and $\epsilon^h(w_j)=\frac{h_j'(w_j)w_j}{h_j}$, thus the optimal consumption formula yields

$$c_j=\frac{\epsilon^h(w_j)+\epsilon^P(w_j)}{(1-\tau)\left(1+\epsilon^h(w_j)+\frac{\epsilon^P(w_j)}{1-\tau}\right)}\cdot MP_{Lj}\cdot h_j+\frac{\frac{\beta}{1+\beta}\cdot\epsilon^P(w_j)}{(1-\tau)\left(1+\epsilon^h(w_j)+\frac{\epsilon^P(w_j)}{1-\tau}\right)}\Delta.$$

Note that with the special preferences in the main text, $\epsilon^h(w_j)=0$ and $\epsilon^P(w_j)=\beta(1-\tau)$, and thus the optimal consumption formula boils down to $c_j=\frac{\beta}{1+\beta}MP_{Lj}\cdot h_j+\Delta$.

We now analyze the optimal progressivity. The first order condition with respect to τ is

$$\sum_j N_j\times\left[\begin{array}{l}\left(1+\xi_j\beta-\gamma\beta\frac{I}{N}\right)\times\frac{\partial V(w_j)}{\partial\tau}+\chi\cdot MP_{Lj}\frac{\partial h_j}{\partial\tau}\\ -\chi\cdot\lambda(1-\tau)w_j^{-\tau}h_j^{1-\tau}\left(\frac{\partial h_j}{\partial\tau}\cdot\frac{w_j}{h_j}+\frac{\lambda'(\tau)}{\lambda}\cdot\frac{w_j}{1-\tau}-\log\left(w_jh_j\right)\cdot\frac{w_j}{1-\tau}\right)\end{array}\right]=0. \quad (A.90)$$

The envelope condition of (26) yields $\frac{\partial V(w_j)}{\partial \tau} = V'(w_j) \times \left(\frac{\lambda'(\tau)}{\lambda} \cdot \frac{w_j}{1-\tau} - \log(w_j h_j) \cdot \frac{w_j}{1-\tau} \right)$, and combining this with the first order condition of the worker's problem $V'(w_j) \cdot \frac{w_j}{h_j} + u_h = 0$, we obtain $\frac{\partial V(w_j)}{\partial \tau} = V'(w_j) \times \left(\frac{\partial h_j}{\partial \tau} \cdot \frac{w_j}{h_j} + \frac{\lambda'(\tau)}{\lambda} \cdot \frac{w_j}{1-\tau} - \log(w_j h_j) \cdot \frac{w_j}{1-\tau} \right) + u_h \cdot \frac{\partial h_j}{\partial \tau}$. Then we can rewrite (A.90) as follows:

$$\sum_j N_j \cdot \Gamma_j \cdot \left(\frac{\partial h_j}{\partial \tau} \cdot \frac{w_j}{h_j} + \frac{\lambda'(\tau)}{\lambda} \cdot \frac{w_j}{1-\tau} - \log(w_j h_j) \cdot \frac{w_j}{1-\tau} \right) + \sum_j N_j \left(1 + \xi_j \beta - \gamma \beta \frac{I}{N} \right) \cdot u_h \cdot \frac{\partial h_j}{\partial \tau}$$
$$= - \sum_j N_j \cdot \chi \cdot MP_{Lj} \cdot \frac{\partial h_j}{\partial \tau}, \tag{A.91}$$

where $\Gamma_j = \left(1 + \xi_j \beta - \gamma \beta \frac{I}{N} \right) \cdot V'(w_j) - \chi \cdot \lambda (1-\tau) w_j^{-\tau} h_j^{1-\tau}$. The two terms of the left-hand side capture marginal benefit of increasing progressivity and the right-hand side reflects the marginal cost of increasing progressivity. By comparing this optimality condition (A.91) to (A.90) with $\phi_j = 0$ (since we analyze the single-type case), we can see that there is additional channel of increase in progressivity under the general preferences. The first term in the marginal benefit captures the net benefit associated with the extensive labor supply elasticity, and this term was zero under the special preferences in the main text (as the first order condition with respect to w_j implies $\Gamma_j = 0$).

C.2 Detailed Analysis of the Special Preferences Example in Section 6.2.2

In Section 6.2.2, we have characterized the optimal markdown under the special preference—Assumption 14—in the presence of intensive labor supply and progressive income taxation. Here we provide detailed derivation.

Extensive Elasticities and Expected Welfare

To derive the extensive elasticity of labor supply, we solve the worker's problem of choosing employers. Given the closed form solution of working hours $h(\tau) = (1-\tau)^{\frac{\epsilon_h}{1+\epsilon_h}}$, the value function from the worker's problem (26) also has the following closed-form:

$$V_s(w_{sj}) = (1-\tau) \log w_{sj} + \frac{\epsilon_h (1-\tau)}{1+\epsilon_h} (1 + \log(1-\tau)) + \log \lambda,$$

with which the optimal worker's choice over the employers is

$$V_s = E \left[\max_j (1-\tau) \log w_{sj} + \epsilon_{ij} \right] + \frac{\epsilon (1-\tau)}{1+\epsilon} (1 + \log(1-\tau)) + \log \lambda.$$

Then the probability of s-skilled worker's choosing firm j is

$$P_{sj} = \frac{w_{sj}^{\beta(1-\tau)}}{\sum_k w_{sk}^{\beta(1-\tau)}} = \frac{c_{sj}^{\beta}}{\sum_k c_{sk}^{\beta}}.$$

Then it is straightforward to show that the extensive elasticity of labor supply is simply

$$\epsilon_s^P(w_{sj}) = \beta(1-\tau), \quad \text{for all } s, j,$$

which implies that the extensive elasticities of labor supply do not depend on w_{sj}. Then the NDC-τ constraint boils down to $\frac{w_{hj}}{MP_{hj}} = \frac{w_{\ell j}}{MP_{\ell j}}$, and in terms of consumption, it is written as $\frac{c_{hj}}{MP_{hj}^{1-\tau}} = \frac{c_{\ell j}}{MP_{\ell j}^{1-\tau}}$. From the Logit problem, we obtain the expected welfare of the s-skilled worker:

$$V_s = \frac{1}{\beta}\log\left(\sum_j w_{sj}^{\beta(1-\tau)}\right) + constant = \frac{1}{\beta}\log\left(\sum_j c_{sj}^{\beta}\right) - \frac{\epsilon(1-\tau)}{1+\epsilon}.$$

Optimal Consumption and Pseudo Markdown $\tilde{\mu}_j$

Using the NDC and the worker's expected welfare derived above, it is straightforward to write down the constrained planner's problem:

$$\begin{aligned}
\max \quad & N_h\left[\frac{1}{\beta}\log\left(\sum_j c_{hj}^{\beta}\right) - \frac{\epsilon(1-\tau)}{1+\epsilon}\right] + N_\ell\left[\frac{1}{\beta}\log\left(\sum_j c_{\ell j}^{\beta}\right) - \frac{\epsilon(1-\tau)}{1+\epsilon}\right] + \psi_f\sum_j u(c_j^e) \qquad \text{(A.92)}\\
(\chi) \quad s.t. \quad & \sum_j c_{hj}N_{hj} + \sum_j c_{\ell j}N_{\ell j} + \sum_j c_j^e = \sum_j z_j\left(\alpha_j\cdot(N_{hj}\cdot h(\tau))^{\frac{\sigma-1}{\sigma}} + (1-\alpha_j)\cdot(N_{\ell j}\cdot h(\tau))^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}\\
(\xi_{sj}) \quad & N_{sj} = \frac{c_{sj}^{\beta}}{I_s}\times N_s, \quad s = h,\ell\\
(\gamma_s) \quad & \sum_j c_{sj}^{\beta} = I_s, \quad s = h,\ell \quad \forall j\\
(\phi_j) \quad & \frac{c_{hj}}{MP_{hj}^{1-\tau}} = \frac{c_{\ell j}}{MP_{\ell j}^{1-\tau}} \quad \forall j,
\end{aligned}$$

where the planner chooses allocation $\{c_{hj}, c_{\ell j}, N_{hj}, N_{\ell j}, c_j^e\}$ and progressivity τ together.

From the first order conditions of the planner's problem (A.92) with respect to allocation, the optimal consumption formulas yield:

$$\begin{aligned}
c_{sj} \quad &= \quad \frac{1}{\Omega_{sj}}\times\left[\frac{\beta}{1+\beta}MP_{sj}\cdot h(\tau) + \Delta_h\right], \quad s = h,\ell \qquad \text{(A.93)}\\
where \quad & \Delta_s = \frac{1-\gamma_s\beta\frac{I_s}{N_s}}{\chi(1+\beta)},\\
& \Omega_{hj} = 1 - \frac{\phi_j}{\chi(1+\beta)}\times\frac{1}{MP_{hj}^{1-\tau}\cdot N_{hj}}\times\left(1-\frac{\beta(1-\tau)}{\sigma}\right), \quad \text{and}\\
& \Omega_{\ell j} = 1 + \frac{\phi_j}{\chi(1+\beta)}\times\frac{1}{MP_{\ell j}^{1-\tau}\cdot N_{\ell j}}\times\left(1-\frac{\beta(1-\tau)}{\sigma}\right).
\end{aligned}$$

Note that the marginal product of labor $MP_{sj} = \frac{\partial y_j}{\partial L_{sj}}$ is the derivative of output with respect to effective labor L_{sj} (not N_{sj}), and thus the derivative of output respect to number of workers is $\frac{\partial y_j}{\partial N_{sj}} = MP_{sj}\times h(\tau)$.

From the optimal consumption formulas (A.93), by multiplying Ω_{sj} to both sides of the formula and using the definition of pseudo markdown $\tilde{\mu}(z_k) = \frac{c_{sj}}{MP_{sj}^{1-\tau}}$, we obtain

$$\Omega_{hj} \cdot \tilde{\mu}_j = \frac{\beta}{1+\beta} \cdot (1-\tau)^{\frac{\epsilon_h}{1+\epsilon_h}} \cdot MP_{hj}^{\tau} + \frac{\Delta_h}{MP_{hj}^{1-\tau}}, \tag{A.94}$$

$$\Omega_{\ell j} \cdot \tilde{\mu}_j = \frac{\beta}{1+\beta} \cdot (1-\tau)^{\frac{\epsilon_h}{1+\epsilon_h}} \cdot MP_{\ell j}^{\tau} + \frac{\Delta_\ell}{MP_{\ell j}^{1-\tau}}. \tag{A.95}$$

Summing up $MP_{hj}^{1-\tau} \cdot L_{hj} \times$(A.94) and $MP_{\ell j}^{1-\tau} \cdot L_{\ell j} \times$(A.95) yields

$$\left(MP_{hj}^{1-\tau} \cdot L_{hj} + MP_{\ell j}^{1-\tau} \cdot L_{\ell j}\right) \times \tilde{\mu}_j = \frac{\beta}{1+\beta} \cdot (1-\tau)^{\frac{\epsilon_h}{1+\epsilon_h}} \cdot \times (MP_{hj} L_{hj} + MP_{\ell j} L_{\ell j}) + L_{hj} \cdot \Delta_h + L_{\ell j} \cdot \Delta_\ell.$$

Dividing both sides by $\left(MP_{hj}^{1-\tau} \cdot L_{hj} + MP_{\ell j}^{1-\tau} \cdot L_{\ell j}\right)$ yields the pseudo optimal markdown equation in the main text:

$$\tilde{\mu}_j = A_j \cdot h(\tau) \cdot \frac{\beta}{1+\beta} + s(\alpha_j) \cdot \frac{\Delta_h}{MP_{hj}^{1-\tau}} + (1 - s(\alpha_j)) \cdot \frac{\Delta_\ell}{MP_{\ell j}^{1-\tau}},$$

where $A = \frac{MP_{hj} L_{hj} + MP_{\ell j} L_{\ell j}}{MP_{hj}^{1-\tau} \cdot L_{hj} + MP_{\ell j}^{1-\tau} \cdot L_{\ell j}} = \frac{y_j}{MP_{hj}^{1-\tau} \cdot L_{hj} + MP_{\ell j}^{1-\tau} \cdot L_{\ell j}}$ and $s(\alpha_j) = \frac{MP_{hj}^{1-\tau} \cdot L_{hj}}{MP_{hj}^{1-\tau} \cdot L_{hj} + MP_{\ell j}^{1-\tau} \cdot L_{\ell j}}$.

Decomposition of $\tilde{\mu}_j$

In the main text, we have decomposed the slope of the pseudo optimal markdown $\tilde{\mu}_j = \tilde{\mu}(z_j)$ by

$$\frac{d\mu_j}{dz_j} = F_\tau + \underbrace{F_{1z} + F_{1_\alpha}}_{=F1} + F2, \quad \text{where}$$

(i) $F_\tau = \frac{dA_j}{dz_j} \times h(\tau) \cdot \frac{\beta}{1+\beta}$, (ii) $F_{1z} = \left.\frac{\partial \tilde{\mu}_j^{net}}{\partial z_j}\right|_{\frac{L_{hj}}{L_{\ell j}}, \alpha_j \text{ given}}$, (iii) $F_{1_\alpha} = \left.\frac{\partial \tilde{\mu}_j^{net}}{\partial \alpha_j}\right|_{\frac{L_{hj}}{L_{\ell j}} \text{ given}} \times \alpha'(z_j)$, and (iv) $F2 = \frac{\partial \tilde{\mu}_j^{net}}{\partial\left(\frac{L_{hj}}{L_{\ell j}}\right)} \times \frac{d\left(\frac{L_{hj}}{L_{\ell j}}\right)}{dz_j}$. We now derive the expression for each component.

The following proposition characterizes each component.

Proposition AP.2. *The decomposition of the* $\frac{d\tilde{\mu}_j}{dz_j}$ *are characterized as follows.*

1. ***Substitution effect:*** *The* F_τ*-term is given by*

$$F_\tau = \left(\frac{\partial A_j}{\partial z_j} + \frac{\partial A_j}{\partial \alpha_j} \cdot \alpha'(z_j) + \frac{\partial A_j}{\partial\left(\frac{L_{hj}}{L_{\ell j}}\right)} \cdot \frac{d\left(\frac{L_{hj}}{L_{\ell j}}\right)}{dz_j}\right),$$

where $\frac{\partial A_j}{\partial z_j} > 0$ *and* $\frac{\partial A_j}{\partial\left(\frac{L_{hj}}{L_{\ell j}}\right)} > 0$*. If* $L_{hj} > L_{\ell j}$*, then* $\frac{\partial A_j}{\partial \alpha_j} > 0$ *and thus* $F_\tau > 0$*.*

2. ***Redistribution across firms:***
The F_{1z}-term is given by $F_{1z} = -\left(s_j \cdot \frac{(1-\tau)\Delta_h}{MP_{hj}^{2-\tau}} \cdot \frac{\partial MP_{hj}}{\partial z_j} + (1-s_j) \cdot \frac{(1-\tau)\Delta_\ell}{MP_{\ell j}^{2-\tau}}\right) < 0$. *And the* $F_{1\alpha}$*-term is given by* $F_{1\alpha} = \frac{(1-\tau)\cdot\tilde{\mu}_j^{net}}{(y_j/z_j)^{\frac{\sigma-1}{\sigma}}} \times \left(\frac{\sigma}{\sigma-1} \cdot \left(L_{hj}^{\frac{\sigma-1}{\sigma}} - L_{\ell j}^{\frac{\sigma-1}{\sigma}} \right) + \frac{L_{hj}^{\frac{\sigma-1}{\sigma}} L_{\ell j}^{\frac{\sigma-1}{\sigma}} \cdot \left(\alpha_j^{-\tau} L_{hj}^{\frac{\tau}{\sigma}} - (1-\alpha_j)^{-\tau} L_{\ell j}^{\frac{\tau}{\sigma}} \right)}{\alpha_j^{1-\tau} L_{hj}^{\frac{\sigma-1+\tau}{\sigma}} + (1-\alpha_j)^{1-\tau} L_{\ell j}^{\frac{\sigma-1+\tau}{\sigma}}} \right)$.
Assume $\sigma > 1$. *Then,* $F_{1\alpha} > 0$ *if* $L_{\ell j} > L_{hj}$.

3. ***Redistribution between skills:***
$F2 = \frac{\partial \tilde{\mu}_j^{net}}{\partial\left(\frac{L_{hj}}{L_{\ell j}}\right)} \times \frac{d\left(\frac{L_{hj}}{L_{\ell j}}\right)}{dz_j}$ *and* $\frac{\partial \tilde{\mu}_j^{net}}{\partial\left(\frac{L_{hj}}{L_{\ell j}}\right)} < 0$. *Thus,* $sign(F2) = -sign\left(\frac{d\left(\frac{L_{hj}}{L_{\ell j}}\right)}{dz_j}\right)$.

We prove Proposition AP.2. First, we start with the substitution effect, F_τ. Note that in the denominator of A_j,

$$MP_{sj}^{1-\tau} \cdot L_{sj} = \left(z_j \cdot \tilde{y}_j^{\frac{1}{\sigma}}\right)^{1-\tau} \cdot \alpha_j^{1-\tau} \cdot L_{sj}^{1-\frac{1-\tau}{\sigma}},$$

where $\tilde{y_j} = \left(\alpha_j L_{hj}^{\frac{\sigma-1}{\sigma}} + (1-\alpha_j) L_{\ell j}^{\frac{\sigma-1}{\sigma}}\right)$. We then rewrite the factor A_j in the following way.

$$A_j = \frac{z_j^\tau \cdot \tilde{y}_j^{\frac{\sigma-1+\tau}{\sigma}}}{\alpha_j^{1-\tau} \cdot L_{hj}^{\frac{\sigma-1+\tau}{\sigma}} + (1-\alpha_j)^{1-\tau} \cdot L_{\ell j}^{\frac{\sigma-1+\tau}{\sigma}}} = \frac{z_j^\tau \cdot \left(\alpha_j \left(\frac{L_{hj}}{L_{\ell j}}\right)^{\frac{\sigma-1}{\sigma}} + 1 - \alpha_j\right)^{\frac{\sigma-1+\tau}{\sigma-1}}}{\alpha_j^{1-\tau} \cdot \left(\frac{L_{hj}}{L_{\ell j}}\right)^{\frac{\sigma-1+\tau}{\sigma}} + (1-\alpha_j)^{1-\tau}}.$$

The F_τ-term is given by

$$F_\tau = \left(\frac{\partial A_j}{\partial z_j} + \frac{\partial A_j}{\partial \alpha_j} \cdot \alpha'(z_j) + \frac{\partial A_j}{\partial\left(\frac{L_{hj}}{L_{\ell j}}\right)} \cdot \frac{d\left(\frac{L_{hj}}{L_{\ell j}}\right)}{dz_j}\right) \times h(\tau) \cdot \frac{\beta}{1+\beta}. \tag{A.96}$$

The first term in the bracket is positive because $\frac{\partial A_j}{\partial z_j} = \tau \cdot z_j^{\tau-1} \cdot \frac{\tilde{y_j}^{\sigma-1+\tau}}{\alpha_j^{1-\tau} \cdot L_{hj}^{\frac{\sigma-1+\tau}{\sigma}} + (1-\alpha_j)^{1-\tau} \cdot L_{\ell j}^{\frac{\sigma-1+\tau}{\sigma}}} > 0$.
The factor of the second term in the bracket of (A.96) is

$$\frac{\partial A_j}{\partial \alpha_j} = \frac{z_j^\tau \cdot \tilde{y}_j^{\frac{\tau}{\sigma}} \times \left[\frac{\sigma-1+\tau}{\sigma-1}\left(L_{hj}^{\frac{\sigma-1}{\sigma}} - L_{\ell j}^{\frac{\sigma-1}{\sigma}}\right)\left(\alpha_j^{1-\tau} L_{hj}^{\frac{\sigma-1+\tau}{\sigma}} + (1-\alpha_j)^{1-\tau} L_{\ell j}^{\frac{\sigma-1+\tau}{\sigma}}\right) - (1-\tau)\tilde{y}_j^{\frac{\sigma-1}{\sigma}}\left(\alpha_j^{-\tau} L_{hj}^{\frac{\sigma-1+\tau}{\sigma}} - (1-\alpha_j)^{-\tau} L_{\ell j}^{\frac{\sigma-1+\tau}{\sigma}}\right)\right]}{\left(\alpha_j^{1-\tau} L_{hj}^{\frac{\sigma-1+\tau}{\sigma}} + (1-\alpha_j)^{1-\tau} L_{\ell j}^{\frac{\sigma-1+\tau}{\sigma}}\right)^2},$$

and to determine its sign we rewrite the numerator (after some algebra):

$$z^\tau \tilde{y}_j^{\frac{\tau}{\sigma}} \times \left[\begin{array}{c} (1-\tau) \cdot L_{hj}^{\frac{\sigma-1}{\sigma}} \cdot L_{\ell j}^{\frac{\sigma-1}{\sigma}} \cdot \left((1-\alpha_j)^{-\tau} L_{\ell j}^{\frac{\tau}{\sigma}} - \alpha_j^{-\tau} L_{hj}^{\frac{\tau}{\sigma}}\right) \\ + \left(L_{hj}^{\frac{\sigma-1}{\sigma}} - L_{\ell j}^{\frac{\sigma-1}{\sigma}}\right) \cdot \left(\alpha_j^{1-\tau} L_{hj}^{\frac{\sigma-1+\tau}{\sigma}} + (1-\alpha_j)^{1-\tau} L_{\ell j}^{\frac{\sigma-1+\tau}{\sigma}}\right) \end{array}\right],$$

which shows that $\frac{\partial A_j}{\partial \alpha_j}$ can be either positive or negative. If $L_{hj} > L_{\ell j}$, then $\frac{\partial A_j}{\partial \alpha_j} > 0$. The factor of the third term in the bracket of (A.96) can be rewritten as follows (after some algebra):

$$\frac{\partial A_j}{\partial\left(\frac{L_{hj}}{L_{\ell j}}\right)} = \frac{\alpha_j \cdot (1-\alpha_j) \cdot z_j^\tau \cdot \frac{\sigma-1+\tau}{\sigma} \cdot \left(\frac{L_{hj}}{L_{\ell j}}\right)^{-\frac{1}{\sigma}} \cdot \left(\alpha_j \left(\frac{L_{hj}}{L_{\ell j}}\right)^{\frac{\sigma-1}{\sigma}} + 1 - \alpha_j\right)^{\frac{\tau}{\sigma-1}} \cdot \left((1-\alpha_j)^{-\tau} - \alpha^{-\tau}\left(\frac{L_{hj}}{L_{\ell j}}\right)^{\frac{\tau}{\sigma}}\right)}{\left(\alpha_j^{1-\tau}\left(\frac{L_{hj}}{L_{\ell j}}\right)^{\frac{\sigma-1+\tau}{\sigma}} + (1-\alpha_j)^{1-\tau}\right)^2} > 0.$$ Note

that $\frac{MP_{hj}}{MP_{\ell j}} = \frac{L_{\ell j}^{\frac{1}{\sigma}}}{L_{hj}^{\frac{1}{\sigma}}} \cdot \frac{\alpha_j}{1-\alpha_j}$, and thus $MP_{hj} > MP_{\ell j}$ implies $\frac{L_{\ell j}^{\frac{1}{\sigma}}}{1-\alpha_j} > \frac{L_{hj}^{\frac{1}{\sigma}}}{\alpha)j}$, which implies $\frac{\partial A_j}{\partial\left(\frac{L_{hj}}{L_{\ell j}}\right)} > 0$.

Next, we derive the expression for $F1 = F_{1z} + F_{1\alpha}$. The F_{1z}-term is obtained by simply taking a partial derivative of $\tilde{\mu}_j^{net}$ with respect to z_j:

$$F_{1z} = -\left(s_j \cdot \frac{(1-\tau)\Delta_h}{MP_{hj}^{2-\tau}} \cdot \frac{\partial MP_{hj}}{\partial z_j} + (1-s_j) \cdot \frac{(1-\tau)\Delta_\ell}{MP_{\ell j}^{2-\tau}}\right) < 0.$$

We then derive the $F_{1\alpha}$ term. Note that

$$F_{1\alpha} = \left[\begin{array}{c} \frac{\partial s_j}{\partial \alpha_j}\Big|_{\frac{L_{hj}}{L_{\ell j}}} \times \left(\frac{\Delta_h}{MP_{hj}^{1-\tau}} - \frac{\Delta_\ell}{MP_{\ell j}^{1-\tau}}\right) \\ -\left(s_j \cdot \frac{(1-\tau)\Delta_h}{MP_{hj}^{2-\tau}} \cdot \frac{\partial MP_{hj}}{\partial \alpha_j}\Big|_{\frac{L_{hj}}{L_{\ell j}}} + (1-s_j) \cdot \frac{(1-\tau)\Delta_\ell}{MP_{\ell j}^{2-\tau}} \cdot \frac{\partial MP_{\ell j}}{\partial \alpha_j}\Big|_{\frac{L_{hj}}{L_{\ell j}}}\right) \end{array}\right] \times \alpha'(z_j) \qquad \text{(A.97)}$$

Note that $s(\alpha_j) = \frac{MP_{hj}^{1-\tau} \cdot L_{hj}}{MP_{hj}^{1-\tau} \cdot L_{hj} + MP_{\ell j}^{1-\tau} \cdot L_{\ell j}} = \frac{\alpha_j^{1-\tau} L_{hj}^{\frac{\sigma-1+\tau}{\sigma}}}{\alpha_j^{1-\tau} L_{hj}^{\frac{\sigma-1+\tau}{\sigma}} + (1-\alpha_j)^{1-\tau} L_{\ell j}^{\frac{\sigma-1+\tau}{\sigma}}}$. Thus

$$\frac{\partial s_j}{\partial \alpha_j}\Big|_{\frac{L_{hj}}{L_{\ell j}}} = \frac{(1-\tau) \cdot \alpha_j^{-\tau} \cdot L_{hj}^{\frac{\sigma-1+\tau}{\sigma}} - s_j \cdot (1-\tau) \cdot \left(\alpha_j^{-\tau} L_{hj}^{\frac{\sigma-1+\tau}{\sigma}} - (1-\alpha_j)^{-\tau} L_{\ell j}^{\frac{\sigma-1+\tau}{\sigma}}\right)}{\alpha_j^{1-\tau} L_{hj}^{\frac{\sigma-1+\tau}{\sigma}} + (1-\alpha_j)^{1-\tau} L_{\ell j}^{\frac{\sigma-1+\tau}{\sigma}}} > 0.$$

Using this expression, the term in (A.97) associated with derivative of MP_{hj} yields

$$\begin{aligned} & s_j \cdot \frac{(1-\tau)\Delta_h}{MP_{hj}^{2-\tau}} \cdot \frac{\partial MP_{hj}}{\partial \alpha_j}\Big|_{\frac{L_{hj}}{L_{\ell j}}} \\ = \; & \frac{\Delta_h}{MP_{hj}^{1-\tau}} \times \left(\frac{(1-\tau) \cdot \alpha_j^{-\tau} \cdot L_{hj}^{\frac{\sigma-1+\tau}{\sigma}}}{\alpha_j^{1-\tau} L_{hj}^{\frac{\sigma-1+\tau}{\sigma}} + (1-\alpha_j)^{1-\tau} L_{\ell j}^{\frac{\sigma-1+\tau}{\sigma}}} + \frac{1}{\sigma-1} \cdot s_j \cdot (1-\tau) \cdot \frac{L_{hj}^{\frac{\sigma-1}{\sigma}} - L_{\ell j}^{\frac{\sigma-1}{\sigma}}}{\alpha_j L_{hj}^{\frac{\sigma-1}{\sigma}} + (1-\alpha_j) L_{\ell j}^{\frac{\sigma-1}{\sigma}}}\right) \\ = \; & \frac{\Delta_h}{MP_{hj}^{1-\tau}} \times \left(\begin{array}{c} \frac{\partial s_j}{\partial \alpha_j}\Big|_{\frac{L_{hj}}{L_{\ell j}}} + \frac{s_j \cdot (1-\tau) \cdot \left(\alpha_j^{-\tau} L_{hj}^{\frac{\sigma-1+\tau}{\sigma}} - (1-\alpha_j)^{-\tau} L_{\ell j}^{\frac{\sigma-1+\tau}{\sigma}}\right)}{\alpha_j^{1-\tau} L_{hj}^{\frac{\sigma-1+\tau}{\sigma}} + (1-\alpha_j)^{1-\tau} L_{\ell j}^{\frac{\sigma-1+\tau}{\sigma}}} \\ + \frac{1}{\sigma-1} \cdot s_j \cdot (1-\tau) \cdot \frac{L_{hj}^{\frac{\sigma-1}{\sigma}} - L_{\ell j}^{\frac{\sigma-1}{\sigma}}}{\alpha_j L_{hj}^{\frac{\sigma-1}{\sigma}} + (1-\alpha_j) L_{\ell j}^{\frac{\sigma-1}{\sigma}}} \end{array}\right) \end{aligned}$$

Similarly, we derive $(1-s_j)\cdot\frac{(1-\tau)\Delta_\ell}{MP_{\ell j}^{2-\tau}}\cdot\left.\frac{\partial MP_{\ell j}}{\partial\alpha_j}\right|_{\frac{L_{hj}}{L_{\ell j}}}$, which yields

$$\begin{aligned}
&-\left[s_j\cdot\frac{(1-\tau)\Delta_h}{MP_{hj}^{2-\tau}}\cdot\left.\frac{\partial MP_{hj}}{\partial\alpha_j}\right|_{\frac{L_{hj}}{L_{\ell j}}}+(1-s_j)\cdot\frac{(1-\tau)\Delta_\ell}{MP_{\ell j}^{2-\tau}}\cdot\left.\frac{\partial MP_{\ell j}}{\partial\alpha_j}\right|_{\frac{L_{hj}}{L_{\ell j}}}\right]\\
=\;&-\left.\frac{\partial s_j}{\partial\alpha_j}\right|_{\frac{L_{hj}}{L_{\ell j}}}\times\left(\frac{\Delta_h}{MP_{hj}^{1-\tau}}-\frac{\Delta_\ell}{MP_{\ell j}^{1-\tau}}\right)\\
&-(1-\tau)\cdot\left(s_j\frac{\Delta_h}{MP_{hj}^{1-\tau}}+(1-s_j)\frac{\Delta_\ell}{MP_{\ell j}^{1-\tau}}\right)\cdot\left(\begin{array}{c}\frac{\alpha_j^{-\tau}L_{hj}^{\frac{\sigma-1+\tau}{\sigma}}-(1-\alpha_j)^{-\tau}L_{\ell j}^{\frac{\sigma-1+\tau}{\sigma}}}{\alpha_j^{1-\tau}L_{hj}^{\frac{\sigma-1+\tau}{\sigma}}+(1-\alpha_j)^{1-\tau}L_{\ell j}^{\frac{\sigma-1+\tau}{\sigma}}}\\+\frac{1}{\sigma-1}\cdot\frac{L_{hj}^{\frac{\sigma-1}{\sigma}}-L_{\ell j}^{\frac{\sigma-1}{\sigma}}}{\alpha_jL_{hj}^{\frac{\sigma-1}{\sigma}}+(1-\alpha_j)L_{\ell j}^{\frac{\sigma-1}{\sigma}}}\end{array}\right).
\end{aligned}$$

By substituting these into (A.97) yields

$$F_{1\alpha}=\frac{(1-\tau)\cdot\tilde{\mu}_j^{net}}{(y_j/z_j)^{\frac{\sigma-1}{\sigma}}}\times\left(\frac{\sigma}{\sigma-1}\cdot\left(L_{hj}^{\frac{\sigma-1}{\sigma}}-L_{\ell j}^{\frac{\sigma-1}{\sigma}}\right)+\frac{L_{hj}^{\frac{\sigma-1}{\sigma}}L_{\ell j}^{\frac{\sigma-1}{\sigma}}\cdot\left(\alpha_j^{-\tau}L_{hj}^{\frac{\tau}{\sigma}}-(1-\alpha_j)^{-\tau}L_{\ell j}^{\frac{\tau}{\sigma}}\right)}{\alpha_j^{1-\tau}L_{hj}^{\frac{\sigma-1+\tau}{\sigma}}+(1-\alpha_j)^{1-\tau}L_{\ell j}^{\frac{\sigma-1+\tau}{\sigma}}}\right).$$

Assume $\sigma>1$. Then, $F_{1\alpha}>0$ if $L_{\ell j}>L_{hj}$.

Lastly, we derive the expression for $F2=\frac{\partial\tilde{\mu}_j^{net}}{\partial\left(\frac{L_{hj}}{L_{\ell j}}\right)}\times\frac{d\left(\frac{L_{hj}}{L_{\ell j}}\right)}{dz_j}$ and $\frac{\partial\tilde{\mu}_j^{net}}{\partial\left(\frac{L_{hj}}{L_{\ell j}}\right)}$. Note that

$$F2=\left[\begin{array}{c}\frac{\partial s_j}{\partial\left(\frac{L_{hj}}{L_{\ell j}}\right)}\cdot\left(\frac{\Delta_h}{MP_{hj}^{1-\tau}}-\frac{\Delta_\ell}{MP_{\ell j}^{1-\tau}}\right)\\-\left(s_j\frac{(1-\tau)\Delta_h}{MP_{hj}^{2-\tau}}\cdot\frac{\partial MP_{hj}}{\partial\left(\frac{L_{hj}}{L_{\ell j}}\right)}+(1-s_j)\frac{(1-\tau)\Delta_h}{MP_{\ell j}^{2-\tau}}\cdot\frac{\partial MP_{\ell j}}{\partial\left(\frac{L_{hj}}{L_{\ell j}}\right)}\right)\end{array}\right]\times\frac{d\left(\frac{L_{hj}}{L_{\ell j}}\right)}{dz_j}\tag{A.98}$$

Note that $s(\alpha_j)=\frac{\alpha_j^{1-\tau}L_{hj}^{\frac{\sigma-1+\tau}{\sigma}}}{\alpha_j^{1-\tau}L_{hj}^{\frac{\sigma-1+\tau}{\sigma}}+(1-\alpha_j)^{1-\tau}L_{\ell j}^{\frac{\sigma-1+\tau}{\sigma}}}=\frac{\alpha_j^{1-\tau}}{\alpha_j^{1-\tau}+(1-\alpha_j)^{1-\tau}\cdot\left(\frac{L_{hj}}{L_{\ell j}}\right)^{-\frac{\sigma-1+\tau}{\sigma}}}$, and thus

$$\frac{\partial s_j}{\partial\left(\frac{L_{hj}}{L_{\ell j}}\right)}=\frac{\sigma-1+\tau}{\sigma}\cdot\alpha_j^{1-\tau}\cdot(1-\alpha_j)^{1-\tau}\cdot\frac{L_{hj}^{\frac{\sigma-1+\tau}{\sigma}}\cdot L_{\ell j}^{\frac{\sigma-1+\tau}{\sigma}}}{\left(\alpha_j^{1-\tau}L_{hj}^{\frac{\sigma-1+\tau}{\sigma}}+(1-\alpha_j)^{1-\tau}L_{\ell j}^{\frac{\sigma-1+\tau}{\sigma}}\right)^2}\cdot\left(\frac{L_{hj}}{L_{\ell j}}\right)^{-1}.$$

We can also derive the following after some algebra:

$$\begin{aligned}
&-\left(s_j\cdot\Delta_h\cdot(1-\tau)\cdot\frac{1}{MP_{hj}^{2-\tau}}\cdot\frac{\partial MP_{hj}}{\partial\left(\frac{L_{hj}}{L_{\ell j}}\right)}+(1-s_j)\cdot\Delta_\ell\cdot(1-\tau)\cdot\frac{1}{MP_\ell^{2-\tau}}\cdot\frac{\partial MP_{\ell j}}{\partial\left(\frac{L_{hj}}{L_{\ell j}}\right)}\right)\\
=\;&\frac{(1-\tau)\cdot\frac{\alpha_j(1-\alpha_j)}{\sigma}\cdot z_j^\tau\cdot L_{hj}^{\frac{\sigma-1}{\sigma}}\cdot L_{\ell j}^{\frac{\sigma-1}{\sigma}}}{\left(\alpha_j^{1-\tau}L_{hj}^{\frac{\sigma-1+\tau}{\sigma}}+(1-\alpha_j)^{1-\tau}L_{\ell j}^{\frac{\sigma-1+\tau}{\sigma}}\right)\cdot\left(\alpha_jL_{hj}^{\frac{\sigma-1}{\sigma}}+(1-\alpha_j)L_{\ell j}^{\frac{\sigma-1}{\sigma}}\right)^{1-\tau}}\cdot\left(\frac{L_{hj}}{L_{\ell j}}\right)^{-1}\cdot\left(\frac{\Delta_h}{MP_{hj}}-\frac{\Delta_\ell}{MP_{\ell j}}\right).
\end{aligned}$$

Thus, by substituting these terms into (A.98), we obtain

$$F2 = \left[\begin{array}{c} \frac{\frac{\sigma-1+\tau}{\sigma}\cdot\alpha_j^{1-\tau}\cdot(1-\alpha_j)^{1-\tau}\cdot L_{hj}^{\frac{\sigma-1+\tau}{\sigma}}\cdot L_{\ell j}^{\frac{\sigma-1+\tau}{\sigma}}}{\left(\alpha_j^{1-\tau}L_{hj}^{\frac{\sigma-1+\tau}{\sigma}}+(1-\alpha_j)^{1-\tau}L_{\ell j}^{\frac{\sigma-1+\tau}{\sigma}}\right)^2}\cdot\left(\frac{L_{hj}}{L_{\ell j}}\right)^{-1}\cdot\left(\frac{\Delta_h}{MP_{hj}^{1-\tau}}-\frac{\Delta_\ell}{MP_{\ell j}^{1-\tau}}\right) \\ +\frac{(1-\tau)\cdot\frac{\alpha_j(1-\alpha_j)}{\sigma}\cdot z_j^{\tau}\cdot L_{hj}^{\frac{\sigma-1}{\sigma}}\cdot L_{\ell j}^{\frac{\sigma-1}{\sigma}}}{\left(\alpha_j^{1-\tau}L_{hj}^{\frac{\sigma-1+\tau}{\sigma}}+(1-\alpha_j)^{1-\tau}L_{\ell j}^{\frac{\sigma-1+\tau}{\sigma}}\right)\cdot\left(\alpha_j L_{hj}^{\frac{\sigma-1}{\sigma}}+(1-\alpha_j)L_{\ell j}^{\frac{\sigma-1}{\sigma}}\right)^{1-\tau}}\cdot\left(\frac{L_{hj}}{L_{\ell j}}\right)^{-1}\cdot\left(\frac{\Delta_h}{MP_{hj}}-\frac{\Delta_\ell}{MP_{\ell j}}\right) \end{array} \right] \times \frac{d\left(\frac{L_{hj}}{L_{\ell j}}\right)}{dz_j}.$$

To determine the sign of the $F2$, we show that

$$\begin{aligned} \frac{\Delta_h}{MP_{hj}^{1-\tau}}-\frac{\Delta_\ell}{MP_{\ell j}^{1-\tau}} &= (\Omega_{hj}-\Omega_{\ell j})\cdot\mu_j-\frac{\beta}{1+\beta}(MP_{hj}-MP_{\ell j}) \\ &= -\frac{\phi_j}{\chi(1+\beta)}\left(1+\frac{\beta(1-\tau)}{\sigma}\right)\cdot\left(\frac{1}{MP_{hj}^{1-\tau}L_{hj}}+\frac{1}{MP_{\ell j}^{1-\tau}L_{\ell j}}\right)-\frac{\beta}{1+\beta}(MP_{hj}-MP_{\ell j})<0 \\ \frac{\Delta_h}{MP_{hj}}-\frac{\Delta_\ell}{MP_{\ell j}} &= \left(MP_{hj}^{-\tau}\Omega_{hj}-MP_{\ell j}^{-\tau}\Omega_{\ell j}\right)\times\mu_j \\ &= \left[\left(\frac{1}{MP_{hj}^{\tau}}-\frac{1}{MP_{\ell j}^{\tau}}\right)-\frac{\phi_j}{\chi(1+\beta)}\left(1+\frac{\beta(1-\tau)}{\sigma}\right)\cdot\left(\frac{1}{MP_{hj}L_{hj}}+\frac{1}{MP_{\ell j}L_{\ell j}}\right)\right]\times\mu_j<0, \end{aligned}$$

where the inequality uses the fact that $MP_{hj}>MP_{\ell j}$. Thus,

$$sign(F2) = -sign\left(\frac{d\left(\frac{L_{hj}}{L_{\ell j}}\right)}{dz_j}\right).$$

C.3 Appendix for the Quantitative Analysis in Section 6

C.3.1 Formulas for the Quantitative Model

Formulas for the Joint Optimal Policy

We present and prove the formula for the optimal markdown and progressivity in the presence of the labor income tax.

The formula for the optimal markdown is given by

$$\begin{aligned} \tilde{\mu}_j &= \frac{\eta y_j}{\sum_m MR_{mj}^{1-\tau}L_{mj}}(1-\tau)^{\frac{\epsilon}{1+\epsilon}}\frac{\beta}{1+\beta}+\sum_m\left(\frac{MR_{mj}^{1-\tau}L_{mj}}{\sum_k MR_{kj}^{1-\tau}L_{kj}}\frac{\Delta_m}{MR_{mj}^{1-\tau}}\right), \\ \text{where}\quad \Delta_m &= \frac{1}{\lambda(1+\beta)}\left(1-\nu_m\frac{I_m}{N_m}\right) \end{aligned} \tag{A.99}$$

where $\tilde{\mu_j}=\phi\mu_j^{1-\tau}$, $MR_{mj}=\eta z_j L_j^{\eta-1}\alpha_{mj}$ is the marginal product of labor, and $I_m=\left(\sum_m c_{mj}^{\beta}g_{mj}^{\beta}\right)^{1/\beta}$ is the wage index for type m workers. Lagrange multipliers for the resource constraint λ and for the wage index equation ν_m satisfies

$$\frac{1}{\lambda}=\frac{\sum_j z_j L_j^{\eta}}{N+\psi_f J},\quad \frac{\nu_m}{\lambda}=\beta\frac{\sum_j(MR_{mj}L_{mj}-c_{mj}N_{mj})}{I_m} \tag{A.100}$$

The formula for the optimal progressivity is given by

$$1-\tau=\frac{\lambda\frac{\eta\sum_j y_j}{N}}{1+\frac{1+\epsilon}{\epsilon}\frac{\sum_j\sum_{m=1}^{M-1}\left(\zeta_{mj}\tilde{\mu}_j\log\frac{MR_{m+1,j}}{MR_{mj}}\right)}{N}} \tag{A.101}$$

where ζ_{mj} is the lagrangian multiplier for the no-discrimination constraints.

To prove the result, we start from stating the planner's problem

$$\max_{\{c_{mj},c_{fj}\}} \quad \sum_m \frac{N_m}{\beta}\log\left(\sum_m c_{mj}^\beta g_{mj}^\beta\right)-N(1-\tau)\frac{\epsilon}{1+\epsilon}+\psi_f\sum_j\log c_{fj} \tag{A.102}$$

$$s.t. \quad \sum_j\sum_m c_{mj}L_{mj}+\sum_j c_{fj}\leq\sum_j z_j\left(\sum_m\alpha_{mj}N_{mj}(1-\tau)^{\frac{\epsilon}{1+\epsilon}}\right)^\eta \tag{A.103}$$

$$I_m=\left(\sum_m c_{mj}^\beta g_{mj}^\beta\right)^{1/\beta}, \quad m=1,\dots,M \tag{A.104}$$

$$\frac{c_{m+1,j}}{(\eta z_j L_j^{\eta-1}\alpha_{m+1,j})^{1-\tau}}=\frac{c_{mj}}{(\eta z_j L_j^{\eta-1}\alpha_{mj})^{1-\tau}}, \quad j=1\dots J, m=1,\dots,M-1 \tag{A.105}$$

and $N_{mj}=N_m\frac{c_{mj}^\beta g_{mj}^\beta}{I_m}$.

Note that the no-discrimination constraint can be rewritten as

$$\frac{c_{m+1,j}}{(\eta z_j L_j^{\eta-1}\alpha_{m+1,j})^{1-\tau}}=\frac{c_{mj}}{(\eta z_j L_j^{\eta-1}\alpha_{mj})^{1-\tau}} \quad \Leftrightarrow \frac{c_{m+1,j}}{\alpha_{m+1,j}^{1-\tau}}=\frac{c_{mj}}{\alpha_{mj}^{1-\tau}} \tag{A.106}$$

From now, we will use the constraint $\frac{c_{m+1,j}}{\alpha_{m+1,j}}=\frac{c_{mj}}{\alpha_{mj}}$ instead and denote the associated lagrange multiplier as $\tilde{\zeta}_{mj}$. If we denote the lagrange multiplier for the no-discrimination constraint as ζ_{mj}, then $\zeta_{mj}=\tilde{\zeta}_{mj}\times(\eta z_j L_j^{\eta-1})^{1-\tau}$.

Lagrangian is given by

$$\begin{aligned}\mathscr{L}=&\sum_m\frac{N_m}{\beta}\log\left(\sum_m c_{mj}^\beta g_{mj}^\beta\right)-N(1-\tau)\frac{\epsilon}{1+\epsilon}+\psi_f\sum_j\log c_{fj}\\&+\lambda\left(\sum_j z_j\left(\sum_m\alpha_{mj}N_{mj}(1-\tau)^{\frac{\epsilon}{1+\epsilon}}\right)^\eta-\sum_j\sum_m c_{mj}N_{mj}+\sum_j c_{fj}\right)\\&+\sum_m\nu_m\left(I_m-\left(\sum_m c_{mj}^\beta g_{mj}^\beta\right)^{1/\beta}\right)+\sum_j\sum_{m=1}^{M-1}\tilde{\zeta}_{mj}\left(\frac{c_{m+1,j}}{\alpha_{m+1,j}^{1-\tau}}-\frac{c_{mj}}{\alpha_{mj}^{1-\tau}}\right)\end{aligned}$$

First order condition with respect to c_{mj} is

$$\frac{N_{mj}}{c_{mj}}+\lambda\left((\eta z_j L_j^{\eta-1}\alpha_{mj}(1-\tau)^{\frac{\epsilon}{1+\epsilon}}-c_{mj})\beta\frac{N_{mj}}{c_{mj}}-N_{mj}\right)-\nu_m\frac{I_m}{N_m}\frac{N_{mj}}{c_{mj}}+\frac{1}{\alpha_{mj}^{1-\tau}}(\tilde{\zeta}_{m-1,j}-\tilde{\zeta}_{m,j})=0 \tag{A.107}$$

Multiply by $\frac{N_{mj}}{c_{mj}}\frac{1}{\lambda(1+\beta)}$ and rearranging yields

$$c_{mj}\left(1+\frac{1}{1+\beta}\frac{1}{\alpha_{mj}^{1-\tau}N_{mj}}\left(\frac{\tilde{\zeta}_{m,j}}{\lambda}-\frac{\tilde{\zeta}_{m-1,j}}{\lambda}\right)\right)=\frac{\beta}{1+\beta}\eta z_j L_j^{\eta-1}\alpha_{mj}(1-\tau)^{\frac{\epsilon}{1+\epsilon}}+\Delta_m \tag{A.108}$$

where we define $\Delta_m=\frac{1}{\lambda(1+\beta)}\left(1-\nu_m\frac{I_m}{N_m}\right)$.

To obtain the expression for $\tilde{\mu}_j=\frac{c_{mj}}{(\eta z_j L_j^{\eta-1}\alpha_{mj})^{1-\tau}}$, divide by $(\eta z_j L_j^{\eta-1}\alpha_{mj})^{1-\tau}$

$$\tilde{\mu_j}\left(1+\frac{1}{1+\beta}\frac{1}{\alpha_{mj}^{1-\tau}N_{mj}}\left(\frac{\tilde{\zeta}_{m,j}}{\lambda}-\frac{\tilde{\zeta_{m-1,j}}}{\lambda}\right)\right)=\frac{\beta}{1+\beta}(\eta z_j L_j^{\eta-1}\alpha_{mj})^{\tau}(1-\tau)^{\frac{\epsilon}{1+\epsilon}}+\frac{\Delta_m}{(\eta z_j L_j^{\eta-1}\alpha_{mj})^{1-\tau}} \tag{A.109}$$

To cancel out the multipliers for no-discrimination constraint $\tilde{\zeta}$, multiply by $\frac{\alpha_{mj}^{1-\tau}N_{mj}}{\sum_k \alpha_{kj}^{1-\tau}N_{kj}}$ and sum over m.

$$\tilde{\mu}_j=\frac{\sum_m \alpha_{mj}^{1-\tau}N_{mj}(\eta z_j L_j^{\eta-1}\alpha_{mj})^{\tau}}{\sum_m \alpha_{mj}^{1-\tau}N_{mj}}(1-\tau)^{\frac{\epsilon}{1+\epsilon}}\frac{\beta}{1+\beta}+\sum_m\frac{\alpha_{mj}^{1-\tau}N_{mj}}{\sum_k \alpha_{kj}^{1-\tau}N_{kj}}\frac{\Delta_m}{(\eta z_j L_j^{\eta-1}\alpha_{mj})^{1-\tau}} \tag{A.110}$$

Using the fact that $MR_{mj}=\eta z_j L_j^{\eta-1}\alpha_{mj}$, $L_{mj}=N_{mj}(1-\tau)^{\frac{\epsilon}{1+\epsilon}}$ and $y_j=z_j L_j^{\eta}$ we have

$$\tilde{\mu}_j=\frac{\eta y_j}{\sum_m MR_{mj}^{1-\tau}L_{mj}}(1-\tau)^{\frac{\epsilon}{1+\epsilon}}\frac{\beta}{1+\beta}+\sum_m\frac{MR_{mj}^{1-\tau}L_{mj}}{\sum_k MR_{kj}^{1-\tau}L_{kj}}\frac{\Delta_m}{MR_{mj}^{1-\tau}} \tag{A.111}$$

which gives us the equation (A.140).

We proceed to obtain equations for the lagrange multipliers ν_m and λ, which are used to define Δ_m. First order condition with respect to I_m is given by

$$\nu_m-\lambda\beta\sum_j\frac{N_{mj}}{I_m}(\eta z_j L_j^{\eta-1}\alpha_{mj}(1-\tau)^{\frac{\epsilon}{1+\epsilon}}-c_{mj})=0 \tag{A.112}$$

Dividing by λ and rearranging yields

$$\frac{\nu_m}{\lambda}=\beta\frac{\sum_j(MR_{mj}(1-\tau)^{\frac{\epsilon}{1+\epsilon}}-c_{mj})N_{mj}}{I_m} \tag{A.113}$$

where we used $MR_{mj}=\eta z_j L_j^{\eta-1}\alpha_{mj}$.

First order condition with respect to c_{fj} is

$$\frac{\psi_f}{c_{fj}}=\lambda \tag{A.114}$$

or equivalently $c_{fj} = \frac{\psi_f}{\lambda}$. Recall that the formula for c_{mj} is given by (A.108). Multiplying by N_{mj} yields

$$\begin{aligned} c_{mj}N_{mj} + \frac{\tilde{\mu}_j(\eta z_j L_j^{\eta-1})^{1-\tau}}{1+\beta}\left(\frac{\tilde{\zeta}_{mj}}{\lambda} - \frac{\tilde{\zeta}_{m-1,j}}{\lambda}\right) \\ = \frac{\beta}{1+\beta}\eta z_j L_j^{\eta-1}\alpha_{mj}L_{mj} + \frac{N_{mj}}{\lambda(1+\beta)} - \frac{\beta}{1+\beta}\frac{\sum_j(\eta z_j L_j^{\eta-1}\alpha_{mj}(1-\tau)^{\frac{\epsilon}{1+\epsilon}} - c_{mj})N_{mj}}{N_m}N_{mj} \end{aligned} \tag{A.115}$$

where we used $\Delta_m = \frac{1}{\lambda(1+\beta)}\left(1 - \nu_m \frac{I_m}{N_m}\right)$ and equation (A.113). Summing up over m, j yields

$$\sum_j \sum_m c_{mj}N_{mj} = \frac{N}{\lambda} \tag{A.116}$$

Substituting $\frac{\psi_f}{c_{fj}} = \lambda$ and $\sum_j \sum_m c_{mj}N_{mj} = \frac{N}{\lambda}$ to the resource constraint yields $\frac{1}{\lambda} = \frac{\sum_j z_j L_j^\eta}{N + \psi_f J}$. This completes the proof for the optimal markdown formula.

We now prove the formula for the optimal progressivity of the labor income tax. Taking first order conditions with respect to τ yields

$$\lambda\eta\left(\sum_j z_j L_j^\eta\right)\frac{\epsilon}{1+\epsilon}\frac{1}{1-\tau} = N\frac{\epsilon}{1+\epsilon} + \sum_j \sum_{m=1}^{M-1} \tilde{\zeta}_{mj}\tilde{\mu}_j(\eta z_j L_j^{\eta-1})^{1-\tau}\log\frac{\alpha_{m+1,j}}{\alpha_{mj}} \tag{A.117}$$

Multiplying by $(1-\tau)\frac{1}{N}\frac{1+\epsilon}{\epsilon}$ and rearrainging yields

$$1 - \tau = \frac{\lambda\frac{\eta\sum_j y_j}{N}}{1 + \frac{1+\epsilon}{\epsilon}\frac{\sum_j\sum_{m=1}^{M-1}\zeta_{mj}\tilde{\mu}_j\log\frac{MR_{m+1,j}}{MR_{mj}}}{N}} \tag{A.118}$$

where we used $MR_{mj} = \eta z_j L_j^{\eta-1}\alpha_{mj}$, $\zeta_{mj} = \tilde{\zeta}_{mj} \times (\eta z_j L_j^{\eta-1})^{1-\tau}$ and $y_j = z_j L_j^\eta$. This proves the equation (A.101).

Decomposition Formula

We derive and present the decomposition of $\frac{d\mu}{d\log z}$ for the model with progressive labor income tax. Recall that the optimal markdown formula is given by

$$\tilde{\mu}_j = \frac{\eta y_j}{\sum_m MR_{mj}^{1-\tau}L_{mj}}(1-\tau)^{\frac{\epsilon}{1+\epsilon}}\frac{\beta}{1+\beta} + \sum_m\left(\frac{MR_{mj}^{1-\tau}L_{mj}}{\sum_k MR_{kj}^{1-\tau}L_{kj}}\frac{\Delta_m}{MR_{mj}^{1-\tau}}\right) \tag{A.119}$$

where $\tilde{\mu}_j = \phi \times \mu_j^{1-\tau}$. Rearranging the terms, we have

$$\begin{aligned} \tilde{\mu}_j &= A_j + \frac{1}{\left(\eta z_j \bar{L}_j^{\eta-1}\left(\sum_m \alpha_{mj}\frac{L_{mj}}{\bar{L}_j}\right)^{\eta-1}\right)^{1-\tau}}\frac{\sum_m \Delta_m \frac{L_{mj}}{\bar{L}_j}}{\sum_m \alpha_{mj}^{1-\tau}\frac{L_{mj}}{\bar{L}_j}}, \\ \text{where} \quad A_j &= \sum_m\left(\frac{\alpha_{mj}^{1-\tau}N_{mj}}{\sum_k \alpha_{kj}^{1-\tau}N_{kj}}MR_{mj}^\tau\right)(1-\tau)^{\frac{\epsilon}{1+\epsilon}}\frac{\beta}{1+\beta} \end{aligned} \tag{A.120}$$

where $\bar{L}_j$ is the total number of employees working at firm j.
The formula for the decomposition of $\frac{d\mu}{d\log z}$ is given as

$$\frac{d\mu_j}{d\log z_j} = F_\tau + F1_z + F1_\alpha + F2, \tag{A.121}$$

$$\text{where} \quad F_\tau = \frac{1}{1-\tau} \times A_j \frac{\mu_j}{\tilde{\mu}_j} \times \frac{d\log A_j}{d\log z_j}, \tag{A.122}$$

$$F1_z = -\left(\mu_j - A_j \frac{\mu_j}{\tilde{\mu}_j}\right) \times \left[1 - (1-\eta) \times \frac{d\log \bar{L}_j}{d\log z_j}\right], \tag{A.123}$$

$$F1_\alpha = -\left(\mu_j - A_j \frac{\mu_j}{\tilde{\mu}_j}\right) \times \left[\sum_m \left(\left(\frac{\alpha_{mj}^{1-\tau} N_{mj}}{\sum_k \alpha_{kj}^{1-\tau} L_{kj}} - (1-\eta)\frac{\alpha_{mj} N_{mj}}{\sum_k \alpha_{kj} N_{kj}}\right) \times \frac{d\log \alpha_{mj}}{d\log z_j}\right)\right], \tag{A.124}$$

$$F2 = \left(\mu_j - A_j \frac{\mu_j}{\tilde{\mu}_j}\right) \times \sum_m \left[\left(\frac{1}{1-\tau} \times \frac{\alpha_{mj}^{1-\tau} N_{mj}}{\sum_k \alpha_{kj}^{1-\tau} N_{mj}} - (1-\eta) \times \frac{\alpha_{mj} L_{mj}}{\sum_k \alpha_{kj} L_{kj}} - \frac{1}{1-\tau} \times \frac{\Delta_m L_{mj}}{\sum_k \Delta_k L_{kj}}\right) \times \frac{d\log \frac{L_{mj}}{\bar{L}_j}}{d\log z_j}\right] \tag{A.125}$$

To calculate $\frac{d\mu_j}{\log z_j}$, we first calculate $\frac{d\tilde{\mu}_j}{\log z_j}$ and then multiply by $\frac{1}{1-\tau}\frac{\mu_j}{\tilde{\mu}_j}$. We start from the derivation of F_τ, which is the term that captures the extent to which the progressive labor income tax directly replaces the role of firm-side subsidies in doing redistribution. Taking derivatives, we have

$$\frac{\partial \tilde{\mu}_j}{\partial A_j}\frac{dA_j}{dz_j} = \frac{dA_j}{dz_j} \tag{A.126}$$

Multiplying by z_j and substituting dA_j, dz_j with $A_j d\log A_j, z_j d\log z_j$ we have

$$\frac{\partial \tilde{\mu}_j}{\partial \log A_j}\frac{d\log A_j}{d\log z_j} = A_j \frac{d\log A_j}{d\log z_j} \tag{A.127}$$

Multiplying by $\frac{1}{1-\tau}\frac{\mu_j}{\tilde{\mu}_j}$ and using $d\mu_j = \frac{1}{1-\tau}\frac{\mu_j}{\tilde{\mu}_j}$ yields

$$\frac{\partial \mu_j}{\partial \log A_j}\frac{d\log A_j}{d\log z_j} = \frac{1}{1-\tau} A_j \frac{\mu_j}{\tilde{\mu}_j}\frac{d\log A_j}{d\log z_j} \tag{A.128}$$

which gives us expression for F_τ.

We derive $F1_z$, which captures the direct effect of increasing z on the optimal markdown schedule. Taking derivatives, we have

$$\frac{\partial \tilde{\mu}_j}{\partial z_j} + \frac{\partial \tilde{\mu}_j}{\partial \bar{L}_j}\frac{d\bar{L}_j}{dz_j} = -\frac{1}{\left(\eta z_j \bar{L}_j^{\eta-1}\left(\sum_m \alpha_{mj}\frac{L_{mj}}{\bar{L}_j}\right)^{\eta-1}\right)^{1-\tau}} \frac{\sum_m \Delta_m \frac{L_{mj}}{\bar{L}_j}}{\sum_m \alpha_{mj}^{1-\tau}\frac{L_{mj}}{\bar{L}_j}} \times \left(\frac{1-\tau}{z_j} - \frac{1-\eta}{\bar{L}_j}\frac{d\bar{L}_j}{dz_j}\right) \tag{A.129}$$

Multiplying by z_j, replacing $d\bar{L}_j, dz_j$ with $\bar{L}_j d\log\bar{L}_j, z_j \log z_j$ and substituting $\frac{1}{\left(\eta z_j \bar{L}_j^{\eta-1}\left(\sum_m \alpha_{mj}\frac{L_{mj}}{\bar{L}_j}\right)^{\eta-1}\right)^{1-\tau}}\frac{\sum_m \Delta_m}{\sum_m \alpha_m^{1-}}$ $\tilde{\mu}_j - A_j$ yields

$$\frac{\partial\tilde{\mu}_j}{\partial\log z_j} + \frac{\partial\tilde{\mu}_j}{\partial\log\bar{L}_j}\frac{d\log\bar{L}_j}{d\log z_j} = -(\tilde{\mu}_j - A_j)\times(1-\tau)\times\left(1-(1-\eta)\times\frac{d\log\bar{L}_j}{d\log z_j}\right) \tag{A.130}$$

Multiplying by $\frac{1}{1-\tau}\frac{\mu_j}{\tilde{\mu}_j}$ and using $d\mu_j = \frac{1}{1-\tau}\frac{\mu_j}{\tilde{\mu}_j}$ yields

$$\frac{\partial\mu_j}{\partial\log z_j} + \frac{\partial\mu_j}{\partial\log\bar{L}_j}\frac{d\log\bar{L}_j}{d\log z_j} = -(\mu_j - A_j\frac{\mu_j}{\tilde{\mu}_j})\times\left(1-(1-\eta)\times\frac{d\log\bar{L}_j}{d\log z_j}\right) \tag{A.131}$$

which gives our expression for $F1_z$.

$F1_\alpha$ captures the indirect effect of increasing z on the markdown schedule through α. Taking derivatives with respect to α_{mj}, we have

$$\begin{aligned}\frac{\partial\tilde{\mu}_j}{\partial\alpha_{mj}}\frac{d\alpha_{mj}}{dz_j} =& -\frac{1}{\left(\eta z_j\bar{L}_j^{\eta-1}\left(\sum_m\alpha_{mj}\frac{L_{mj}}{\bar{L}_j}\right)^{\eta-1}\right)^{1-\tau}}\frac{\sum_m\Delta_m\frac{L_{mj}}{\bar{L}_j}}{\sum_m\alpha_{mj}^{1-\tau}}\\ &\times\left((1-\tau)\times\frac{\alpha_{mj}^{-\tau}\frac{L_{mj}}{\bar{L}_j}}{\sum_k\alpha_{kj}^{1-\tau}\frac{L_{kj}}{\bar{L}_j}}\frac{d\alpha_{mj}}{dz_j} - (1-\tau)\times(1-\eta)\times\frac{\frac{L_{mj}}{\bar{L}_j}}{\sum_k\alpha_{kj}\frac{L_{kj}}{\bar{L}_j}}\frac{d\alpha_{mj}}{dz_j}\right)\end{aligned} \tag{A.132}$$

Multiplying by z_j, replacing $d\bar{\alpha}_{mj}, dz_j$ with $\alpha_{mj}d\log\alpha_{mj}, z_j\log z_j$ and substituting $\frac{1}{\left(\eta z_j\bar{L}_j^{\eta-1}\left(\sum_m\alpha_{mj}\frac{L_{mj}}{\bar{L}_j}\right)^{\eta-1}\right)^{1-\tau}}\frac{\sum}{\sum}$ $\tilde{\mu}_j - A_j$ yields

$$\frac{\partial\tilde{\mu}_j}{\partial\alpha_{mj}}\frac{d\alpha_{mj}}{dz_j} = -(\tilde{\mu}_j - A_j)\times\left((1-\tau)\times\frac{\alpha_{mj}^{1-\tau}L_{mj}}{\sum_k\alpha_{kj}^{1-\tau}L_{kj}}\frac{d\log\alpha_{mj}}{d\log z_j} - (1-\tau)\times(1-\eta)\times\frac{\alpha_{mj}L_{mj}}{\sum_k\alpha_{kj}L_{kj}}\frac{d\log\alpha_{mj}}{d\log z_j}\right) \tag{A.133}$$

Summing up over m gives

$$\sum_m\left(\frac{\partial\tilde{\mu}_j}{\partial\alpha_{mj}}\frac{d\alpha_{mj}}{dz_j}\right) = -(\tilde{\mu}_j - A_j)\times(1-\tau)\times\sum_m\left[\left(\frac{\alpha_{mj}^{1-\tau}L_{mj}}{\sum_k\alpha_{kj}^{1-\tau}L_{kj}} - (1-\eta)\frac{\alpha_{mj}L_{mj}}{\sum_k\alpha_{kj}L_{kj}}\right)\times\frac{d\log\alpha_{mj}}{d\log z_j}\right] \tag{A.134}$$

Multiplying by $\frac{1}{1-\tau}\frac{\mu_j}{\tilde{\mu}_j}$ and using $d\mu_j = \frac{1}{1-\tau}\frac{\mu_j}{\tilde{\mu}_j}$ yields

$$\sum_m\left(\frac{\partial\mu_j}{\partial\alpha_{mj}}\frac{d\alpha_{mj}}{dz_j}\right) = -(\mu_j - A_j\frac{\mu_j}{\tilde{\mu}_j})\times(1-\tau)\times\sum_m\left[\left(\frac{\alpha_{mj}^{1-\tau}L_{mj}}{\sum_k\alpha_{kj}^{1-\tau}L_{kj}} - (1-\eta)\frac{\alpha_{mj}L_{mj}}{\sum_k\alpha_{kj}L_{kj}}\right)\times\frac{d\log\alpha_{mj}}{d\log z_j}\right] \tag{A.135}$$

which gives our expression for $F1_\alpha$.

Finally, we derive the term $F2$. Taking derivatives with respect to $\frac{L_{mj}}{\bar{L}_j}$,

$$\frac{\partial\tilde{\mu}_j}{\partial\frac{L_{mj}}{\bar{L}_j}}\frac{d\frac{L_{mj}}{\bar{L}_j}}{dz_j} = -\frac{1}{\left(\eta z_j \bar{L}_j^{\eta-1}\left(\sum_m \alpha_{mj}\frac{L_{mj}}{\bar{L}_j}\right)^{\eta-1}\right)^{1-\tau}}\frac{\sum_m \Delta_m \frac{L_{mj}}{\bar{L}_j}}{\sum_m \alpha_{mj}^{1-\tau}}$$
$$\times\left[\frac{\alpha_{mj}^{1-\tau}}{\sum_k \alpha_{kj}^{1-\tau}\frac{L_{kj}}{\bar{L}_j}}\frac{d\frac{L_{mj}}{\bar{L}_j}}{dz_j} - \frac{\Delta_m}{\sum_k \Delta_{kj}\frac{L_{kj}}{\bar{L}_j}}\frac{d\frac{L_{mj}}{\bar{L}_j}}{dz_j} - (1-\tau)(1-\eta)\frac{\alpha_{mj}}{\sum_k \alpha_{kj}\frac{L_{kj}}{\bar{L}_j}}\frac{d\frac{L_{mj}}{\bar{L}_j}}{dz_j}\right] \quad \text{(A.136)}$$

Multiplying by z_j, replacing $d\frac{L_{mj}}{\bar{L}_j}, dz_j$ with $\frac{L_{mj}}{\bar{L}_j}d\log\frac{L_{mj}}{\bar{L}_j}$, $z_j\log z_j$ and substituting $\frac{1}{\left(\eta z_j \bar{L}_j^{\eta-1}\left(\sum_m \alpha_{mj}\frac{L_{mj}}{\bar{L}_j}\right)^{\eta-1}\right)^{1-\tau}}\frac{\sum$ $\tilde{\mu}_j - A_j$ yields

$$\frac{\partial\tilde{\mu}_j}{\partial\log\frac{L_{mj}}{\bar{L}_j}}\frac{d\log\frac{L_{mj}}{\bar{L}_j}}{d\log z_j} = -(\tilde{\mu}_j - A_j)$$
$$\times\left[\frac{\alpha_{mj}^{1-\tau}L_{mj}}{\sum_k \alpha_{kj}^{1-\tau}L_{kj}}\frac{d\log\frac{L_{mj}}{\bar{L}_j}}{d\log z_j} - \frac{\Delta_m L_{mj}}{\sum_k \Delta_{kj}L_{kj}}\frac{d\log\frac{L_{mj}}{\bar{L}_j}}{d\log z_j} - (1-\tau)(1-\eta)\frac{\alpha_{mj}L_{mj}}{\sum_k \alpha_{kj}L_{kj}}\frac{d\log\frac{L_{mj}}{\bar{L}_j}}{d\log z_j}\right] \quad \text{(A.137)}$$

Summing up over m and n gives

$$\sum_m\left(\frac{\partial\tilde{\mu}_j}{\partial\log\frac{L_{mj}}{\bar{L}_j}}\frac{d\log\frac{L_{mj}}{\bar{L}_j}}{d\log z_j}\right) = -(\tilde{\mu}_j - A_j)$$
$$\times\sum_m\left[\left(\frac{\alpha_{mj}^{1-\tau}L_{mj}}{\sum_k \alpha_{kj}^{1-\tau}L_{kj}} - \frac{\Delta_m L_{mj}}{\sum_k \Delta_{kj}L_{kj}} - (1-\tau)(1-\eta)\frac{\alpha_{mj}L_{mj}}{\sum_k \alpha_{kj}L_{kj}}\right)\times\frac{d\log\frac{L_{mj}}{\bar{L}_j}}{d\log z_j}\right] \quad \text{(A.138)}$$

Multiplying by $\frac{1}{1-\tau}\frac{\mu_j}{\tilde{\mu_j}}$ and using $d\mu_j = \frac{1}{1-\tau}\frac{\mu_j}{\tilde{\mu}_j}$ yields

$$\sum_m\left(\frac{\partial\mu_j}{\partial\log\frac{L_{mj}}{\bar{L}_j}}\frac{d\log\frac{L_{mj}}{\bar{L}_j}}{d\log z_j}\right) = -(\mu_j - A_j\frac{\mu_j}{\tilde{\mu}_j})$$
$$\times\sum_m\left[\left(\frac{1}{1-\tau}\frac{\alpha_{mj}^{1-\tau}L_{mj}}{\sum_k \alpha_{kj}^{1-\tau}L_{kj}} - \frac{1}{1-\tau}\frac{\Delta_m L_{mj}}{\sum_k \Delta_{kj}L_{kj}} - (1-\eta)\frac{\alpha_{mj}L_{mj}}{\sum_k \alpha_{kj}L_{kj}}\right)\times\frac{d\log\frac{L_{mj}}{\bar{L}_j}}{d\log z_j}\right] \quad \text{(A.139)}$$

which is the $F2$ term.

C.3.2 Algorithm for Computing the Optimal Allocation in the Presence of Income Tax

In this section we explain how we numerically compute the optimal markdown schedule and progressivity. We first calculate the optimal markdown schedule and associated endogeneous variables,

taking the level of progressivity τ as given. Then we calculate the associated utilitarian social welfare and find the optimal progressivity using the grid search method. [6]

To compute the optimal markdown, we use the following formula

$$\tilde{\mu}_j = \frac{\eta y_j}{\sum_m MR_{mj}^{1-\tau} L_{mj}} (1-\tau)^{\frac{\epsilon}{1+\epsilon}} \frac{\beta}{1+\beta} + \sum_m \left(\frac{MR_{mj}^{1-\tau} L_{mj}}{\sum_k MR_{kj}^{1-\tau} L_{kj}} \frac{\Delta_m}{MR_{mj}^{1-\tau}} \right),$$
$$\text{where} \quad \Delta_m = \frac{1}{1+\beta} \frac{\sum_j y_j}{N + \psi_f J} - \frac{\beta}{1+\beta} \frac{\sum_j (MR_{mj} L_{mj} - c_{mj} N_{mj})}{N_m} \tag{A.140}$$

The algorithm is as follows : First, start with the guess for $\{\mu_j\}$. Second, obtain the values of $\{L_j\}$ which are consistent with our guess of the (possibly non-optimal) markdown scheudle $\{\tilde{\mu}_j\}$. Third, calculate the rest of the endogenous variables $\{c_{mj}\}, \{N_{mj}\}, \{L_{mj}\}, \{MR_{mj}\}, \{y_j\}$ and update $\{\tilde{\mu}_j\}$ by calculating the right hand side of equation (A.140). Fourth, iterate on $\{\tilde{\mu}_j\}$ until convergence. Last, recover $\{\mu_j\}$ from $\{\tilde{\mu}_j\}$.

We explain how to calculate $\{L_j\}$ which are consistent with the (possibly non-optimal) markdown schedule $\{\tilde{\mu}_j\}$. We only make use of the fact that workers optimally choose their preferred workplace after observing the wages and amenities of each firm. Since $\tilde{\mu}_j = \frac{c_{mj}}{MR_{mj}^{1-\tau}}$, we can express c_{mj} as

$$\begin{aligned} c_{mj} &= \tilde{\mu}_j MR_{mj}^{1-\tau} \\ &= \tilde{\mu}_j \eta^{1-\tau} z_j^{1-\tau} L_j^{-(1-\tau)(1-\eta)} \alpha_{mj}^{1-\tau} \end{aligned} \tag{A.141}$$

Worker's optimization decisions imply that given $\{c_{mj}\}$, type m workers' labor supply to firm j is equal to

$$N_{mj} = N_m \frac{c_{mj}^\beta g_{mj}^\beta}{\sum_k (c_{mk}^\beta g_{mk}^\beta)} = N_m \frac{\tilde{\mu}_j^\beta z_j^{\beta(1-\tau)} L_j^{-\beta(1-\tau)(1-\eta)} \alpha_{mj}^{\beta(1-\tau)}}{\sum_{j'} (\tilde{\mu}_{j'}^\beta z_{j'}^{\beta(1-\tau)} L_{j'}^{-\beta(1-\tau)(1-\eta)} \alpha_{mj'}^{\beta(1-\tau)})} \tag{A.142}$$

Using the fact that $L_j = (1-\tau)^{\frac{\epsilon}{1+\epsilon}} \times \sum_m (\alpha_{mj} N_{mj})$, we have

$$\begin{aligned} L_j &= (1-\tau)^{\frac{\epsilon}{1+\epsilon}} \times \sum_m (\alpha_{mj} N_{mj}) \\ &= (1-\tau)^{\frac{\epsilon}{1+\epsilon}} \times \sum_m \left(N_m \frac{\tilde{\mu}_j^\beta z_j^{\beta(1-\tau)} L_j^{-\beta(1-\tau)(1-\eta)} \alpha_{mj}^{1+\beta(1-\tau)}}{\sum_{j'} (\tilde{\mu}_{j'}^\beta z_{j'}^{\beta(1-\tau)} L_{j'}^{-\beta(1-\tau)(1-\eta)} \alpha_{mj'}^{\beta(1-\tau)})} \right) \end{aligned} \tag{A.143}$$

Define a new variable $H_j = L_j \tilde{\mu}_j^{-\frac{\beta}{1+\beta(1-\eta)(1-\tau)}} z_j^{-\frac{\beta(1-\tau)}{1+\beta(1-\eta)(1-\tau)}}$ and use H_j to substitute for L_j. Simple algebra shows that

$$H_j = \left[(1-\tau)^{\frac{\epsilon}{1+\epsilon}} \times \sum_m \left(N_m \frac{\alpha_{mj}^{1+\beta(1-\tau)} g_{mj}^\beta}{\sum_{j'} \left(\tilde{\mu}_{j'}^{\frac{\beta}{1+\beta(1-\eta)(1-\tau)}} z_{j'}^{\frac{\beta(1-\tau)}{1+\beta(1-\eta)1-\tau}} H_{j'}^{-\beta(1-\tau)(1-\eta)} \alpha_{mj'}^{\beta(1-\tau)} g_{mj'}^\beta \right)} \right) \right]^{\frac{1}{1+\beta(1-\tau)(1-\eta)}} \tag{A.144}$$

[6] The utilitarian social welfare is given by $\sum_m \left(\frac{N_m}{\beta} \log \left(\sum_j c_{mj}^\beta g_{mj}^\beta \right) \right) - N \frac{\epsilon(1-\tau)}{1+\epsilon}$

Define $\tilde{H}_j = H_j^{1-\eta}$ and we have

$$\tilde{H}_j = \left[(1-\tau)^{\frac{\epsilon}{1+\epsilon}} \times \sum_m \left(N_m \frac{\alpha_{mj}^{1+\beta(1-\tau)} g_{mj}^{\beta}}{\sum_{j'} \left(\tilde{\mu}_{j'}^{\frac{\beta}{1+\beta(1-\eta)(1-\tau)}} z_{j'}^{\frac{\beta(1-\tau)}{1+\beta(1-\eta)(1-\tau)}} \tilde{H}_{j'}^{-\beta(1-\tau)} \alpha_{mj'}^{\beta(1-\tau)} g_{mj'}^{\beta} \right)} \right) \right]^{\frac{1-\eta}{1+\beta(1-\tau)(1-\eta)}} \tag{A.145}$$

We numerically solve the fixed point equation (A.145) to obtain the sequence of $\{\tilde{H}_j\}$. [7] We recover $\{H_j\}$ from $H_j = \tilde{H}_j^{\frac{1}{1-\eta}}$ and use $L_j = H_j \tilde{\mu}_j^{\frac{\beta}{1+\beta(1-\eta)(1-\tau)}} z_j^{\frac{\beta(1-\tau)}{1+\beta(1-\eta)(1-\tau)}}$ to recover $\{L_j\}$. We calculate $\{MR_{mj}, y_j\}$ using $MR_{mj} = \eta z_j L_j^{\eta-1} \alpha_{mj}$ and $y_j = z_j L_j^{\eta}$. We recover $\{c_{mj}\}$ using $c_{mj} = \tilde{\mu}_j MR_{mj}^{1-\tau}$ and use $N_{mj} = N_m \frac{c_{mj}^{\beta} g_{mj}^{\beta}}{\sum_{j'} c_{mj'}^{\beta} g_{mj'}^{\beta}}, L_{mj} = N_{mj} \times (1-\tau)^{\frac{\epsilon}{1+\epsilon}}$ to calculate $\{N_{mj}, L_{mj}\}$. Using the aforementioned variables we can update $\{\tilde{\mu}_j\}$ by calculating the right hand side of the equation (A.140).

It remains to recover $\{\mu_j\}$ from $\{\tilde{\mu}_j\}$. Note that

$$\mu_j = \left(\frac{\tilde{\mu}_j}{\phi} \right)^{\frac{1}{1-\tau}} \tag{A.146}$$

so we need to find the value of ϕ which clears the government's budget. We assume that the government operates a separate budget for firm-side subsidy and progressive labor income tax, i.e.

$$\sum_j S_j(y_j) = 0 \tag{A.147}$$

Since $c_{fj} = y_j + S_j(y_j) - \sum_m w_{mj} L_{mj}$, by substituting $S_j(y_j)$ we have

$$\sum_j \sum_m w_{mj} L_{mj} + \sum_j c_{fj} = \sum_j y_j \tag{A.148}$$

Using $\sum_j c_{fj} = \frac{J\psi_f}{\lambda} = \frac{J\psi_f}{N+J\psi_f} \sum_j y_j$

$$\sum_j \sum_m w_{mj} L_{mj} = \frac{N}{N+J\psi_f} \sum_j y_j \tag{A.149}$$

Substituting $w_{mj} = \mu_j MR_{mj} = \left(\frac{\tilde{\mu}_j}{\phi} \right)^{\frac{1}{1-\tau}} MR_{mj}$ and rearranging term yields

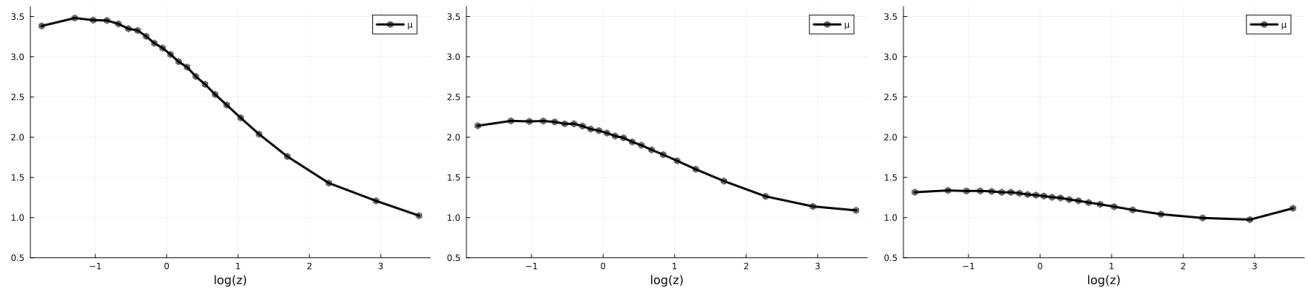
$$\phi = \left(\frac{\sum_j \sum_m MR_{mj} L_{mj} \tilde{\mu}_j^{\frac{1}{1-\tau}}}{\frac{N}{N+\psi_f J} \sum_j y_j} \right)^{1-\tau} \tag{A.150}$$

Using the calculated value of ϕ, we can recover $\{\mu_j\}$.

[7] Lamadon, Mogstad, and Setzler (2022) provides the proof for the convergence of the fixed point equation (A.145).

C.3.3 Quantitative Investigation of the Joint Optimal Policy

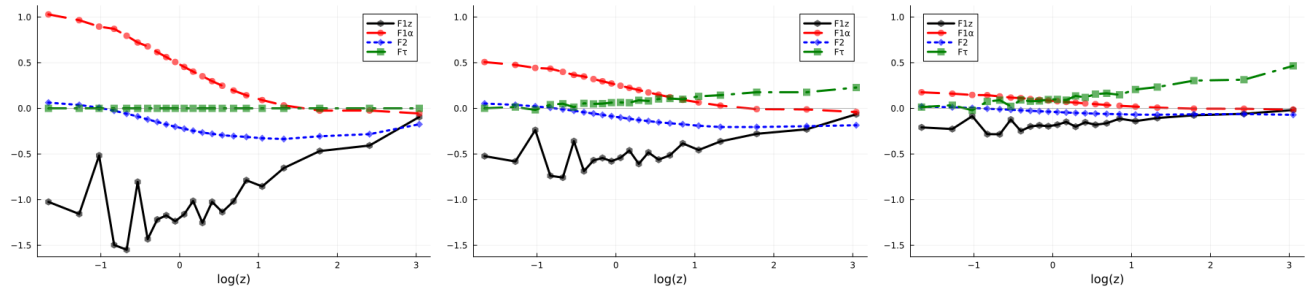
In this section we present the results from the numerical exercises with a progressive labor income tax. Throughout our numerical analysis, we assume that the Frisch elasticity of labor supply is given by 0.5.



(a) $\tau = 0.0$ (b) $\tau = 0.25$ (c) $\tau = 0.5$

Figure A.9: Optimal Markdown, varying τ

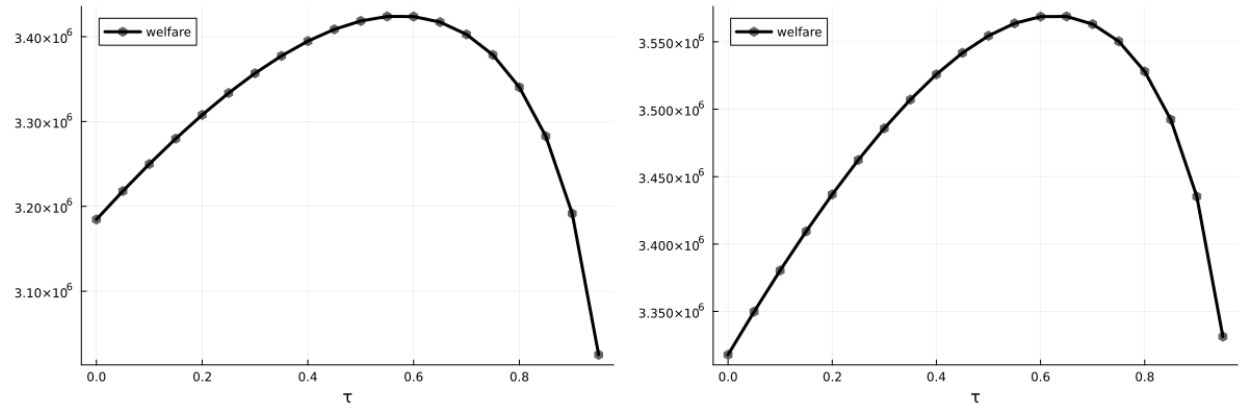
We first compute the optimal markdown schedule and the decomposition of its slope for a given value of labor income tax progressivity. Figure A.9a, Figure A.9b and Figure A.9c plot the optimal markdown schedule when the labor income tax progressivity is given by $\tau = 0.0, 0.25, 0.5$. Note that as labor income tax progressivity increases, the optimal markdown schedule becomes less steeper across the firm's productivity distribution implying less wage compression with higher progressivity.



(a) $\tau = 0.0$ (b) $\tau = 0.25$ (c) $\tau = 0.5$

Figure A.10: Decomposition, varying τ

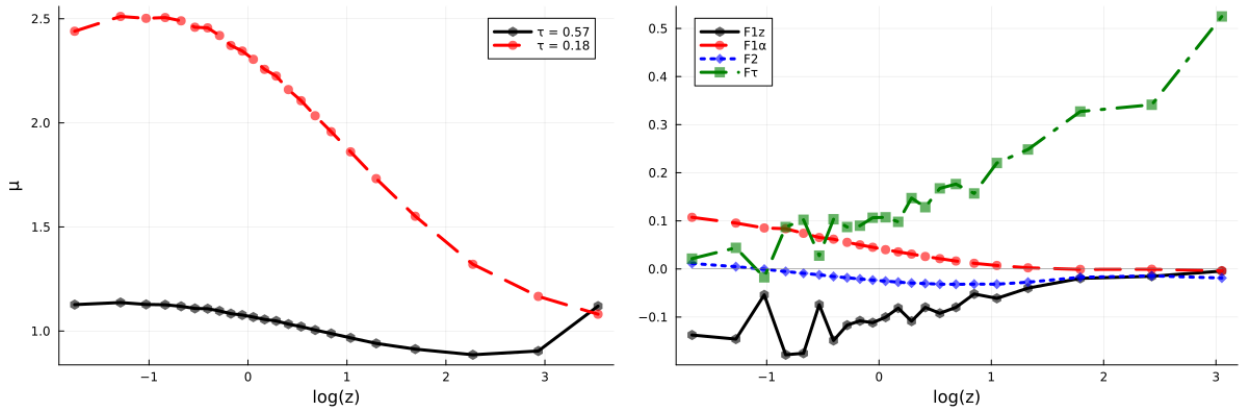
Figure A.10a, Figure A.10b and Figure A.10c plot the decomposition of the slope of the optimal pseudo markdown under the labor income tax progressivity $\tau = 0.0, 0.25, 0.5$. Observe that the magnitude of the direct substitution term, F_τ, increases relative to the other terms ($F1_z$, $F1_\alpha$ and $F2$) as τ increases. The planner can provide more redistribution across workers with more progressive income tax, and with more substitution, the planner sets less steeper markdown schedule across firms.



(a) $\epsilon_h = 0.5$ (b) $\epsilon_h = 0.25$

Figure A.11: Social Welfare

We now present the optimal labor tax progressivity in the joint optimal policy. Figure A.11a and Figure A.11b plot the utilitarian social welfare of workers for each value of labor income tax progressivity τ under the Frisch elasticity $\epsilon_h = 0.5$ and 0.25 respectively.[8] Under $\epsilon = 0.5$, the social welfare attains its maximum at $\tau = 0.57$, which is considerably larger than the current labor income tax progressivity level. Under the lower $\epsilon = 0.25$, distortionary effects of taxes on the labor supply are smaller, leading to higher optimal labor income tax progressivity $\tau = 0.63$.



(a) Markdown (b) Decomposition

Figure A.12: Optimal Progressivity, $\tau = 0.18$

Figure A.12a plots the markdown schedule across the firm's productivity distribution under the optimal labor income tax progressivity level $\tau = 0.57$ compared to the current US progressivity level $\tau = 0.18$. Compared to the markdown schedule under the current US progressivity level, the optimal markdown is considerably less steeper. Figure A.12b plots the decomposition of the slope

[8] We set the social welfare weight on firm owners, ψ_f, to zero.

of the optimal markdown under $\tau = 0.57$. Extremly large magnitude of $F\tau$ demonstrates that the planner does the majority of the redistribution using progressive labor income tax, resulting in a flat schedule of optimal markdown.

D Appendix for Section 7.2

D.1 Economy with Single Skill Type

In Section 7.2 of the main text, we have investigated the extension to the model with product market power. We provide detailed derivation and further investigation in this section.

We first derive the planner's optimality condition (34). The first order condition of the Lagrangian of the planner's problem (33) with respect to consumption at firm j c_j yields

$$\frac{N}{\beta} \cdot \frac{\beta c_j^{\beta-1}}{\sum_j c_j^{\beta}} + \lambda \times \left[y^{\frac{1}{\theta}} \cdot y_j^{-\frac{1}{\theta}} \cdot z_j \cdot \frac{dL_j}{dc_j} - L_j - w_j \frac{dL_j}{dc_j} \right] - \gamma \cdot \frac{dL_j}{dc_j} = 0.$$

By multiplying $\frac{c_j}{L_j}$ on both sides of the first order condition, and using $\frac{N}{L_j} \cdot \frac{c_j^{\beta}}{I} = 1$ and $\frac{dL_j}{dc_j} \cdot \frac{c_j}{L_j} = \beta$, we obtain

$$\lambda(1+\beta)c_j = \lambda\beta \left(\frac{y}{y_j} \right)^{\frac{1}{\theta}} z_j + (1 - \gamma\beta),$$

which yields (34).

Implication of Product Market Power on the Degree of Wage Compression

What is the implication of introducing product market power on the degree of wage compression? The answer depends on the reference point of the wage compression. For example, with linear production, the slope of markdown schedule across firms' productivities is determined by

$$\frac{\partial \mu_j^{SP}}{\partial z_j} = -\frac{\Delta}{p_j z_j^2} + \frac{1}{\theta} \times \frac{\Delta}{p_j z_j^2} \times \left(1 + \frac{dlogL_j}{dlogz_j} \right). \tag{A.151}$$

Due to the introduction of imperfectly competitive product market ($\theta < \infty$), additional positive term arises in the slope of markdown expression (A.151), which makes the markdown schedule $\mu^{SP}(z_j)$ flatter.

Does this imply less wage compression? If we think of the wage distribution relative to the distribution of the marginal revenue product of labor, then the answer is yes. As the entrepreneur's products are less substitutable, the efficiency cost of wage compression increases. Thus, more severe equity-efficiency trade-off implies less wage compression. If we consider the wage compression relative to the equilibrium wage distribution, then the answer is ambiguous. The correction of

inverse markup amplifies the cross subsidization, which implies intensified wage compression, thus the answer depends on which force dominates.

Understanding the Result in the Lens of Kroft, Luo, Mogstad, and Setzler (2023)

Kroft, Luo, Mogstad, and Setzler (2023) study the imperfect competition in both labor market and product market jointly. They show that the impacts of market power in one market are attenuated by the market power in other market relative to the case when market power exists only in one market. We now relate our results to the implications of Kroft, Luo, Mogstad, and Setzler (2023).

The solution of the planner's problem (33) can be implemented with the two-step process: (i) step 1: correction of the inverse markup, and (ii) step 2: redistributive subsidization. Consider a planner's problem in the second step, which is the problem after correcting the inverse markup. Then the effect of imperfect competition in the product market ($\theta < \infty$) on the degree of planner's intervention can be understood as the attenuation mechanism emphasized by Kroft, Luo, Mogstad, and Setzler (2023). Compared to the case with labor market power only ($\theta = \infty$), in the presence of market power in both markets, the planner's intervention in the second step—increase in labor supply by the subsidization—is attenuated as increased employment and output will decreases the price of output. We can show this more clearly using the graph.

D.2 Economy with Multiple Skill Types

In Section 7.2 of the main text, we briefly discussed the results with multiple skill type in the presence of product market power. The results can be obtained by solving the planner's problem who faces the same no-differentiation constraint as in the economy with only labor market power. In this section, we provides the motivation for the same NDC in this economy.

Under the assumption that the planner cannot differentiate the subsidy rate between the skill types, the two types of workers working at the same firm face the same subsidy rate $1 + S'(p_j y_j) = \frac{c'(y_j))}{p_j \cdot (1 - \frac{1}{\theta})}$, where

$$\begin{aligned} c(y_j) &= \min_{L_{hj}, L_{\ell j}} w_{hj}(L_{hj}) L_{hj} + w_{\ell j}(L_{\ell j}) L_{\ell j} \\ (\lambda_c) \quad s.t. &\qquad y_j = z_j f(L_{hj}, L_{\ell j}). \end{aligned}$$

By combining the first order conditions of the cost-minimization problem and the envelope condition $c'(y_j) = \lambda_c$, we get

$$c'(y_j) = \left(1 + \frac{1}{\beta}\right) \cdot \frac{w_{sj}}{MPL_{sj}}, \quad \forall s \in \{h, \ell\},$$

and thus, the same subsidy rate of both workers implies the following no-differentiation constraints (NDC)

$$\frac{w_{hj}}{MPL_{hj}} = \frac{w_{\ell j}}{MPL_{\ell j}},$$

which also implies the same optimal markdown $\mu_j^{SP} = \frac{w_{sj}}{p_j \cdot MPL_{sj}}$ for both types of workers working in the same firm. Thus the introduction of the imperfect competition in the product market does not change the NDC, and the only change in the planner's problem is once again the final good production technology in the resource constraint.

From the first order conditions of the planning problem, we get the optimal consumption expression, which is essentially the same as in the benchmark economy with perfectly competitive product market:

$$c_j^{SP} = \frac{1}{\Omega_{sj}} \times \left(\frac{\beta}{1+\beta} \times p_j \cdot MPL_{sj} + \Delta_s \right),$$

for $s \in \{h, \ell\}$. By following the same process as in the benchmark model, we can rewrite the optimal markdown:

$$\mu_j^{SP} = \mu_j^E \times \frac{1}{1-\frac{1}{\theta}} \times \left(1 + s_j(\alpha_j) \cdot \frac{\Delta_h \cdot \frac{1+\beta}{\beta}}{p_j \cdot MPL_{h_j}} + (1 - s_j(\alpha_j)) \cdot \frac{\Delta_l \cdot \frac{1+\beta}{\beta}}{p_j \cdot MPL_{l_j}} \right). \tag{A.152}$$

This expression shows that the introduction of the production market power amplifies the subsidy rate without changing the direction of the original intervention driven by redistribution across firms and across skill types.

References

Kroft, K., Y. Luo, M. Mogstad, and B. Setzler (2023): "Imperfect competition and rents in labor and product markets: The case of the construction industry," Discussion paper.

Lamadon, T., M. Mogstad, and B. Setzler (2022): "Imperfect Competition, Compensating Differentials, and Rent Sharing in the US Labor Market," *American Economic Review*, 112(1), 169–212.