Optimal Redistribution Policy with Labor Market Monopsony*

Hanvit Kim† Yena Park‡

March 24, 2026

Abstract

We study the optimal redistribution policy in the presence of imperfect competition in the labor market compensating differentials. When firms have monopsony power due to workers' idiosyncratic preferences for non-wage characteristics of firms, the key efficiency cost of a redistribution is misallocation of workers across firms. To minimize the cost, a redistribution policy should be designed to generate a pass-through schedule from firm-productivity to wage that increases in firm-productivity, which can be implemented by a firm regulation in the goods market. In the presence of vertical amenity and compensating differentials, however, a redistribution policy generates multi-dimensional trade-offs and it is (constrained) best to subsidize firm's productivity and to (indirectly) tax amenity for more productive firms, whose effectiveness depends on the sources of inequality.

Keywords: Imperfect Labor Market, Monopsony, Redistribution Policy, Firm Regulation
JEL Classification: D61, H21, H24, H25, J22, J31, J42

*We would like to thank Arpad Abraham, Florian Scheuer, Yongsung Chang, Magne Mogstad, Allen Head, Gianluca Violante, Barış Kaymak, André Luduvice, and Kjetil Storesletten for helpful comments. We also appreciate the great comments and suggestions of the participants at various conferences and seminars.

†University of Chicago, Department of Economics, hanvit@uchicago.edu

‡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? The standard framework for the design of redistribution policy is a competitive labor market in which firms take market wages as given. However, there has been an accumulation of evidence on the role of firms in wage inequality. For example, Bonhomme, Holzheu, Lamadon, Manresa, Mogstad, and Setzler (2022) show that firm effects—both firm premium and assortative sorting of workers—explain 30% to 50% of the wage inequality in developed countries.[1] Many studies based on a model of Abowd, Kramarz, and Margolis (1999) report that firm premiums account for about 5-20% of wage inequality.[2,3]

When the firm plays a role in wage inequality, how should we rethink the optimal redistribution policy? This paper asks the following questions. Does the nature of the key trade-offs and optimal property of redistribution policy change? In the presence of redistributive income tax-transfer policy, do we need additional firm regulation?

We answer these questions by studying the optimal redistribution in a monopsonistic labor market model which builds on Manning (2013), Card, Cardoso, Heining, and Kline (2017), and Lamadon, Mogstad, and Setzler (2022). Workers have heterogeneous preferences over the horizontal amenities (such as firm location and job characteristics) of firms, which generate upward-sloping labor supply curves. Differentiated firms have monopsony power and mark down the wages relative to the marginal revenue products of labor. In the benchmark, workers have a single skill type, while firms are heterogeneous in terms of productivity and vertical amenities.[4] Then, identically skilled workers in different firms earn different

[1] Card, Heining, and Kline (2013) also show that between-firm variation explains the larger share of increasing wage inequality in West Germany over time. Similarly, using US tax data, Song, Price, Guvenen, Bloom, and von Wachter (2019) shows that two thirds of the increase in earnings inequality occurred due to a rise in the dispersion between firms.

[2] Abowd, Kramarz, and Margolis (1999) make an identification using movers, whose estimation tends to be upward biased due to the small number of movers. Recent studies correcting this bias show 5–15% firm premium effects (e.g., Bonhomme, Lamadon, and Manresa (2019) and Kline, Saggio, and Sølvsten (2020)).

[3] There is also a related literature estimating the elasticities of wage with respect to exogenous shocks to firm, which are in the range of 0.05–0.15 (Abowd and Lemieux (1993), Reenen (1996), and Reenen (1996)).

[4] Vertical amenities are amenities that are commonly valued by worker such as fringe benefits (e.g. paid leave, health insurance, and pension benefits), and they also include non-pecuniary traits related to working

wages because of (i) pass-through (PT)—higher wage at a more productive firm—and (ii) compensating differentials (CD)—more compensation at a firm with smaller vertical amenity.

Since we want to highlight the trade-offs generated by a redistribution, we consider a benchmark economy where monopsony equilibrium is efficient, abstracting from the motive of correcting the monopsony-driven distortions.[5] Regardless of the competitiveness of the labor market, an introduction of redistribution generates a trade-off between equity and efficiency. The nature of the trade-off, however, depends on the market structure. In a monopsonistic market with differentiated firms, the key efficiency cost of a redistribution is the misallocation of workers between firms. A redistribution policy can improve worker's consumption inequality by reducing the rate of pass-through from firm's productivity to worker's consumption, but this is at the cost of misallocation. With a redistribution, amenities (idiosyncratic preferences) become more important than wages when workers choose firms, thus workers are reallocated from high- to low-productivity firms.

To provide a redistribution at the minimum efficiency cost, how should we design the policy? Our goal is to figure out an optimal property of redistribution that has general policy implications not specific to the tool restriction. Thus, we consider a constrained planner's problem who faces the key trade-offs but is not restricted to the specific policy. The planner maximizes social welfare—weighted sum of welfares of workers and firm owners (entrepreneurs). The planner then has redistribution motives across workers as well as between workers and entrepreneurs. The planner, however, is constrained in the sense that she cannot directly control the assignment of workers across firms, generating the key equity-efficiency trade-off. We also assume that the planner cannot directly tax amenities.[6]

We then highlight the optimal property of redistribution under this key trade-off. We start with the case with no differences in vertical amenities. Since misallocation is more costly in more productive firms, the planner sets a redistribution policy so that PT is less reduced in more productive firms. That is, it is optimal to reduce labor distortion by subsidizing high-productivity firms with relatively higher PT. This feature of optimally tilted PT looks more

conditions and human capital investments.

[5]Incorporating additional motive of correcting inefficiency is straightforward extension; see Section 3.6.

[6]The planner can indirectly tax vertical amenities, as they are correlated with firm productivity (or correlated with wages), while horizontal amenities are uncorrelated.

pronounced when compared to the optimal income tax with constant rate of progressivity, which also lowers the PT but is restricted to the constant rate across firms. Designing a redistribution policy with this optimal property requires a firm regulation.

Next, we incorporate heterogeneous vertical amenities and compensating differentials. Then social welfare depends on three factors: aggregate efficiency, consumption inequality, and effective amenities, and a redistribution policy generates multi-dimensional trade-offs through the two channels—(i) pass-through (PT) of productivity-consumption and (ii) compensating differentials (CD). A redistribution can improve welfare by reducing the consumption inequality with lowered PT and also by increasing the effective amount of vertical amenities with lowered CD—less compensation for lower amenities can increase effective amenity by inducing workers to choose in favor of firms with better amenities.[7] However, both channels are at the cost of distorting allocation of workers across firms. We theoretically show that the degree of misallocation is completely determined by the PT and CD gaps from the efficient levels. Reduction in PT reallocates workers from high- to low-productivity firms, and reduction in CD reallocates workers from low- to high-amenity firms, both implying misallocation of workers to firms with lower marginal products of labor.

Then, what is the desirable property under this multi-dimensional trade-off? Ideally, it would be optimal to increase PT at high-productivity firms and increase CD at low-amenity firms, which subsidize high-productivity and low-amenity firms, as these are firms with high marginal products of labor. With no direct control over vertical amenities, however, the planner has limited ability to adjust PT and CD separately, and it is best to set the policy so that both PT and CD increase in firm productivity. As long as productivity is more important determinant than amenity for the marginal product of labor, a redistribution policy with this desirable property can improve the misallocation at firms with higher marginal products of labor. But its effectiveness depends on the correlation between productivity and vertical amenity—it is effective when the correlation is not too high. When they are highly correlated, the two goals of the planner—subsidizing high productivity and subsidizing low amenities—are more conflicting, limiting the role of refined PT and CD schedules.

[7]Note that redistribution weakens both PT and CD as it reduces the transmission of the dispersion in marginal product of labor to that of wage.

Labor income tax-transfer policy alone cannot implement the constrained optimal allocation because income tax can achieve the tilted PT and CD schedules only by making the labor supply elasticity increase in wage rate, which generates additional efficiency cost. To generate the desirable property of the PT and CD schedules, we need a non-proportional wedge between marginal product of labor and wage, which calls for firm-side policy. Since exact decentralization of optimal allocation requires firm specific policies, we investigate the effectiveness of realistic approximating policies. We show that a linear employment-dependent subsidy to firms can be effective in mitigating the misallocation of the redistribution policy.

We extend our analysis to an economy with multiple types of worker skills to show that the key results from the single-type case still apply with some modifications. To capture the realism, we assume that the planner can only use skill-independent policy by introducing the no-discrimination constraint (NDC) that the markdown of differently skilled workers within a same firm should be the same. Then the constrained planner has an additional redistribution motive between skill types, but without a direct tool, the markdown schedule should be set to indirectly provide this redistribution. Moreover, with the productivity-skill complementarity, the redistribution against the dispersion in marginal revenue products across firms does not necessarily imply wage compression, as the average marginal revenue products, the average between the skill types within a firm, does not always increase in firm productivity. These additional considerations of the planner generate additional forces which makes the optimal schedules of markdown nonmonotone, but the key trade-offs of a redistribution do not change and the optimal property through PT and CD still holds.

We perform a numerical analysis for three purposes: (i) to show the quantitative importance of analytical results, (ii) to decompose the welfare gains from a redistribution policy into each of the three factors (efficiency, consumption inequality, and effective amenities), and (iii) to examine how well a realistic policy can approximate the optimal allocation. The constrained planner can do better than the optimal progressive labor income tax through the effective provision of vertical amenities at a low efficiency cost. When firm productivity and vertical amenity are highly correlated, however, the welfare gain is significantly weakened as the planner's ability to exploit the refined PT and CD schedules becomes limited. According to our quantitative analysis, a labor income tax schedule with a constant rate of

progressivity can achieve almost 99% of welfare gains when the correlation is close to one.

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 (e.g. Heathcote, Storesletten, and Violante (2017)). Recently, there has been a growing number of empirical evidences that firms have some degree of wage-setting power (see the discussion in Card (2022)).[8] This paper departs from the competitive labor market and studies optimal redistribution policy under labor market monopsony. We build on a labor market model developed by Card, Cardoso, Heining, and Kline (2017) and Lamadon, Mogstad, and Setzler (2022), where wage-setting power arises from workers' idiosyncratic preferences across workplaces and investigate an optimal redistribution policy in this environment.

Despite the development in the positive literature, there are relatively few normative studies that analyze optimal policies when firms have wage-setting power, and our paper contributes to the normative literature.[9] Golosov, Maziero, and Menzio (2013) study the optimal redistribution of residual wage inequality in a directed-search model where firms are heterogeneous with respect to productivity and wage-setting power is generated by search frictions.[10] Following Berger, Herkenhoff, and Mongey (2022) and Jarosch, Nimczik, and Sorkin (2024), several papers study optimal policies in an oligopsony environment: Berger, Herkenhoff, and Mongey (2025) analyze the optimal minimum wage, Berger, Hasenzagl, Herkenhoff, Mongey, and Posner (2025) study the optimal merger policy, and Gottfries and Jarosch (2023) investigate the ban on noncompete agreements. In contrast, our paper mainly considers an environment in which such strategic interactions among a small number of agents are absent. With a relatively simple environment, we analyze the constrained planner's allocation and

[8]Following the approach pioneered by Abowd, Kramarz, and Margolis (1999) and extended by Song, Price, Guvenen, Bloom, and von Wachter (2019), Bonhomme, Holzheu, Lamadon, Manresa, Mogstad, and Setzler (2022), and Kline, Saggio, and Sølvsten (2020), studies using regressions with both worker and firm fixed effects find that firm pay premiums account for $5 - 20\%$ of the overall wage dispersion.

[9]Our paper is also related to a literature studying optimal policies in discrete choice environments, such as in Ales and Sleet (2022) and Mongey and Waugh (2024).

[10]In Golosov, Maziero, and Menzio (2013), redistribution does not take place between identically skilled workers who search for jobs offering different wages, as higher wages compensate for lower job-finding probabilities which makes workers indifferent. Different from their analysis, our planner engages in redistribution even between identically skilled workers.

its decentralization rather than committing to a particular policy tool.

The study that is most closely related to ours is Berger, Herkenhoff, and Oppenheimer (2025). They do a Ramsey-style normative analysis in an economy with labor market monopsony, restricting their attention to the income tax function which is the same as in Heathcote, Storesletten, and Violante (2017). Their main channel is that raising the progressivity of labor income tax lowers the elasticity of labor supply, as workers become less sensitive to pre-tax wage variation across firms. We analyze the constrained planner's problem without functional form assumptions on the policy tools. In particular, we show that the labor income tax alone cannot implement the constrained planner's allocation and that we can improve welfare by using firm-side policies that do not distort elasticities of labor supply. Another important difference is that we allow for multiple sources of inequality—productivity pass-through and compensating differentials—and show that the underlying inequality mechanism is central for the size of welfare gains that can be obtained with additional firm regulations.

Layout The remainder of the paper is organized as follows. In Section 2, we introduce our baseline model. In Section 3, we study the optimal redistribution policy with a single skill type and investigate this numerically in Section 4. We extend to the multiple types of skills in Section 5 and do a quantitative analysis in Section 6. Section 7 provides various extensions. Section 8 concludes.

2 Simple Model—Labor Market Monopsony

In this section, we present the simple environment with a single skill type of workers and define the equilibrium under a monopsonistic labor market without redistribution policy.[11] The economy consists of J firm owners and N workers. Each firm j pays wages w_j and exogenously provides vertical amenities g_j. The utility of a worker working in a firm j is

$$v(\varepsilon, j) = \log w_j + \log g_j + \varepsilon_j \tag{1}$$

where ε_j is the idiosyncratic preference (or horizontal amenities) of a worker that is unobserved by the firm. The worker's $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_J) \in \mathbb{R}^J$ is independently drawn from a

[11] The description of the model follows closely Lamadon, Mogstad, and Setzler (2022).

Gumbel distribution with scale $1/\beta$.[12] For workers, the values of working at a given firm differ due to idiosyncratic preferences. Thus, a firm does not lose all of its employees when it lowers the wage, which gives rise to the firm's wage-setting power.

A worker with preference shock ε decides where to work to maximize his utility, taking the distribution of wages and vertical amenities $\{w_j, g_j\}$ as given:

$$j(\varepsilon) = \text{argmax}_{k \in \{1,\ldots,J\}} \quad v(\varepsilon, k). \tag{2}$$

Workers' optimal decisions give rise to an upward-sloping labor supply curve. Due to the Logit structure, the number of workers who work in the firm j, L_j, is derived as

$$L_j = N \cdot \mathbb{P}[j = \text{argmax}_k \;\; v(\varepsilon, k)] = N \cdot \frac{(w_j g_j)^\beta}{I},$$

where $\mathbb{P}$ is the probability and $I = \sum_k (w_k g_k)^\beta$ is the aggregate benefit. Firms paying higher wages and providing better vertical amenities employ more workers. Note that horizontal amenities (ϵ) generate upward-sloping labor supply curves, while vertical amenities (g_j) can shift labor supply curves. The elasticity of labor supply with respect to wages is β, which is the inverse of the scale for the distribution of idiosyncratic preferences.[13]

The firm j with productivity z_j produces the output y_j using technology $y_j = z_j L_j^\eta$, where η is the returns to scale. The firm j facing the labor supply curve $L_j(w_j)$ solves the following profit maximization problem.[14]

$$\pi_j = \max_{w_j, L_j} \qquad z_j L_j^\eta - w_j L_j \qquad s.t. \qquad L_j = L_j(w_j) = N \tfrac{(w_j g_j)^\beta}{I}. \tag{3}$$

Then the firm j sets the wages as markdown from the workers' marginal revenue products:

$$w_j = \frac{\beta}{1+\beta} \cdot MR_j = \frac{\beta}{1+\beta} \cdot \eta z_j L_j^{\eta-1}. \tag{4}$$

Importantly, equation (4) shows that identically skilled workers working in different firms earn different wages. Firm productivity spills over to workers' wages as more productive

[12] We denote the joint distribution of the idiosyncratic preference shocks as $F(\varepsilon)$ and its density as $f(\varepsilon)$.

[13] Note that β is the elasticity of firm-level labor supply, and at lower values of β workers respond less to changes in wages as workers' idiosyncratic preferences across firms are more dispersed.

[14] We assume that there is no strategic interaction, i.e., firms ignore the impact of their wage choice on the wage index I.

firms set higher wages to employ and produce more. We also remark that firms that provide lower vertical amenities compensate workers by paying higher wages. Firms with lower vertical amenities employ less workers, which implies higher marginal products and wages since marginal products decrease in labor ($\eta < 1$).

We assume that firm's vertical amenity g_j and productivity z_j are positively correlated, while ϵ_j is uncorrelated with z_j. This leads to a positive correlation between wage and vertical amenity. Later, when we set the redistribution policy, we can indirectly tax vertical amenities by exploiting this correlation even if we do not have a direct tool to tax amenities.

We formally define the monopsony equilibrium without a redistribution policy.

Definition 1. *A monopsony equilibrium is defined as a set of worker decisions $j(\varepsilon)$, wage indices $\{I\}$, labor demand $\{L_j\}$, labor supply curves $\{L_j(w)\}$ and wages $\{w_j\}$ such that*

1. *Each worker takes the distribution of wages and amenities $\{w_j, g_j\}$ as given and decides where to work $j(\varepsilon)$ to maximize utility as in* (2).
2. *Firms facing upward-sloping labor supply curves $\{L_j(w)\}$ set wages $\{w_j\}$ and choose labor demand $\{L_{sj}\}$ to maximize profits as in* (3).
3. *Wage indices $\{I\}$ satisfy $I = \sum_j (w_j g_j)^\beta$, and labor markets clear: $L_j = L_j(w)\ \forall j$.*

3 Key Trade-offs and Optimal Redistribution

In this section, we study the key trade-offs of a redistribution policy and policy implications in a monopsonistic labor market. The proofs can be found in Online Appendix A.

3.1 Equity-Efficiency Trade-offs of a Redistribution Policy

Regardless of the competitiveness of the labor market, introducing a redistribution policy generates equity-efficiency trade-offs. The nature of trade-offs, however, depends on the market structure. We investigate the trade-offs of a redistribution policy in a monopsonistic labor market to show that the key efficiency cost of a redistribution is misallocation.

We begin with a social welfare function and a welfare decomposition. The welfare function is the weighted sum of the utilities of workers and firm owners (entrepreneurs). We denote the indirect utility of a worker with a preference draw ε by $V(\varepsilon; \{c_j, g_j\}) = \max_j \log c_j + \log g_j + \varepsilon_j$ given the distribution of consumption and amenities across firms. The function V takes into account the optimal choice of firms by workers. Then the social welfare function is given by

$$SWF = N \cdot \int_{\varepsilon} V(\varepsilon; \{c_j, g_j\}) dF(\epsilon) + J \cdot G\left(\bar{U}_f\right), \quad \text{where} \quad \bar{U}_f = \frac{1}{J} \sum_j U(c_{fj}),$$

c_{fj} is the consumption of the firm j's entrepreneur, U is the entrepreneur's utility function, and a function G captures the social preferences on the entrepreneurs. Note that SWF reflects the redistributive motives across workers as well as between workers and firm owners.

We can rewrite the worker's welfare in the social welfare function as

$$W = N \times \int_{\epsilon} \left[\sum_j \mathbb{1}_{\{j(\epsilon)=j\}} \times (\log c_j + \log g_j + \epsilon_j) \right] dF(\epsilon), \tag{5}$$

and the following proposition decomposes the sources of welfare gain and loss in this economy.

Proposition 2. *The worker's welfare (5) can be decomposed into*

$$\begin{aligned}
\frac{W}{N} &= \log C + \sum_j \frac{L_j}{N} \log\left(\frac{c_j}{C}\right) + \widetilde{W}, \quad \textit{where} \\
\widetilde{W} &= \widetilde{W}_g + \widetilde{W}_\epsilon, \quad \textit{where} \\
\widetilde{W}_g &= \sum_j \frac{L_j}{N} \log g_j, \quad \textit{and} \\
\widetilde{W}_\epsilon &= \int_{\epsilon} \left[\sum_j \mathbb{1}_{\{j(\epsilon)=j\}} \times \epsilon \right] dF(\epsilon).
\end{aligned}$$

where $C = \frac{1}{N} \sum_j c_j L_j$ *is the average consumption of workers.*[15]

Proposition 2 shows that there are three sources of welfare: (i) aggregate efficiency, (ii) consumption inequality, and (iii) effective amenities. The first two sources are standard, but the third component $\widetilde{W}$ is a new source that arises due to differentiated firms. $\widetilde{W}$ can be further decomposed into terms associated with vertical amenity $\widetilde{W}_g$ and horizontal amenity $\widetilde{W}_\epsilon$ respectively. The vertical amenities provided by firms are given exogenously, but the

[15] In Online Appendix A, we further decompose $\widetilde{W}_\epsilon$ to see its sources.

effective amount of vertical amenities actually enjoyed by the workers is determined by the worker's decision of where to work, which in turn depends on the planner's redistribution.[16]

An introduction of any redistribution policy generates trade-offs through the three sources in Proposition 2. For now, we focus only on the equity-efficiency trade-off assuming that there is no difference in vertical amenities across firms: $g_j = g, \forall j$, and later we will extend the analysis to multi-dimensional trade-offs with heterogeneous g_j.

Consider a general redistribution policy that reduces consumption inequality. In this simple economy with single skill type of workers, worker's consumption inequality is completely driven by the firm's productivity dispersion. Thus, any redistribution policy should reduce the rate of pass-through from firm's productivity to worker's consumption to improve the inequality, and this results in misallocation of workers.

Proposition 3. *We denote the probability of a worker working for a firm j by P_j. We then define the misallocation of labor between firm j and firm k under a redistribution policy r by*

$$D^r_{j,k} \equiv \log\left(\frac{P^r_j}{P^r_k}\right) - \log\left(\frac{P^e_j}{P^e_k}\right). \tag{6}$$

Assume $g_j = g, \forall j$. The distortion of labor $D^r_{j,k}$ is determined as follows:

$$D^r_{j,k} = \beta \times \int_{z_k}^{z_j} \left(PT^r(z) - PT^e(z)\right) \frac{1}{z} dz, \quad \textit{where} \tag{7}$$

$PT^r(z) = \frac{d\log c^r(z)}{d\log z}$ and $PT^e(z) = \frac{d\log c^e(z)}{d\log z}$ are the pass-through rates under the policy r and in equilibrium, respectively. Under a redistribution r, $PT^r(z) \leq PT^e(z), \forall z$ and $D^r_{j,k} < 0$.

Proposition 3 shows that any redistribution policy that improves worker's inequality leads to misallocation of workers compared to efficient allocation in equilibrium, which is the nature of the equity-efficiency trade-off in the monopsonistic labor market. Redistribution makes workers value amenities—idiosyncratic preferences—more relative to wages when they choose firms, reallocating workers from high- to low-productivity firms.

Before proceeding to optimal policy analysis, let us discuss why we need a redistribution policy in the first place when the equilibrium is efficient and inequality is generated by the

[16]Similarly, the distribution of workers' idiosyncratic preferences across firms is fixed($F(\epsilon)$), but the effective horizontal amenities enjoyed by the workers depend on the allocation of workers to firms.

taste of workers. First of all, as in the classic optimal taxation literature (with competitive labor market), even without market failures, the public cares about inequality. The social welfare function captures this redistribution motive. Second, we remark that the key trade-offs of a redistribution is not specific to this simple setup, where we abstract other sources of inequality for a sharp analysis. The result applies to extensions with an additional motive of redistribution (e.g. redistribution against workers' skill dispersion), as we will show later.

3.2 Constrained Planner's Problem

We have shown that any redistribution policy generates the equity-misallocation trade-offs. Then what is the best way to provide redistribution? If one takes a specific form of policy as in the Ramsey approach, the answer will be limited, as the policy restricts the optimal balance of the trade-offs. Since we are interested in optimal property of a redistribution that can be generally applied, we consider a constrained planner's problem who faces the key trade-offs of a redistribution without specific restriction on the policy. This allows us to highlight the optimal property of redistribution policy to minimize the efficiency cost.

Our planner is constrained in the sense that given the planner's choice of consumption allocation $\{c_j\}$, each worker chooses where to work, and the planner should espect the optimal workplace decision of workers. Thus, the constrained planner faces the same upward-sloping labor supply curve as a firm does, which generates the key equity-efficiency trade-off. The planner also does not have a direct tool for taxing amenities, and she can only indirectly affect the effective amount of amenities by the consumption redistribution.[17] We define our social planner problem.

[17]Note that in the following planner's problem, each firm's amenities are exogenously given, and workers can enjoy amenities with no taxes or subsidies on them.

Definition 4. *The social planner's problem with one type of labor is defined as the program*

$$\begin{aligned} max_{\{c_j,L_j\},\{c_{fj}\},I} \quad & N \cdot \int_{\varepsilon} V(\varepsilon;\{c_j,g_j\}) + J \cdot G\left(\frac{1}{J}\sum_j U(c_{fj})\right) \\ s.t. \quad (\lambda) \quad & \sum_j (c_j L_j + c_{fj}) \leq \sum_j z_j L_j^{\eta} \qquad (8) \\ (\nu) \quad & I = \sum_j (c_j g_j)^{\beta} \qquad (9) \\ & L_j = N\frac{(c_j g_j)^{\beta}}{I}, \quad \forall j, \qquad (10) \end{aligned}$$

where (8) *is resource constraint,* (9) *is wage index, and* (10) *is firm j's labor supply curve.*

In this section, we provide the optimal consumption formula by solving the the planner problem in Definition 4. Then, we will characterize the optimal properties of redistribution from the next section to provide more tangible policy implications.[18]

Proposition 5. *The optimal consumption allocation of workers in the firm j, c_j, is given by*

$$c_j = \frac{\beta}{1+\beta} \times MR_j + \Delta, \quad \Delta = \frac{1}{(1+\beta)\lambda} \cdot \left(1 - \beta\nu\frac{I}{N}\right) \qquad (11)$$

where $MR_j = \eta z_j L_j^{\eta-1}$ is the marginal product of labor in firm j, and λ and ν are the Lagrange multipliers on (8) *and* (9) *respectively. The Lagrange multipliers satisfy $\frac{1}{\lambda} = \frac{\sum_j c_j L_j}{N} = G'(\bar{U}_f) \cdot U'(c_{fj})$ and $\nu\frac{I}{N} = \eta\frac{\sum_j z_j L_j^{\eta}}{\sum_j c_j L_j} - 1$.*

If we assume that $G(\bar{U}_f) = \psi \cdot \log\left(\bar{U}_f\right)$ and $U(c_{fj}) = c_{fj}$, we can further characterize the optimal allocation as follows.

[18]The trade-off of our constrained planner can be better understood by comparing our planner's problem to an alternative First-Best planner's problem. For simplicity, we assume $g_j = 1, \forall j$ and $\eta = 1$. The First-Best planner solves the following problem (see Mongey and Waugh (2024) for details):

$$\begin{aligned} max_{\{j(\epsilon)\},\{c_j(\varepsilon),c_{fj}\}} \quad & N\int_{\varepsilon}\left(\sum_j \mathbb{1}_{\{j(\epsilon)=j\}}(\log c_j(\varepsilon) + \varepsilon_j)\right) dF(\varepsilon) + G(\bar{U}_f) \\ s.t. \quad (\lambda) \quad & N\int_{\varepsilon}\left(\sum_j \mathbb{1}_{\{j(\epsilon)=j\}} c_j(\varepsilon)\right) dF(\varepsilon) + \sum_j c_{fj} \leq \sum_j z_j\left(N\int_{\varepsilon}\mathbb{1}_{\{j(\epsilon)=j\}} dF(\varepsilon)\right) \end{aligned}$$

The First-Best planner does not face the upward-sloping labor supply curve and can therefore choose the allocation of workers across firms $\{j(\epsilon)\}$ independently of the consumption allocation. Then, the First-Best planner provides complete redistribution and allocates workers only considering the efficiency of production and effective horizontal amenities—the optimal $j(\varepsilon)$ is $j(\varepsilon) = j$ if $\lambda z_j + \varepsilon_j \geq \max_k[\lambda z_k + \varepsilon_k]$.

(i) The optimal aggregate income ratio between workers and entrepreneurs is determined by the Pareto weight ψ: $\frac{\frac{1}{J}\sum_j c_{fj}}{\frac{1}{N}\sum_j c_j L_j} = \psi$.

(ii) The common transfer term Δ is positive if $\psi < \psi_e$, where $\psi_e = \frac{\frac{1}{J}\sum_j c^e_{fj}}{\frac{1}{N}\sum_j c^e_j L^e_j}$ is the equilibrium aggregate income ratio. If $\psi = \psi_e$, $\Delta = 0$ and the optimal allocation is equivalent to the equilibrium allocation.

Equation (11) shows that optimal consumption for workers working in the firm j is the sum of a fraction of the marginal product of labor $\frac{\beta}{1+\beta}MR_j$ and the common transfer term Δ. The size of Δ measures the degrees of redistribution for both (i) from entrepreneurs to workers and (ii) across workers at different firms.[19] Throughout the paper, we assume that Δ is positive, which is satisfied if the $G(\cdot)$-function does not put too much welfare weight on the entrepreneurs. With the special social welfare function $G(\bar{U}_f) = \psi \cdot \log\left(\bar{U}_f\right)$, the first motive for redistribution is determined by the size of the Pareto weight on entrepreneurs ψ—Lower ψ implies more redistribution toward workers. The second redistribution motive is achieved as the consumption inequality of workers working in different firms is reduced by the common transfer term.[20]

We also define the optimal markdown of a firm j by

$$\mu_j \equiv \frac{c_j}{MR_j}, \tag{12}$$

which is the ratio of wage relative to the marginal revenue product of labor.[21] A lower μ_j implies that the wage is marked down more widely relative to MR_j. The planner achieves wage compression by setting a wider markdown—lower μ_j—at more productive firms.

Proposition 6. *The optimal markdown is given by*

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

[19] The Lagrange multipliers λ and $\nu\frac{I}{N}$ are shadow prices of additional resources and extra worker, respectively. The optimal Δ decreases in λ and $\nu\frac{I}{N}$, because the efficiency cost of additional redistribution is larger when the shadow prices are higher.

[20] Under this special welfare function, redistribution across workers is provided using the transfer ($\Delta > 0$) from entrepreneurs to workers only when $\psi < \psi_e$. If $\psi = \psi_e$, the equilibrium (with $\Delta = 0$) is optimal.

[21] In decentralization without income tax, μ_j is also equal to $\frac{w_j}{MR_j}$ since $c_j = w_j$.

The optimal markdown schedule is downward-sloping with respect to firm productivity:

$$\left.\frac{d\log\mu_j}{d\log z_j}\right|_{g_j} = -\frac{\frac{\Delta}{MR_j}}{\frac{\beta}{1+\beta}+\frac{\Delta}{MR_j}} \times \frac{d\log MR_j}{d\log z_j} < 0.$$

3.3 Properties of Optimal Redistribution without g-heterogeneity

We proceed further to highlight the key properties of redistribution policy that can optimally balance the trade-offs. We do this step by step. In this section, we consider the case with no differences in vertical amenity across firms, which allows us to focus on how to optimally balance the equity-efficiency trade-off (abstracting from the mechanism through the effective amenity). In Section 3.4, we extend the analysis to allow heterogeneity in vertical amenity.

Assume $g_j = g$ for all j. In Proposition 3, we have shown that any redistribution policy causes misallocation through the pass-through rates. Thus, when designing the redistribution policy, we need to be careful about its impact on the pass-through rate schedule. The next proposition characterizes the optimal property of the PT-schedule by comparing the schedules in equilibrium and under the two redistribution policies—the constrained planner and the HSV income tax with the functional form $T(w) = w - \lambda \cdot w^{1-\tau}$ (Heathcote, Storesletten, and Violante (2017), hereafter HSV).

Proposition 7. *The schedules of pass-through rates $\frac{d\log c_j}{d\log z_j}$ are characterized as follows.*

1. *In equilibrium without redistribution policy, $\quad PT_j^e = \frac{1}{1+(1-\eta)\cdot\beta}, \quad \forall j.$*
   Under the HSV income tax and constrained planner, the firm j sets

$$PT_j^\tau = \frac{1-\tau}{1+(1-\tau)(1-\eta)\beta}, \quad \textit{and} \quad PT_j^o = \frac{\Phi_j^o}{1+\Phi_j^o(1-\eta)\beta},$$

   where $\Phi_j^o \equiv \frac{\frac{\beta}{1+\beta}MR_j^o}{\frac{\beta}{1+\beta}MR_j^o+\Delta} < 1$, showing that redistribution reduces the pass-through rates: $PT_j^e > PT_j^o, \quad \textit{and} \quad PT_j^e > PT_j^\tau$.

2. *Under the constrained planner, the pass-through rate PT_j^o increases in z_j.*

3. *$PT_j^o \geq (<) PT_j^\tau, \quad \textit{for} \quad z_j \geq (<)\bar{z}$, where $\bar{z} = \frac{1+\beta}{\beta\eta}\left(\frac{N}{I}\right)^{1-\eta}\left(\frac{\Delta}{\tau}\right)^{\beta(1-\tau)+1}(1-\tau)g^{\beta(1-\eta)}$.*

Under both redistribution policies, productivity pass-through rates are lower than those in equilibrium, which improves inequality at the cost of misallocation. More important, Proposition 7 highlights how the constrained planner optimally balances the equity-efficiency trade-off. The planner sets the PT-schedule that increases in firm productivity so that the degrees of reduction in PT caused by the redistribution are smaller at more productive firms where the efficiency cost of such reduction is high. This optimal property is more pronounced when it is compared to the schedule under the HSV income tax. At more (less) productive firms, the PT rates of the constrained planner are higher (lower) than those of the HSV income tax. We explicitly show how this tilted PT-schedule mitigates the distortion.

Proposition 8. *Assume $g_j = g$ for all j. Consider two firms with $z_j > z_k$. The distortion $D^{\tau}_{j,k}$ defined in (6) satisfies the following.*

1. *Under the HSV income tax ($c = \phi w^{1-\tau}$, $0 < \tau < 1$), $MR^{\tau}_j > MR^{\tau}_k$, and*

$$D^{\tau}_{j,k} = -\frac{\tau\beta}{1+\beta(1-\eta)} \cdot \log\left(\frac{MR^{\tau}_j}{MR^{\tau}_k}\right) < 0, \tag{13}$$

   implying a misallocation from the firm j to the firm k (under-allocation to j).

2. *In the constrained planner's optimal allocation,*

$$\begin{aligned} D^{o}_{j,k} &= D^{\tau}_{j,k} + D^{gap}_{j,k}, \quad \textit{where} \\ D^{gap}_{j,k} &= \beta \times \int_{z_k}^{z_j} \left(PT^{o}(z) - PT^{\tau}(z)\right) \frac{1}{z} dz. \end{aligned} \tag{14}$$

   If $z_k \geq \bar{z}$ ($\bar{z}$ is defined in Proposition 7), then $D^{gap}_{j,k} \geq 0$ and $D^{\tau}_{j,k} \leq D^{o}_{j,k} < 0$, implying that the misallocation from firm j to k is mitigated by the constrained planner.

Proposition 8 considers a case where firm j has higher productivity than firm k. A redistributive HSV income tax causes a misallocation ($D^{\tau}_{j,k} < 0$) from a less productive firm k to a more productive firm j (Proposition 8.1). The tilted PT-schedule of the constrained planner mitigates this misallocation ($D^{\tau}_{j,k} \leq D^{o}_{j,k} < 0$) for firms with high productivity above the cut-off $\bar{z}$ (Proposition 8.2). In this region with high marginal products of labor, the misallocation is especially costly, and we can minimize the cost by designing a redistribution so that PT is raised in this region (relative to PT under the HSV income tax).

We have shown that we can lower the efficiency cost of redistribution by designing a policy that generates a PT-schedule increasing in firm productivity. Then, what kind of policy do we need to implement this optimal property? We investigate possible implementation.

Proposition 9. *An optimal allocation can be implemented with various policies by combining labor income tax and firm regulations.*

1. *The labor income tax alone cannot implement the optimal allocation.*

2. *The following policies can implement optimal allocation.*

   (a) ***Revenue-dependent subsidy and progressive income tax:***
   Given an income tax with the form $T(w_j) = w_j - \phi \cdot w_j^{1-\tau}$*, a sum-to-zero subsidy schedule* $\{S_j(y)\}$ *to the firm can implement the constrained optimal allocation if it satisfies* $\phi \cdot \left(1 + S_j'(y_j)\right)^{1-\tau} \times \left(\frac{(1-\tau)\beta}{1+(1-\tau)\beta}\right) \times MR_j^{-\tau} = \frac{\beta}{1+\beta} + \frac{\Delta}{MR_j}$*. If* $\frac{\tau}{1-\tau} < \frac{1+\beta}{\beta} \cdot \frac{\Delta}{MR_j}$ $\forall j$*,* $S_j'(y_j)$ *decreases in firm productivity* z_j*.*

   (b) ***Employment-dependent subsidy to the firm only:***
   A sum-to-zero subsidy schedule $\tilde{S}(L)$ *that is common to all firms can implement the constrained optimal allocation if it satisfies* $\tilde{S}'(L_j) = \frac{1+\beta}{\beta} \cdot \Delta$*.*

Given that the planner provides redistribution against workers' inequality, one might conjecture that optimal allocation could be implemented using a redistributive income tax only. Proposition 9 shows that this is not possible. The fundamental difficulty is that the labor income tax can affect the PT-schedule only through its impact on the labor supply elasticity, but the change in the labor supply elasticity generates additional efficiency cost. In Section 3.5, We revisit this limitation and compare the effectiveness of various realistic policies.

To generate the tilted PT-schedule, we need a wedge between the marginal product of labor and wage, which calls for a firm regulation. Proposition 9 considers a subsidy to the revenue or to the hiring cost.[22] Note that the first decentralization encompasses infinite combinations

[22]With a revenue-dependent subsidy $\{S_j(y)\}$ which sums to zero—although named subsidies, are actually taxes for some firms, the firm j solves the following problem given $S_j(y)$:

$$max_{y_j, w_j, L_j} \quad y_j + S_j(y_j) - w_j L_j \qquad\qquad s.t. \quad y_j = z_j L_j^{\eta}, \quad \text{and} \quad L_j = N \frac{(w_j g_j)^{\beta}}{I}.$$

of the income tax (with various levels of τ) and (firm-specific) revenue-dependent subsidy. The marginal subsidy rate decreases in firm productivity, but decreases less in the region with high marginal products of labor. Alternatively, linear employment-dependent subsidies common to all firms can implement the optimal allocation. However, when a progressive income tax already exists, a common schedule is not enough for decentralization. Since firm-specific subsidies are not realistic, we compare the effectiveness of realistic (approximating) policies in Section 3.5 and investigate them numerically in Section 4.

3.4 Properties of Optimal Redistribution with g-heterogeneity

Now, we are back to the economy with heterogeneous vertical amenities. Recall that there are three sources of welfare—efficiency, consumption inequality, and effective amenity. When firms are also different in their vertical amenities, a redistribution can improve welfare through two channels: (i) reducing the consumption inequality and (ii) increasing the effective amount of vertical amenities. Any redistribution policy reduces the transmission of the dispersion in the marginal product of labor to the dispersion in consumption. This lowers not only pass-through (PT) rates but also compensating differentials (CD)—the consumption compensation for lower level of amenities—with heterogeneous g.[23]

Then, we need to consider multi-dimensional trade-offs when designing a redistribution policy. A decrease in PT still reduces the consumption inequality at the cost of misallocation. In addition, a reduction in CD increases the effective vertical amenity by making workers value amenities more relative to wages, also reallocating workers across firms. The next proposition revisits the labor distortion in Proposition 3 by rewriting $D^r_{j,k}$.

With this subsidy, the optimal wage choice of firm j is $w_j = (1 + S'_j(y_j)) \times \frac{\beta}{1+\beta} MR_j$. Alternatively, with an employment-dependent subsidy $\tilde{S}(L)$, the firm j solves the following problem given $\tilde{S}(L)$:

$$max_{y_j, w_j, L_j} \quad y_j + \tilde{S}(L_j) - w_j L_j \qquad\qquad s.t. \quad y_j = z_j L_j^{\eta}, \quad \text{and} \quad L_j = N \frac{(w_j g_j)^{\beta}}{I}.$$

Then the optimal wage choice of the firm j is $w_j = \frac{\beta}{1+\beta} MR_j + \frac{\beta}{1+\beta} \tilde{S}'(L_j)$.

[23]The mechanism of compensating differentials is that a lower amenity of a firm decreases labor supply, leading to higher marginal product of labor, thus higher wage (consumption) of workers at that firm. Thus, a weaker link between the marginal product of labor and consumption reduces the compensating differentials.

Proposition 10. *In the presence of heterogeneous vertical amenities, the labor distortion under policy r, $D^r_{j,k}$, is determined by both pass-through and compensating-differential rates:*

$$\begin{aligned} D^r_{j,k} &= D^r_1 + D^r_2, \quad \textit{where} \\ D^r_1 &= \beta \times \int_{z_k}^{z_j} \left(PT^r(z, g_j) - PT^e(z, g_j)\right) \frac{1}{z} dz, \quad \textit{and} \\ D^r_2 &= -\beta \times \int_{g_k}^{g_j} \left(CD^r(z_k, g) - CD^e(z_k, g)\right) \frac{1}{g} dg, \end{aligned} \tag{15}$$

where $PT^p(z,g) = \frac{d\log c^p(z,g)}{d\log z}\Big|_g$ *and* $CD^p(z,g) = \frac{d\log c^p(z,g)}{d\log g}\Big|_z$, $p \in \{e, r\}$.

Proposition 10 shows that with heterogeneous g, a redistribution policy causes misallocation through changes in both pass-through and compensating-differentials. Thus, to provide redistribution at a lower efficiency cost, we need to design a policy considering its impact on the PT and CD together. We now characterize the optimal properties of PT and CD.

Proposition 11. *With heterogeneous $\{g_j\}$, the equations for the pass-through rates are the as as in Proposition 7. The compensating differentials ($-\frac{d\log c_j}{d\log g_j}\Big|_{z_j}$) are characterized as follows.*

1. *In equilibrium without redistribution policy, $CD^e_j = \frac{(1-\eta)\beta}{1+(1-\eta)\beta} \quad \forall j$.*
   Under the HSV income tax and the constrained planner, firm j sets

$$CD^\tau_j = \frac{(1-\tau)(1-\eta)\beta}{1+(1-\tau)(1-\eta)\beta} \quad \forall j, \quad \textit{and} \quad CD^o_j = \frac{\mathbf{\Phi}^o_j(1-\eta)\beta}{1+\mathbf{\Phi}^o_j(1-\eta)\beta},$$

   where $\mathbf{\Phi}^o_j \equiv \frac{\frac{\beta}{1+\beta}MR^o_j}{\frac{\beta}{1+\beta}MR^o_j+\Delta} < 1$, showing that redistribution reduces both PT and CD: $PT^e_j > PT^r_j$, and $CD^e_j > CD^r_j$, $\quad r \in \{\tau, o\}$.

2. *Assume $z_j > z_k$, then $D^r_1 < 0$, $r \in \{\tau, o\}$. On the other hand, $D^r_2 \leq (>) 0$ if $g_k \geq (<) g_j$.*

3. *Under the constrained planner, both $PT^o(z,g)$ and $CD^o(z,g)$ increase in $MR^o(z,g)$, implying that they increase in z given g, while they decrease in g given z.*

4. *Given g, both $PT^o(z,g)$ and $CD^o(z,g)$ cross $PT^\tau(z,g)$ and $CD^\tau(z,g)$ from below: $PT^o(z,g) \geq (\leq) PT^\tau(z,g)$ and $CD^o(z,g) \geq (\leq) CD^\tau(z,g)$ if $z \geq (\leq) \bar{z}(g)$, where $\bar{z}(g) = \frac{1+\beta}{\beta\eta}\left(\frac{N}{I}\right)^{1-\eta}\left(\frac{\Delta}{\tau}\right)^{\beta(1-\tau)+1}(1-\tau)g^{\beta(1-\eta)}$ is g-dependent cut-off productivity.*

Any redistribution reduces PT and CD, improving the consumption inequality and increasing effective amenity (Proposition 11.1). Reduction in PT reallocates workers from high- to low-productivity(z) firms given g, and reduction in CD reallocates workers from low- to high-amenity(g) firms given z, implying misallocation to firms with lower marginal products of labor in each margin (lower z_j or higher g_j implies lower MR_j). But with heterogeneity in both z and g, whether additional reallocation through the CD mitigates or worsens misallocation through the PT depends on the correlation between firm's productivity and vertical amenity. Proposition 11.2 shows that under a positive correlation, reallocation from low-g to high-g implies reallocation from low-z to high-z, mitigating the misallocation through PT.

Given the multi-dimensional trade-offs, how can we provide redistribution at a low efficiency cost? Proposition 11.3 and 11.4 characterize the desirable properties of PT and CD by showing how the constrained planner sets these schedules to minimize the cost of misallocation when providing redistribution. The planner sets the PT and CD schedules that increase in z given g and decrease in g given z, mitigating misallocation at firms with higher marginal products of labor where misallocation is especially costly. At productive firms above the (g-dependent) cut-off productivity $\bar{z}(g)$, the planner's PT and CD are higher than those under the HSV income tax (which are constant for all firms).

Note that this is the best that a constrained planner can do. Ideally, to minimize the efficiency cost of redistribution, it would be optimal to increase PT at high-productivity firms (subsidizing high-z firms) and increase CD at low-amenity firms (subsidizing low-g firms). However, the planner without a direct tool to control vertical amenities can only achieve this partially, as she cannot separately control the PT and CD. For example, when firms with high productivity also have high vertical amenities, subsidizing high-z firms (relative to low-z firms) and subsidizing low-g firms (relative to high-g firms) are not compatible. The next proposition shows that at high-productivity firms, the planner's PT-schedule improves the misallocation, but the CD-schedule does not necessarily mitigate the misallocation.

Proposition 12. *Consider two firms that have productivity-amenity pairs* (z_j, g_j) *and* (z_k, g_k). *Assume that*

$$z_j > z_k, \quad \textit{and} \quad \frac{z_j}{z_k} > \left(\frac{g_j}{g_k}\right)^{(1-\eta)\beta}. \tag{16}$$

Under the HSV income tax, $D^{\tau}_{j,k} < 0$ as in (13). The labor distortion decomposition under the constrained planner (14) also applies, but $D^{gap}_{j,k}$ should now satisfy

$$\begin{aligned} D^{gap}_{j,k} &= D^{gap}_1 + D^{gap}_2, \quad \textit{where} \\ D^{gap}_1 &= \beta \times \int_{z_k}^{z_j} \left(PT^o(z, g_j) - PT^{\tau}(z, g_j)\right) \frac{1}{z} dz, \\ D^{gap}_2 &= -\beta \times \int_{g_k}^{g_j} \left(CD^o(z_k, g) - CD^{\tau}(z_k, g)\right) \frac{1}{g} dg, \end{aligned} \tag{17}$$

We denote the cut-off productivity associated with $\bar{g} = \max\{g_k, g_j\}$ by $\bar{z}(\bar{g}) \equiv \frac{1+\beta}{\beta\eta} \left(\frac{N}{I}\right)^{1-\eta} \left(\frac{\Delta}{\tau}\right)^{\beta(1-\tau)+1} (1-\tau)\bar{g}^{\beta(1-\eta)}$. If $z_k \geq \bar{z}(\bar{g})$, then $D^{gap}_1 \geq 0$ (implying mitigated distortion), and the sign of D^{gap}_2 depends on the size of g_j and g_k:

(i) $D^{gap}_2 \geq 0$ if $g_k \geq g_j$ (mitigated distortion), and

(ii) $D^{gap}_2 < 0$ if $g_j > g_k$ (stronger distortion).

Under the condition of (16), the firm j has a higher marginal product of labor than the firm k. The negative distortion $D^{\tau}_{j,k}$ implies that a redistributive HSV income tax causes a misallocation of labor from the firm j to k. Proposition 12 shows how the planner improves this distortion. For firms with productivity above the cut-off $\bar{z}(\bar{g})$, a higher PT with the tilted PT-schedule mitigates the misallocation, but a higher CD with the tilted CD-schedule does not necessarily improve the misallocation, showing the planner's limited ability. Higher CD mitigates distortion only when g is negatively correlated with z ($z_j > z_k$ and $g_j < g_k$).

When z and g are positively correlated ($z_j > z_k$ and $g_j > g_k$), a higher CD above the cutoff $\bar{z}(\bar{g})$ reduces the relative consumption of workers in the higher-amenity firm j, which induces workers to move from j to k. But the firm j has a higher-productivity with a positive $corr(z, g)$. This implies that the increase in CD by the planner generates additional distortion ($D^{gap}_2 < 0$) counterbalancing the good effects of a higher PT ($D^{gap}_1 > 0$) above the cutoff $\bar{z}(\bar{g})$. As long as productivity is more important in determining the firm's marginal product of labor—assumed by (16), the channel through PT dominates the channel through CD, and it is best to set PT and CD schedules that increase in firm productivity (z).

Role of $corr(z, g)$ We summarize the role of $corr(z, g)$. First, any redistribution policy reduces CD, but its trade-off depends on the correlation. With a higher correlation, reduced

CD can better mitigate misallocation through PT (Proposition 11.2), which increases the optimal amount of redistribution. However, a higher correlation makes worker's workplace decision respond less to a reduction in CD, leading to a smaller increase in the effective amenity. Second, how much the planner can do better with the refined PT and CD schedules also depends on the correlation. With a higher correlation, the planner's adjustments of tilting PT and CD are less powerful in both mitigating misallocation and increasing effective amount of amenity. In Section 4, we quantitatively investigate the role of $corr(z, g)$.

3.5 Policy Implications—Comparison of Approximating Policies

Since exact decentralization of the constrained planner's allocation requires unrealistic firm-specific policies (Proposition 9), we consider realistic (approximating) policies and investigate their effectiveness. As we have pointed out, an important property of the redistribution policy is its impact on the PT and CD schedules. The effectiveness of policies depend on their ability to generate the optimal property of the PT and CD.

We rewrite the PT and CD to see how policies can affect them in general:

$$PT_j \equiv \frac{d\log c_j}{d\log z_j} = \frac{d\log c_j}{d\log w_j} \times \frac{d\log w_j}{d\log MR_j} \times \frac{d\log MR_j}{d\log z_j}, \quad \text{and} \tag{18}$$

$$CD_j \equiv \frac{d\log c_j}{d\log g_j} = \frac{d\log c_j}{d\log w_j} \times \frac{d\log w_j}{d\log MR_j} \times \left(-\frac{d\log MR_j}{d\log g_j}\right), \tag{19}$$

which clearly shows the possible channels of the policies. Note that both $\frac{d\log MR_j}{d\log z_j}$ and $\frac{d\log MR_j}{d\log g_j}$ are determined by technology (without direct control over the amenities). Thus, policies can affect PT and CD through $\frac{d\log c_j}{d\log w_j}$ or through $\frac{d\log w_j}{d\log MR_j}$. First, firm regulations can do the job by generating a non-proportional wedge between the wage (w_j) and the marginal revenue product of labor (MR_j), which can be done by taxing/subsidizing either revenue or labor cost. Second, redistributive income taxes have a direct impact on $\frac{d\log c_j}{d\log w_j}$ and an indirect impact on $\frac{d\log w_j}{d\log MR_j}$ through the change in the elasticity of labor supply. The next proposition investigates the effectiveness of these policies in the PT and CD schedules.

Proposition 13. *Under each policy r, PT_j^r and CD_j^r are expressed in the following forms:*

$$PT_j^r = \frac{\frac{\partial \log c_j^r}{\partial \log z_j}}{1+\beta(1-\eta)\cdot\left(-\frac{d\log c_j^r}{d\log L_j}\right)} \quad \textit{and} \quad CD_j^r = \frac{\beta\cdot\left(-\frac{d\log c_j^r}{d\log L_j}\right)}{1+\beta\cdot\left(-\frac{d\log c_j^r}{d\log L_j}\right)}. \tag{20}$$

In the following two cases, we consider firm regulations given the existing HSV income tax

1. *Employment-dependent subsidy $\tilde{S}(L_j)$ (given HSV income tax)*

   Consider a subsidy that is linear in employment: $\tilde{S}(L_j) = \tilde{s}\cdot L_j$. Under $\tilde{S}(L_j)$,

   $$\frac{\partial \log c_j^{\tilde{S}}}{\partial \log z_j} = (1-\tau)\frac{MR_j}{MR_j+\tilde{s}}, \quad -\frac{d\log c_j^{\tilde{S}}}{d\log L_j} = (1-\tau)\frac{(1-\eta)MR_j}{MR_j+\tilde{s}}, \textit{ implying the following.}$$

   (a) *Both PT_j and CD_j decrease in $\tilde{s}$.*

   (b) *If $\tilde{s}>0$, both PT_j and CD_j increase in MR_j.*

2. *Revenue-dependent subsidy $S(y_j)$ (given HSV income tax)*

   Consider a revenue-dependent subsidy schedule $S(y_j)$ for firms. Under $S(y_j)$,

   $$\frac{\partial \log c_j^S}{\partial \log z_j} = (1-\tau)\cdot(1-r(y_j)), \quad \textit{and} \quad -\frac{d\log c_j^S}{d\log L_j} = (1-\tau)\cdot(1-\eta+\eta\cdot r(y_j)), \quad \textit{where}$$

   $r(y) = -\frac{S''(y)y}{1+S'(y)}$ measures the rate of progressivity of $S(y)$, which implies the following

   (a) *The level of the marginal subsidy rate $S'(y_j)$ does not have an impact on PT_j and CD_j. $S(y_j)$ affects PT and CD only through progressivity $r(y)$.*

   (b) *Higher $r(y_j)$ results in lower PT_j but higher CD_j.*

3. *Finally, we consider a case with labor income tax $T(w_j)$ only. Under $T(w_j)$,*

   $$\frac{\partial \log c_j^T}{\partial \log z_j} = -\frac{1}{1-\eta}\cdot\frac{d\log c_j^T}{d\log L_j} = \rho(w_j)\times\frac{1}{1-\frac{1}{1+\tilde{\beta}(w_j)}\cdot\frac{d\log\tilde{\beta}(w_j)}{d\log w_j}}, \quad \textit{where}$$

   $\rho(w_j) = \frac{1-T'(w_j)}{1-T(w_j)/w_j}$ and $\tilde{\beta}(w_j) = \beta\times\rho(w_j)$, which implies the following.

   (a) *Lower progressivity through higher $\rho(w_j)$ affects PT and CD through two channels.*

   i. Direct: higher $\rho(w_j)$ increases PT_j and CD_j controlling for $\tilde{\beta}(w_j)$.

   ii. Indirect: higher $\tilde{\beta}(w_j)$ decreases PT_j and CD_j controlling for $\rho(w_j)$ and $\frac{d\log\tilde{\beta}}{d\log w_j}$.

(b) If $T(w_j)$ satisfies $\rho'(w_j) > 0$, then

i. PT_j and CD_j increase in w_j controlling for $\tilde{\beta}(w_j)$, and

ii. $\tilde{\beta}(w_j)$ counterbalances (i) controlling for $\frac{d \log \tilde{\rho}(w_j)}{d \log w_j}$.

As we discussed above, a firm regulation which generates a non-proportional wedge between wage and marginal product of labor can generate the schedules of PT and CD which increase in firm productivity. Proposition 13 shows that a linear employment-dependent subsidy can do the job, while we need a nonlinear revenue-dependent subsidy to tilt either schedule. In fact, a progressive revenue-dependent subsidy is not effective, as an increase in progressivity moves the PT and CD in opposite directions.

Proposition 13 also shows the role of income tax. Linear tax has no impact on the PT and CD, and constant progressivity implies constant schedules of them. An income tax schedule whose progressivity decreases in income can provide direct forces of making PT and CD increase in firm productivity, but the same tax schedule also provides the opposite force through has indirect effect—causes the elasticity of labor supply to increase in wage—which makes the income tax less effective in lowering the efficiency cost of redistribution. The effectiveness of this additional firm regulation depends on the sources of inequality ($corr(z, g)$), which is quantitatively investigated in Section 4.

3.6 Extensions to Monopsony Distortion Correction

So far, to focus on the trade-offs generated by a redistribution policy, our analysis has been carried out in an economy where there is no distortion in a monopsony equilibrium. In this section, we show that the key results extend to an economy with monopsony distortion (both misallocation and aggregate distortion). see Online Appendix B for a detailed derivation.

3.6.1 Monopsony-Distortion 1: Misallocation

First, we consider an extension where the dispersion of monopsony power generates misallocation in equilibrium. We assume that each firm j belongs to a market $m(j)$ and denote the set

of firms in the market m by J_m. Worker's ϵ is drawn from a distribution which has a nested-logit structure: $F(\epsilon) = \exp\left(-\sum_m \left(\sum_{j\in J_m} \exp\left(-\frac{\epsilon}{\rho_m}\right)\right)^{\rho_m}\right)$, where $\rho_m = \sqrt{1 - corr(\epsilon_j, \epsilon_{j'})}$ captures the degree of independence within the market if $m(j) = m(j') = m$.

Then, the labor supply curve of a firm j is $L_j = N \times P_m \times P_{j|m} = N \times \frac{I_m^{\rho_m}}{\sum_{\tilde{m}} I_{\tilde{m}}^{\rho_{\tilde{m}}}} \times \frac{w_j^{\beta/\rho_m} \cdot g_j^{\beta/\rho_m}}{I_m}$ where $I_m = \sum_{k\in J_m} w_k^{\beta/\rho_m} \cdot g_k^{\beta/\rho_m}$, implying that the elasticity of labor supply is β/ρ_m, and the markdown in equilibrium is $\mu_j^e \equiv \frac{w_j^e}{MR_j^e} = \frac{\beta/\rho_m}{1+\beta/\rho_m}$. Assuming a variation in ρ_m across markets, heterogeneous elasticities generate a dispersion in markdown μ_j^e, leading to misallocation between markets in an equilibrium. The planner then has an additional goal of improving the misallocation driven by the ρ_m-dispersion when designing a redistribution policy.

In this extended economy, there can be two types of misallocation: (i) between firms within a market and (ii) across markets. The next proposition characterizes misallocations in equilibrium and under the constrained planner.

Proposition 14. *We denote the probability of a worker working for a firm j and a market m in a competitive labor market—where $w_j^* = MR_j^*$—by P_j^* and P_m^*, respectively. Under a policy p (including an equilibrium with no policy), we define the misallocation between firms j and k within a market m by*

$$D_{j,k}^p \equiv \log\left(\frac{P_j^p}{P_k^p}\right) - \log\left(\frac{P_j^*}{P_k^*}\right), \quad for \quad m(j) = m(k), \tag{21}$$

and we define the misallocation between markets m and m' by

$$D_{m,m'}^p \equiv \log\left(\frac{P_m^p}{P_{m'}^p}\right) - \log\left(\frac{P_m^*}{P_{m'}^*}\right). \tag{22}$$

These labor distortions are characterized as follows.

1. *Within a market, an equilibrium allocation is efficient with $D_{j,k}^e = 0$, while the constrained planner's redistribution generates misallocation with $D_{j,k}^o \neq 0$. All the results in Propositions 10, 11, and 12 apply with the only change in the elasticity term—β is replaced by β/ρ_m in all equations.*

2. *To characterize the misallocation between markets, we further assume that $\eta = 1$ ($y_j = z_j \cdot L_j$) and consider two markets with $\rho_m > \rho_{m'}$.*

(a) In equilibrium, misallocation between markets arises with labor distortion

$$D^e_{m,m'} = \beta \times \left(\log\left(\frac{\beta/\rho_m}{1+\beta/\rho_m} \right) - \log\left(\frac{\beta/\rho_{m'}}{1+\beta/\rho_{m'}} \right) \right) < 0,$$

which implies under-allocation to less correlated market with higher ρ_m.

(b) Under the constrained planner, the consumption of workers at firm j satisfies the formula: $c^o_j = \frac{\beta/\rho_m}{1+\beta/\rho_m} \cdot MR^o_j + \Delta_r$. The planner can then mitigate the misallocation of the equilibrium by setting $\Delta_m > \Delta_{m'}$ so that $D^e_{m,m'} < D^o_{m,m'} < 0$

Proposition 14 shows that in the extension with nested-logit, the redistribution within a market has exactly the same trade-offs and optimal properties of a redistribution policy as in the single-nest case, implying that the key results are robust to the economy with monopsony distortion. When the monopsony power itself generates misallocation, however, the planner has an additional goal—the policy should also correct the misallocation across markets.

3.6.2 Monopsony-Distortion 2: Aggregate Distortion

In a monopsonistic labor market, aggregate markdown can generate distortion if there is endogenous unemployment or intensive labor supply decision. Here, we consider an extension with an intensive labor supply. For simplicity, we assume that workers have preferences with no income effects in labor supply:

$$u(c,h) = \log\left(c - \frac{1}{1+\frac{1}{\epsilon}} h^{1+\frac{1}{\epsilon}} \right), \tag{23}$$

where h is working hours. A worker solves the two-stage problem: (i) choose where to work and (ii) decides working hour. Solving backwards, after choosing firm j, the worker solves

$$V(w_j) = \max_{c,h} \log\left(c - \frac{1}{1+\frac{1}{\epsilon}} h^{1+\frac{1}{\epsilon}} \right) + \log g_j \quad \text{s.t.} \quad c = w_j \cdot h,$$

and the worker chooses a firm in the first stage: $V = E\left[\max_j V(w_j) + \epsilon_{ij} + \log(g_j)\right]$, where ϵ_{ij} is again drawn from the Gumbel distribution.

With the special preferences, the optimal working hour has a closed form solution, $h(w_j) = w_j^\epsilon$, and the firm j's labor supply curve is derived as $L_j = N_j \times h_j = N \times \frac{w_j^{\beta(1+\epsilon)} \cdot g_j^\beta}{I} \times w_j^\epsilon$, where

$I = \sum_j w_j^{\beta(1+\epsilon)} \cdot g_j^{\beta}$. Then we denote the total elasticity of labor supply by $\epsilon^t \equiv \frac{d \log L_j}{d \log w_j} = \epsilon^N + \epsilon^h = \beta(1+\epsilon) + \epsilon$, which is constant for all firms. In an equilibrium of this economy, firms with monopsony power set wages with a constant markdown: $w_j^e = \frac{\epsilon^t}{1+\epsilon^t} \cdot MR_j^e$, so there is no misallocation of workers. However, lowered wages due to markdown decrease workers' working hours relative to the efficient working hours in the competitive labor market. Then the planner has an additional goal of improving the distortion driven by the aggregate markdown when designing a redistribution policy.

The next proposition characterizes the labor distortion—misallocation and aggregate distortion—in equilibrium and under the constrained planner. These distortions are measured by comparing the allocation with the efficient allocation in a competitive labor market.

Proposition 15. *Consider an economy with intensive labor supply with worker's preferences (23). We define the misallocation $D_{j,k}^p$ between firms j and k under a policy p as in (21), and we also define the distortion of labor at firm j due to the aggregate markdown by*

$$D_j^{A,p} \equiv \log h_j^p - \log h_j^*, \tag{24}$$

where h_j^p and h_j^* are the working hours of workers in firm j under a policy p and in the competitive market (with $w_j^* = MR_j^*$), respectively. We characterize these labor distortions.

1. ***Distortion due to Misallocation:***

   There is no misallocation of workers in equilibrium ($D_{j,k}^e = 0$), while a redistribution causes misallocation. The characterization of misallocation in Proposition 10 applies.

   (a) *The optimal properties of PT and CD in Proposition 11 apply to modified PT and CD which include additional feedback effects through the working hours.*

   (b) *As long as $\beta(1-\eta) > \frac{\epsilon\eta}{1+\epsilon}$, the planner's role in mitigating misallocation in Proposition 12 applies.*

2. ***Distortion due to Aggregate Markdown:***

   (a) *In equilibrium, working hours of workers are reduced due to the markdown, generating distortion: $D_j^{A,p} = \frac{\epsilon}{1+(1-\eta)\epsilon} \times \log\left(\mu_j^e\right) < 0$, where $\mu_j^e = \frac{\epsilon^t}{1+\epsilon^t} < 1$.*

(b) Under the constrained planner, the distortions in working hours satisfy

$$D_j^{A,o} = D_j^{A,e} + D_j^{A,gap}, \quad \textit{where} \quad D_j^{A,gap} = \epsilon \cdot \log\left(\frac{\mu_j^o \cdot MR_{Lj}^o}{\mu_j^e \cdot MR_{Lj}^e}\right),$$

where $\mu_j^o = \frac{\epsilon^t}{1+\epsilon^t} + (1+\epsilon) \cdot \frac{\Delta}{MR_{Lj}^o h_j^o}$. $D_j^{A,gap} > 0$ *for firms with* $P_j^e > P_j^o$*, which implies that the distortion is mitigated for firms with high productivity.*

Proposition 15 shows that in the extension with intensive labor supply, the planner still faces the same trade-offs of a redistribution as in the benchmark case, but now has an additional consideration of correcting the monopsony distortion. The optimal properties of the PT and CD that increase in firm productivity is maintained to lower the efficiency cost of a redistribution, but the planner also mitigates the distortion in working hours generated by the monopsony power by increasing $\mu_j^o = \frac{w_j^o}{MR_{Lj}^o}$ for high productivity firms.

4 Numerical Analysis

So far, we have analyzed the optimal properties of a redistribution policy through PT and CD schedules and their effectiveness. We now confirm these results numerically and quantitatively decompose the welfare gains from the redistribution policy. We also investigate how much of the optimal welfare gain can be approximated with more realistic policies.

We set the mass of workers and firms as $N = 0.88$ and $J = 0.12$.[24] We set the elasticity of the labor supply $\beta = 4.5$ and return to the scale parameter $\eta = 0.75$. We assume that the joint distribution of firm productivity and amenity is log-normal:

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

We calibrate the parameters of the distribution $\sigma_z = 0.93, \sigma_g = 0.76, \rho_{zg} = 0.95$ to match the variance decomposition of firm pay premiums as in Lamadon, Mogstad, and Setzler (2022).[25]

[24] (N, J) matches the share of workers and entrepreneurs in the population as in Boar and Midrigan (2025).

[25] The firm pay premium φ_j is the sum of the productivity spill-over component φ_z and the compensating differential component. See the Online Appendix E.1 for details.

4.1 Schedules of Pass-through and Compensating Differentials

We consider the utilitarian social welfare function with $\psi = 1$ as a benchmark.[26] Throughout this section, we mainly compare three cases: (1) equilibrium without redistribution policies, (2) constrained planner's allocation, and (3) allocation under the optimal HSV labor income tax.[27] We also consider other approximating policies where the progressivity of HSV labor income tax is fixed at the current US level ($\tau = 0.18$) and the planner has access to either (4) HSV revenue-dependent subsidies to firms, or (5) linear employment-dependent subsidies to firms (see Online Appendix D.5 for exact form of the policies).[28]

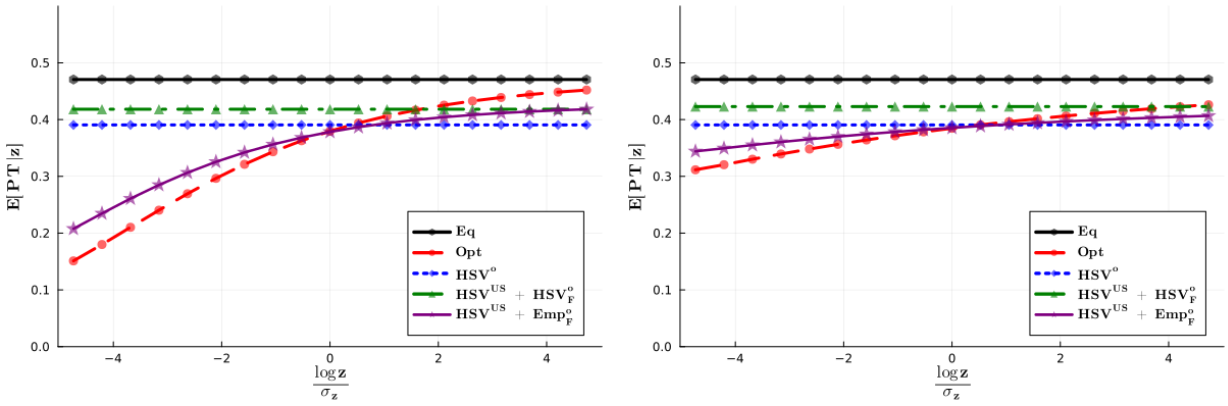
Figure 1 and Figure 2 plot the PT and CD schedules under various policies for low and high ρ_{zg}'s. In equilibrium (1), both PT and CD are constant across firms. All redistributive policies reduce the levels of PT and CD relative to those in equilibrium to improve consumption inequality and increase the effective level of amenities. The constrained planner (2) sets tilted schedules that increase in firm productivity to minimize efficiency costs, while the HSV income tax (3) is restricted to impose constant schedules, which confirms the results of Propositions 7 and 11. On the other hand, the planner with HSV revenue-dependent subsidy to firms (4), also confined to set constant schedules, cannot reduce the PT rate as much as the rate under the HSV income tax, because lowering PT by raising the progressivity of revenue-dependent subsidy inevitably increases CD as shown in Proposition 13. Providing linear employment subsidies to firms (5) best approximates the constrained planner's allocation by both sufficiently reducing the levels and generating the tilted schedules.

The correlation between firm productivity and amenities, ρ_{zg}, determines the extent to which the constrained planner can perform better than the optimal HSV income tax using

[26] Throughout the quantitative analysis, we abstract from the redistribution across entrepreneurs by assuming that $G\left(\frac{1}{J}\sum_j U(c_{fj})\right) = \psi \cdot \log\left(\frac{1}{J}\sum_j c_{fj}\right)$. In reality, firms' financial portfolios are well diversified. See Online appendix D.2 for further discussion. See Online Appendix E.5 for the case with $\psi = 0$.

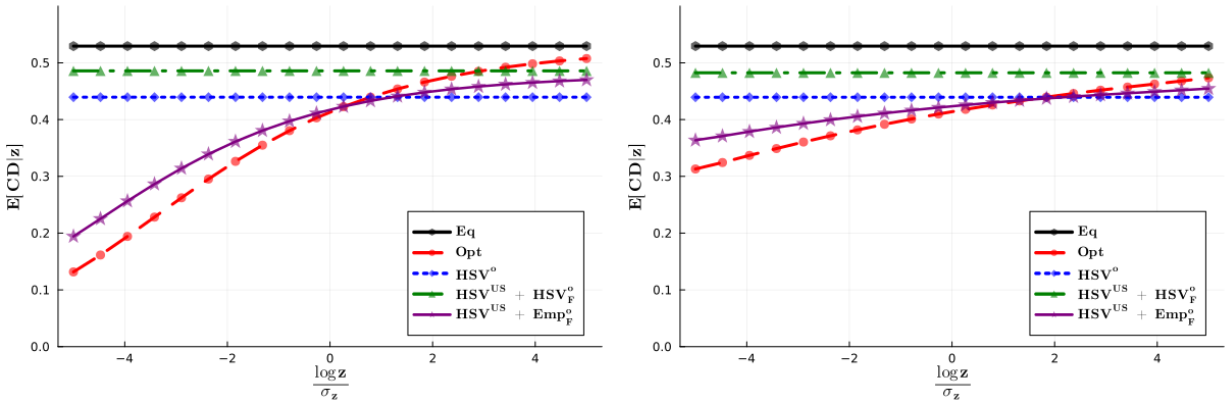
[27] To focus on the redistribution across workers, we allow the planner to adjust the aggregate labor income share in all five cases (by having access to linear revenue-dependent subsidies to firms). Here, the utilitarian planner optimally sets the labor income share to $\frac{N}{N+\psi J} = 0.88$.

[28] We consider such policies because in reality the progressivity of the labor income tax is determined by various other factors (e.g. redistribution between differently-skilled workers), so it may seem implausible to approximate the constrained planner's allocation by varying the progressivity of labor income tax only.



(a) $\text{corr}(\log z, \log g) = 0.25$ (b) $\text{corr}(\log z, \log g) = 0.75$

Figure 1: Productivity pass-through rate



(a) $\text{corr}(\log z, \log g) = 0.25$ (b) $\text{corr}(\log z, \log g) = 0.75$

Figure 2: Degree of compensating differentials

refined PT and CD schedules. As we have shown in Proposition 12, at lower values of ρ_{zg} (but still positive), the two goals of the planner—subsidizing high-productivity firms and subsidizing low-amenity firms—are less conflicting, allowing the planner to set steeper schedules of PT and CD in Figure 1 and 2. This explains why the constrained planner achieves larger welfare gains when ρ_{zg} is lower, which we show through the welfare decomposition below.

4.2 Welfare Gain Decomposition

How much can we improve welfare by optimally balancing the trade-offs of a redistribution policy? We now evaluate welfare gains and decompose them by their sources—efficiency,

ρ_{zg}	Total Gains	Decomposition			
		Efficiency	Redistribution	Amenity	
				Vertical	Horizontal
0.25	7.403	-17.894	2.359	26.853	0.743
	(100%)	(-276.1%)	(32.6%)	(333.1%)	(10.4%)
0.50	4.631	-11.755	2.062	14.259	1.675
	(100%)	(-276.2%)	(45.1%)	(294.5%)	(36.7%)
0.75	2.254	-5.977	1.263	5.421	1.874
	(100%)	(-276.5%)	(56.3%)	(236.8%)	(83.3%)
0.95	0.565	-1.547	0.366	0.045	1.727
	(100%)	(-276.6%)	(64.7%)	(8.0%)	(303.8%)

(a) Welfare Gains from Constrained Planner's Allocation

ρ_{zg}	Total Gains	Decomposition				Gains rel. to CP(%)
		Efficiency	Redistribution	Amenity		
				Vertical	Horizontal	
0.25	6.328	-15.608	4.253	19.183	1.402	85.5
	(100%)	(-276.6%)	(67.8%)	(286.0%)	(22.7%)	
0.50	4.197	-10.751	2.825	11.811	1.548	90.6
	(100%)	(-276.6%)	(67.8%)	(271.5%)	(37.4%)	
0.75	2.148	-5.712	1.449	5.076	1.629	95.3
	(100%)	(-276.7%)	(67.7%)	(233.0%)	(76.1%)	
0.95	0.558	-1.529	0.377	0.088	1.646	98.8
	(100%)	(-276.7%)	(67.7%)	(15.7%)	(293.3%)	

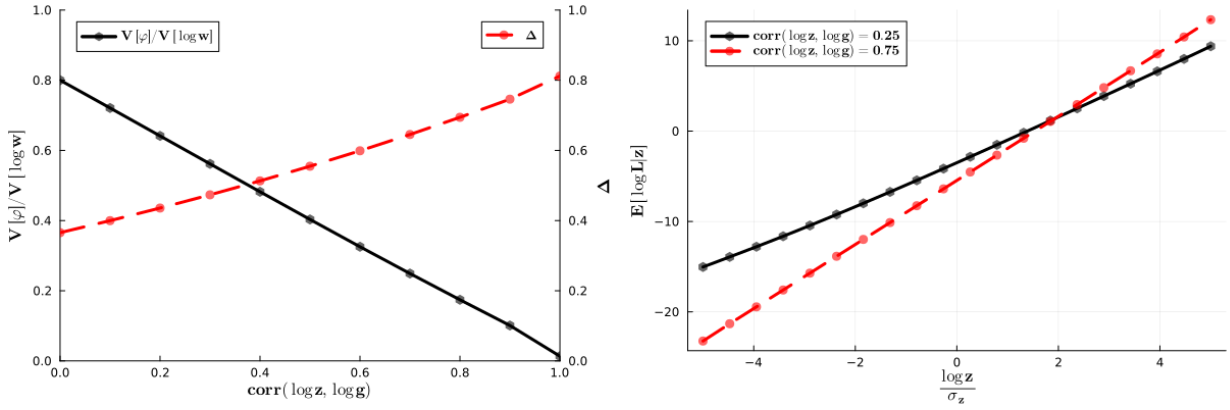
(b) Welfare Gains from the Optimal HSV Labor Income Tax

Table 1: Decomposition of Welfare Gains

Notes: The welfare gains are reported in consumption-equivalent variations which satisfy $1 + CEV = (1 + CEV_e)(1 + CEV_r)(1 + CEV_a)$. Numbers in parentheses refer to the decomposition of welfare change ($dW = dW_e + dW_r + dW_a$). See Online Appendix E.4 for the details.

consumption inequality, and effective amenity—as in Proposition 2. Table 1a and Table 1b report the welfare gains from the constrained planner's allocation relative to the equilibrium and the gains from the optimal HSV labor income tax relative to the equilibrium, respectively. They verify that redistribution achieves welfare gains by improving worker's consumption inequality and increasing effective amenities, at the cost of lowering production efficiency.

The correlation between firm productivity and amenities plays its role in two aspects, as we have discussed in Section 3.4. First, the standard equity-efficiency trade-offs of any redistri-



(a) Firm Premium and Δ (b) $\mathbb{E}[\log L|z]$

Figure 3: Sources of Inequality and the Optimal Policy

bution policy become relaxed as the correlation increases. Figure 3a shows that the amount of redistribution measured by Δ increases in ρ_{zg}, even though the role of firm premium in wage inequality—measured by $\frac{\mathbb{V}[\varphi]}{\mathbb{V}[\log w]}$—decreases in ρ_{zg}. This is because with higher ρ_{zg}, reallocation through the reduced CD mitigates misallocation as shown in Proposition 11.2. Tables 1a, 1b and Figure 3b confirm mitigated inefficiency when ρ_{zg} is higher.[29,30]

Second, and more important, higher correlation limits the planner's ability to balance the multi-dimensional trade-offs. Comparison of Tables 1a and 1b shows that the differences in welfare gains between the constrained planner's allocation and the optimal HSV labor income tax are indeed greater at lower values of ρ_{zg}, demonstrating that the planner's refined PT and CD schedules can work more effectively when the correlation is low. The optimal HSV income tax achieves 85.9% of the constrained planner's welfare gains at $\rho_{zg} = 0.25$ whereas 98.9% at $\rho_{zg} = 0.95$. Recall that compared to the optimal HSV income tax, a lower CD of the planner in the low productivity region increases the effective vertical amenity, while a higher CD in the high productivity region generates inefficient misallocation (Proposition 12). As ρ_{zg} increases, the planner's welfare gain from this refined CD schedule decreases, since a higher ρ_{zg} results in a less increase in effective-g (with a weakened response of workers

[29]Figure 3b shows that when ρ_{zg} is higher, productive firms employ more workers despite greater Δ.

[30]Table 1a also confirms that higher Δ at higher correlation does not lead to greater welfare gain as reduced CD leads to much less increase in effective vertical amenity with higher correlation.

to CD) and an increased inefficiency (with further misallocation from high- to low z).[31]

In Online Appendix E.4 we also present the decomposition of the welfare gain from implementing the approximating policies. The HSV revenue-dependent subsidy achieves a smaller portion of the welfare gains, as lowering PT for wage compression raises the level of CD and is therefore inefficient at increasing vertical amenities. The linear employment-dependent subsidy achieves most of the welfare gains regardless of the values of ρ_{zg}.

5 Extension to Multiple Skill-Types of Workers

We extend the analysis to multiple types of workers with different skills. There are two types of workers, high- and low-skilled, with a total mass $N = N_h + N_\ell$. Without a redistribution policy, each firm j pays skill-specific wages w_{sj} for $s \in \{h, \ell\}$, and g_{sj} is the value that workers of skill s place on vertical amenities g_j provided by the firm j. From the worker's optimal decision, the labor supply curve of s-skill in the firm j is derived as $L_{sj} = N_s \cdot \frac{(w_{sj} g_{sj})^\beta}{I_s}$. The firm j produces output using production technology $y_j = z_j \cdot \tilde{y}_j^\eta$, where $\tilde{y}_j = \left(\alpha_j L_{hj}^{\frac{\sigma-1}{\sigma}} + (1-\alpha_j) L_{\ell j}^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}$, where α_j is the intensity of the skill, η is the returns to scale and σ is the elasticity of substitution. In this section, we assume a CRS production ($\eta = 1$), and allow $\eta < 1$ in the quantative analysis. In equilibrium, the firm j, which maximizes profit, sets the wages as markdown from the workers' marginal revenue products:

$$w_{hj} = \frac{\beta}{1+\beta} \cdot \underbrace{\eta z_j \tilde{y}_j^{\eta-1} \cdot \alpha_j \left(\frac{L_{hj}}{\tilde{y}_j}\right)^{-\frac{1}{\sigma}}}_{MR_{hj}}, \quad w_{\ell j} = \frac{\beta}{1+\beta} \cdot \underbrace{\eta z_j \tilde{y}_j^{\eta-1} \cdot (1-\alpha_j) \left(\frac{L_{\ell j}}{\tilde{y}_j}\right)^{-\frac{1}{\sigma}}}_{MR_{\ell j}}. \tag{25}$$

Note that our model also allows for systematic sorting of worker skills across firms. First, demand-side sorting can arise due to the complementarity between firm productivity and worker skill. High-productivity firms have a higher demand for high-skilled workers ($corr(\alpha_j, z_j) >$

[31] The effect on horizontal amenities can go in either direction. A reduction in PT makes the workers choose firms in favor of amenities, thereby increasing horizontal amenities. A reduction in CD makes the workers value wages more and lowers horizontal amenities. The net effect depends on the value of ρ_{zg}.

0). Second, there can be supply-side sorting if differently skilled workers value amenities differently. If high-productivity firms provide more amenities, high-skilled workers who value the amenities more can sort into more productive firms.

We now analyze optimal policy. The trade-offs of a redistribution in Section 3 are still alive, but in the presence of multiple types of workers, there is an additional motivation for redistribution between different types of skills. In the presence of type-dependent subsidy, the planner can separate this redistribution between skill types from the redistribution across workers at different firms within skill type, thus the properties of optimal allocation are the same as in the economy with only single type (see Online Appendix C for more details).

However, the assumption of a type-specific policy is surely unrealistic. We commonly observe size-dependent policies for firms in many countries, but skill-dependent policiess are rarely used. It is hard to implement policies that condition on the worker skill, since our definition of skill includes not only the level of education but also the general income generating ability.[32] Thus, we assume that the planner can only use type-independent policy.

5.1 No Discrimination Constraint and the Constrained Planner's Problem

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

$$W = N_h \int_\varepsilon V_h(\varepsilon; \{c_{hj}\}) dF(\varepsilon) + N_\ell \int_\varepsilon V_\ell(\varepsilon; \{c_{\ell j}\}) dF(\varepsilon) + \psi \cdot J \cdot \log\left(\frac{1}{J}\sum_j c_{fj}\right).$$

It reflects that with multiple skill types, the planner's motives of redistribution across workers combine both the redistribution across different firms and redistribution between skill types.

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

$$\frac{c_{hj}}{MR_{hj}} = \frac{c_{\ell j}}{MR_{\ell j}}, \quad \forall j = 1, \ldots, J. \tag{26}$$

The constraint requires that the markdown—the ratio of consumption to marginal revenue products—should be the same across differently skilled workers within a firm. This implies

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

that the planner cannot direct all of its resources to a particular type of worker when subsidizing a firm. To increase the consumption of a low-skilled worker, the planner should also increase the consumption of a high-skilled worker working in the same firm.

The form of the no-discrimination constraint (26) is motivated by decentralization with skill-independent policy. Suppose that the government can use revenue-dependent subsidy $S_j(y)$, then the firm j maximizes profit $y_j + S_j(y_j) - (w_{hj}L_{hj} + w_{\ell j}L_{\ell j})$ subject to the labor supply curves. The optimal wage setting of the firm implies that

$$(1 + S_j'(y_j))\frac{\beta}{1+\beta} = \frac{w_{hj}}{MR_{hj}} = \frac{w_{\ell j}}{MR_{\ell j}}, \tag{27}$$

which derives our no-discrimination constraint (26).

Importantly, we want to remark that the specific form of the NDC depends on the policy tools the planner can use in decentralization. Note that we have derived the NDC (26) from the decentralization with subsidy to firms only, implicitly assuming that there is no income tax-transfer policy in the decentralization. In Section 7.1, we extend our analysis to the case with the income tax-transfer policy to show the robustness of our results.

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

Definition 16. *The Constrained planner's problem with no-discrimination constraints is defined as*

$$\max_{\substack{\{c_{hj},c_{\ell j},c_{fj}\},\\ \{L_{hj},L_{\ell j}\},I_h,I_\ell}} \quad N_h \int_\varepsilon V_h(\varepsilon;\{c_{hj},g_{hj}\})dF(\varepsilon) + N_\ell \int_\varepsilon V_\ell(\varepsilon;\{c_{\ell j},g_{\ell j}\})dF(\varepsilon) + \psi \cdot J \cdot \log\left(\frac{1}{J}\sum_j c_{fj}\right)$$

$$s.t. \quad \sum_j (c_{hj}L_{hj} + c_{\ell j}L_{\ell j} + c_{fj}) \leq \sum_j z_j(\alpha_j L_{hj}^{\frac{\sigma-1}{\sigma}} + (1-\alpha_j)L_{\ell j}^{\frac{\sigma-1}{\sigma}})^{\frac{\sigma}{\sigma-1}} \tag{28}$$

$$\frac{c_{hj}}{MR_{hj}} = \frac{c_{\ell j}}{MR_{\ell j}}, \quad \forall j = 1,\ldots,J \tag{29}$$

$$I_h = \sum_j (c_{hj}g_{hj})^\beta, \quad I_\ell = \sum_j (c_{\ell j}g_{\ell j})^\beta \tag{30}$$

$$L_{hj} = N_h\frac{(c_{hj}g_{hj})^\beta}{I_h}, \quad L_{\ell j} = N_\ell\frac{(c_{\ell j}g_{\ell j})^\beta}{I_\ell}, \quad \forall j = 1,\ldots,J. \tag{31}$$

5.2 Optimal Allocation and Markdown

The next proposition characterizes the optimal consumption (see Online Appendix A for the proof).

Proposition 17. *The optimal consumption allocation c_{sj} for each type of workers at firm j is*

$$c_{sj} = \frac{1}{\Omega_{sj}} \times \left(\frac{\beta}{1+\beta} MR_{sj} + \Delta_s \right), \quad s \in \{h, \ell\}, \quad \textit{where} \tag{32}$$

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

and $\Delta_s = \frac{1}{\lambda(1+\beta)} \left(1 - \beta \nu_s \frac{I_s}{N_s}\right)$. *$\lambda$, ζ_j and ν_s are Lagrange multipliers on the resource constraint* (28)*, the no-discrimination constraint* (29)*, and the index equations* (30)*, respectively.*

The optimal consumption formula (32) closely resembles the formula with a single skill type, except for modification by Ω_{sj}'s. When the NDC is not binding, $\zeta_j = 0$ implies $\Omega_{sj} = 1$, which exactly goes back to the result with the type-specific policy. Note that the equality of the NDC (29) 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 multiplier ζ_j is nonnegative. That is, the planner wants to redistribute from the high-skill type to the low-skill type within a firm j, but she cannot do this when the NDC is binding as $\Omega_{hj} < 1$ and $\Omega_{\ell j} > 1$ adjust c_{hj} upward and $c_{\ell j}$ downward.

Then how does the constrained planner achieve redistribution across skill types when she does not have a direct tool? The planner can indirectly provide this redistribution by adjusting the markdown schedule according to the firm's labor share of the skill types. On the other hand, with the multiple skill types and productivity-skill complementarity($\alpha'(z_j) > 0$), the motive of redistribution across workers at different firms does not necessarily imply that the markdown schedule (μ_j-schedule) decreases in firm productivity, which is different from Proposition 6. The planner should consider how the *average* marginal revenue products, the average between the skill types within a firm, varies across firms with different productivities. How do these additional considerations shape the optimal markdown schedule?

We define the (type-independent) optimal markdown in firm j by $\mu_j = \frac{c_{sj}}{MR_{sj}}$, $s \in \{h, \ell\}$. The next proposition characterizes the μ_j-schedule (see Online Appendix A for the proof).

Proposition 18. *Optimal markdown of a firm j is given by*

$$\mu_j = \omega_j \times \left(\frac{\beta}{1+\beta} + \frac{\Delta_h}{MR_{hj}} \right) + (1-\omega_j) \times \left(\frac{\beta}{1+\beta} + \frac{\Delta_\ell}{MR_{\ell j}} \right)$$

where the weight ω_j of high-skilled worker within firm j is

$$\omega_j = \frac{\alpha_j L_{hj}^{\frac{\sigma-1}{\sigma}}}{\alpha_j L_{hj}^{\frac{\sigma-1}{\sigma}} + (1-\alpha_j) L_{\ell j}^{\frac{\sigma-1}{\sigma}}}. \tag{33}$$

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

$$\begin{aligned} \frac{d\log \mu_j}{d\log z_j} = &\quad F1 + F2, \\ \textit{where} \quad F1 = \left. \frac{d\log \mu_j}{d\log z_j} \right|_{\frac{L_{hj}}{L_{\ell j}} \textit{ given}} &\quad \textit{and} \quad F2 = \frac{\partial \log \mu_j}{\partial \log \left(\frac{L_{hj}}{L_{\ell j}} \right)} \times \frac{d\log \left(\frac{L_{hj}}{L_{\ell j}} \right)}{d\log z_j}. \end{aligned} \tag{34}$$

Proposition 18 shows that the optimal markdown is given by the weighted sum of $\frac{\beta}{1+\beta} + \frac{\Delta_h}{MR_{hj}}$ and $\frac{\beta}{1+\beta} + \frac{\Delta_\ell}{MR_{\ell j}}$. It also shows that the slope is shaped by two distinct channels. The $F1$-term captures the planner's motive to redistribute across identically skilled workers working at different firms, while the $F2$-term reflects the motive to redistribute between different skill types. We investigate each term one by one.

Proposition 19. *The term $F1$ in the slope of the optimal markdown $\mu(z_j)$ is given by*

$$\begin{aligned} F1 &= F_{1z} + F_{1\alpha} \qquad (35) \\ \textit{where} \quad F_{1z} &= -\left[\frac{\omega_j \frac{\Delta_h}{MR_{hj}}}{\mu_j} \times \frac{\partial \log MR_{hj}}{\partial \log z_j} + \frac{(1-\omega_j)\frac{\Delta_\ell}{MR_{\ell j}}}{\mu_j} \times \frac{\partial \log MR_{\ell j}}{\partial \log z_j} \right] < 0, \quad \textit{and} \\ F_{1\alpha} &= \begin{cases} -\frac{\sigma}{\sigma-1} \cdot \frac{\alpha_j \cdot (L_{hj}^{\frac{\sigma-1}{\sigma}} - L_{\ell j}^{\frac{\sigma-1}{\sigma}})}{\alpha_j L_{hj}^{\frac{\sigma-1}{\sigma}} + (1-\alpha_j) L_{\ell j}^{\frac{\sigma-1}{\sigma}}} \times \left(\frac{\omega_j \frac{\Delta_h}{MR_{hj}}}{\mu_j} + \frac{(1-\omega_j)\frac{\Delta_\ell}{MR_{\ell j}}}{\mu_j} \right) \times \frac{d\log \alpha_j}{d\log z_j}, & \sigma \neq 1 \\ -\alpha_j \cdot \log \left(\frac{L_{hj}}{L_{\ell j}} \right) \times \left(\frac{\omega_j \frac{\Delta_h}{MR_{hj}}}{\mu_j} + \frac{(1-\omega_j)\frac{\Delta_\ell}{MR_{\ell j}}}{\mu_j} \right) \times \frac{d\log \alpha_j}{d\log z_j}, & \sigma = 1 \end{cases} \end{aligned}$$

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:

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

Proposition 19 shows that the term $F1$—redistribution across workers at different firms—can be further decomposed into two channels (see Online Appendix A for the proof). The first term, 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 a single skill type. Focusing on the case where $\Delta_s > 0$ for all s, the sign of F_{1z} is negative. The planner closes the wage gap of workers within a skill type by compressing the wage distribution across firms, which provides a force for a wider markdown of wage—lower μ_j—for large firms with high productivities.

On the other hand, the second term, $F_{1\alpha}$, captures the effect of increasing z_j on the markdown μ_j *through the change in skill intensity* $\alpha(z_j)$ (keeping the ratio of employment $\frac{L_{hj}}{L_{\ell j}}$ constant). Unlike F_{1z}, $F_{1\alpha}$ can be 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 when the skill intensity increases in firm productivity ($\alpha'(z_j) > 0$), increased firm productivity can lead to a higher or lower average marginal revenue product of the firm, depending on the share of employment. An increase in skill intensity reduces the average marginal product of labor of a firm in the region where low-skilled workers have a higher share of employment $L_{hj} < L_{\ell j}$, while it increases the average marginal product in the region with $L_{hj} > L_{\ell j}$. The 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.

We next move to the investigation of the second term $F2$.

Proposition 20. *The term $F2$ in the slope of the optimal markdown $\mu(z_j)$ is given by*

$$F2 = \frac{\omega_j(1-\omega_j)}{\mu_j} \cdot \left(\frac{\Delta_h}{MR_{hj}} - \frac{\Delta_\ell}{MR_{\ell j}} \right) \cdot \left[\frac{\partial \log\left(\frac{L_{hj}}{L_{\ell j}}\right)}{\partial \log \alpha_j} \cdot \frac{d\log \alpha_j}{d \log z_j} + \frac{\partial \log\left(\frac{L_{hj}}{L_{\ell j}}\right)}{\partial \log\left(\frac{g_{hj}}{g_{\ell j}}\right)} \cdot \frac{d \log\left(\frac{g_{hj}}{g_{\ell j}}\right)}{d \log z_j} \right], \quad (36)$$

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

If the no-discrimination constraint is not binding ($\zeta_j = 0$), 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:

$$sign(F2) = -sign\left(\frac{\partial \log\left(\frac{L_{hj}}{L_{\ell j}}\right)}{\partial \log \alpha_j} \cdot \frac{d \log \alpha_j}{d \log z_j} + \frac{\partial \log\left(\frac{L_{hj}}{L_{\ell j}}\right)}{\partial \log\left(\frac{g_{hj}}{g_{\ell j}}\right)} \cdot \frac{d \log\left(\frac{g_{hj}}{g_{\ell j}}\right)}{d \log z_j}\right).$$

Proposition 20 shows that when the NDC is not binding, the term $F2$ does not appear, which implies that the term $F2$ appears to provide redistribution between different types of workers. If $\zeta_j > 0$, the planner wants to redistribute from high- to low-skilled workers but cannot directly achieve this due to NDC. The planner indirectly provides this redistribution 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 a lower subsidy (or wider markdown of wage with lower μ_j) to such high-productivity firms, generating forces for a negative slope of the $\mu(z_j)$—negative $F2$. However, if the labor share of each type of worker is constant across firms, $F2 = 0$, because despite redistributive concerns between workers of different types, it is impossible to close the wage gap by subsidizing a particular firm.

In sum, negative terms F_{1z} and $F2$ make the $\mu(z_j)$-schedule decrease in firm productivity, which provides 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 6.

5.3 Optimal Balance of the Trade-offs with Multiple Skill Types

We show that when there are multiple skill types, the planner still uses the refined PT and CD schedules to reduce the efficiency cost of redistribution, but this does not necessarily imply that PT and CD are increasing in firm productivity.

We denote the pass-through rate of the skill type s by $PT_{sj} = \frac{d \log c_{sj}}{d \log z_j}\Big|_{g_j}$. The next proposition characterizes the pass-through rates.

Proposition 21. *PT_{sj} can be decomposed into a common term and an α-related term:*

$$PT^r_{sj} = PT^{c,r}_j + PT^{\alpha,r}_{sj}, \quad s \in \{h, \ell\}, \quad r \in \{e, o\}.$$

1. In equilibrium without any redistribution policy,

$$\begin{aligned} PT_j^{c,e} &= 1, \\ PT_{sj}^{\alpha,e} &= \frac{\sigma}{\sigma+\beta}\left(p_{sj}^{\alpha} + \frac{\omega_j^e - \alpha_j}{(\sigma-1)(1-\alpha_j)}(1+\beta)\right)\cdot\frac{d\log\alpha_j}{d\log z_j}, \quad s\in\{h,\ell\} \\ \textit{where} \quad & p_{hj}^{\alpha}=1, \quad p_{\ell j}^{\alpha} = -\frac{\alpha_j}{1+\alpha_j}, \textit{ and } \omega_j^e \textit{ is (33) evaluated at equilibrium.} \end{aligned}$$

2. Under the constrained planner,

$$\begin{aligned} PT_j^{c,o} &= \frac{\beta}{1+\beta}\cdot\frac{1}{\mu_j^o} < 1, \quad \textit{and denoting the weight (33) by } \omega_j^o, \\ PT_{sj}^{\alpha,o} &= \frac{\sigma}{\sigma+\beta}\left(p_{sj}^{\alpha} + \frac{\omega_j^o - \alpha_j}{(\sigma-1)(1-\alpha_j)}(1+\beta)\right)\cdot\frac{d\log\alpha_j}{d\log z_j} \quad + \quad \Lambda_{\alpha,j}^o, \quad s\in\{h,\ell\}, \\ \textit{where} \quad & \Lambda_{\alpha,j}^o = \left(\frac{\sigma\beta}{\sigma+\beta}\cdot\frac{1}{1-\alpha_j}\cdot\frac{\partial\log\mu_j^o}{\partial\log\left(\frac{L_{hj}}{L_{\ell j}}\right)} + \frac{\partial\log\mu_j^o}{\partial\log\alpha_j}\right)\cdot\frac{d\log\alpha_j}{d\log z_j}. \end{aligned}$$

This implies that

(i) Suppose that there is no complementarity between firm's productivity and worker's skill. Then $PT_{sj}^r = PT_j^{c,r}$, and the pass-through rate under the constrained planner is smaller than that in equilibrium. The planner's PT schedule increases in firm productivity.

(ii) Suppose that there is complementarity between firm's productivity and worker's skill. The gap of PT's under the constrained planner and the equilibrium, $PT_{sj}^o - PT_{sj}^e$, can be non-monotone in firm productivity due to (a) non-monotone $\frac{1}{\mu_j}$-schedule in $PT_j^{c,o}$ and (b) an additional non-monotone term $\Lambda_{\alpha,j}^o$ in $PT_{sj}^{\alpha,o}$.

Proposition 21 shows that, as in the case of a single type of workers, a redistribution policy lowers the pass-through rate ($PT_j^{c,o} < PT_j^{c,e}$) through the weaker link between the marginal product of labor and consumption, which generates the efficiency cost of redistribution. To reduce this cost, the planner still wants to increase the PT rate for firms with high marginal products of labor, but in the presence of complementarity between firm productivity and worker skill, this does not necessarily imply that the PT schedule increases in firm productivity. As we have discussed when explaining the $F_{1\alpha}$-term, the firm's average marginal product of labor (average between the skill types) can decrease in firm productivity in the

region where low-skilled workers are employed more, because higher skill intensity lowers the marginal product of low-skilled workers. As a result, μ_j-schedule is non-monotone in firm productivity, which also makes $PT_j^{c,o}$ non-monotone. In the region where μ_j decreases in firm productivity, $PT_j^{c,o}$ still increases in firm productivity as in the single type case.

Proposition 21 also shows that with multiple types of skills, there is an additional channel in the pass-through rate through the complementarity between the firm productivity and the worker's skill. The gap of PT's under the planner and the equilibrium through this new channel is determined mainly by $\Lambda^o_{\alpha,j}$ which captures the change in consumption caused by the change in markdown through the complementarity ($F_{1\alpha}$-term). Thus, the PT schedule also inherits the shape of $F_{1\alpha}$, which is another force of making the PT schedule non-monotone. We investigate the PT schedules numerically in Section 6.

Next, we denote the (negative) compensating differentials of the skill type s by $\widetilde{CD}_{sj} \equiv \frac{d\log c_{sj}}{d\log g_j}$, and the absolute value of $\widetilde{CD}_{sj}$ is denoted by $\overline{CD}_{sj} \equiv \left|\frac{d\log c_{sj}}{d\log g_j}\right|$. The next proposition characterizes the compensating differentials with multiple types of skills.

Proposition 22. *The compensating differential schedules are characterized as follows.*

1. *Suppose that $g_{hj} = g_{\ell j} = g_j$. Then, $\overline{CD}^e_{sj} = \overline{CD}^o_{sj} = 0, \quad s \in \{h, \ell\}$.*

2. *Suppose that $\frac{d\log g_{hj}}{d\log g_j} > \frac{d\log g_{\ell j}}{d\log g_j}$. The CD schedules are then characterized as follows.*

   (a) *In equilibrium, $\widetilde{CD}^e_{sj} = \tilde{\gamma}^e_{sj} \cdot \frac{d\log\left(g_{hj}/g_{\ell j}\right)}{d\log g_j}$, where $\tilde{\gamma}^e_{hj} = -\frac{\beta(1-\omega^e_j)}{\sigma+\beta}$ and $\tilde{\gamma}^e_{\ell j} = \frac{\beta\omega^e_j}{\sigma+\beta}$. This implies $\widetilde{CD}^e_{hj} < 0 < \widetilde{CD}^e_{\ell j}$.*

   (b) *Under the constrained planner, $\widetilde{CD}^o_{sj} = \tilde{\gamma}^o_{sj} \cdot \frac{d\log\left(g_{hj}/g_{\ell j}\right)}{d\log g_j} + \tilde{\Phi}^o_j$, $s \in \{h, \ell\}$, where $\tilde{\Phi}^o_j \leq 0$ and becomes strictly negative when the NDC binds.*

      - *The additional term $\tilde{\Phi}^o_j$ reinforces the degree of negative $\widetilde{CD}$ for $s = h$.*
      - *The additional term $\tilde{\Phi}^o_j$ mitigates the degree of positive $\widetilde{CD}$ for $s = \ell$.*

Note that under CRS technology, the standard CD channel discussed above that works through the level of labor supply does not exist, as the marginal product of labor does not depend on the level of labor. When $g_{hj} = g_{\ell j} = g_j$, the CD channel that works through

the share of employment of differently skilled workers does not also exist, as the ratio $\frac{L_{hj}}{L_{\ell j}}$ is independent of g_j. Thus, with CRS technology and no-supply side sorting, the planner does not use the CD schedule to balance the trade-offs. In Section 6, we numerically investigate the case with multiple types of skills under DRS technology and no supply-side sorting to show that $\overline{CD}$ increases in firm productivity as in the case of single-skill type.

On the other hand, if we allow supply-side sorting with skill-dependent valuation of amenity, a new channel can arise through the change in ratio $\frac{L_{hj}}{L_{\ell j}}$, which results in $\widetilde{CD}^e_{hj} < 0 < \widetilde{CD}^e_{\ell j}$.[33] A redistribution by the planner reinforces the degree of negative $\widetilde{CD}$ for the high-skilled worker, while mitigating the degree of positive $\widetilde{CD}$ for the low-skilled worker. This is because with the NDC, the transmission from the marginal product of labor to consumption is strengthened for the high-skilled, while it is weakened for the low-skilled.

6 Quantitative Analysis

We now numerically investigate the optimal markdown and the balance of the trade-offs with multiple types of skills. We extend the model in Section 5 to allow for more than two types of workers. There are $M = 10$ types of workers, each with a mass N_m. The firm j employs $\{L_{mj}\}$ workers and produces y_j using a technology $y_j = z_j \tilde{y}_j^{\eta}$ where $\tilde{y}_j = \sum_m \alpha_{mj} L_{mj}$.

6.1 Parametrization and Calibration

We parameterize the efficiency of a worker of type m in firm j, α_{mj}, as

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

Thus, $\log \alpha_{mj}$ is the product of worker skill $\log X_m$ and firm-specific skill intensity θ_j. We assume that $\log X_m$ follows the normal distribution $N(0, \sigma_x^2)$. The intensity of skill in the firm j θ_j increases in productivity z_j if $\theta > 0$, where θ governs the complementarity between

[33] In an equilibrium without redistribution policy, lower amenity of a firm is compensated for by higher consumption for the high-skilled worker ($\widetilde{CD}^e_{hj} < 0$), while it is negatively compensated for by lower consumption for the low-skilled worker ($\widetilde{CD}^e_{\ell j} > 0$). This is because lower amenity decreases the ratio of $\frac{L_{hj}}{L_{\ell j}}$, which increases MP_{hj} and decreases $MP_{\ell j}$.

Parameters	Values	Target Moments
σ_z	1.00	$\mathbb{V}[\varphi_z] = 0.17$
σ_g	0.47	$\mathbb{V}[\varphi_g] = 0.13$
ρ_{zg}	0.95	$2\mathbb{C}[\varphi_z, \varphi_g] = -0.27$
σ_x	0.57	$\mathbb{V}[\log X] = 0.31$
θ	0.38	$\text{corr}[\log X, \varphi] = 0.37$

Table 2: Calibration

the skill of the worker and the productivity of the firm. We assume that differently-skilled workers value vertical amenities of a firm identically $g_{mj} = g_j, \forall m$.[34,35]

We set $N = 0.88$ and $J = 0.12$ as in Section 4. We set the elasticity of labor supply $\beta = 6$ and return to scale parameter $\eta = 0.75$ (see Online Appendix D.10 for the sensitivity analysis). We calibrate σ_z, σ_g, ρ_{zg}, σ_x and θ to match the variance decomposition of the worker's wage from Lamadon, Mogstad, and Setzler (2022). Table 2 shows the value of the calibrated parameters and the target moments.[36]

6.2 Optimal Markdown Schedule

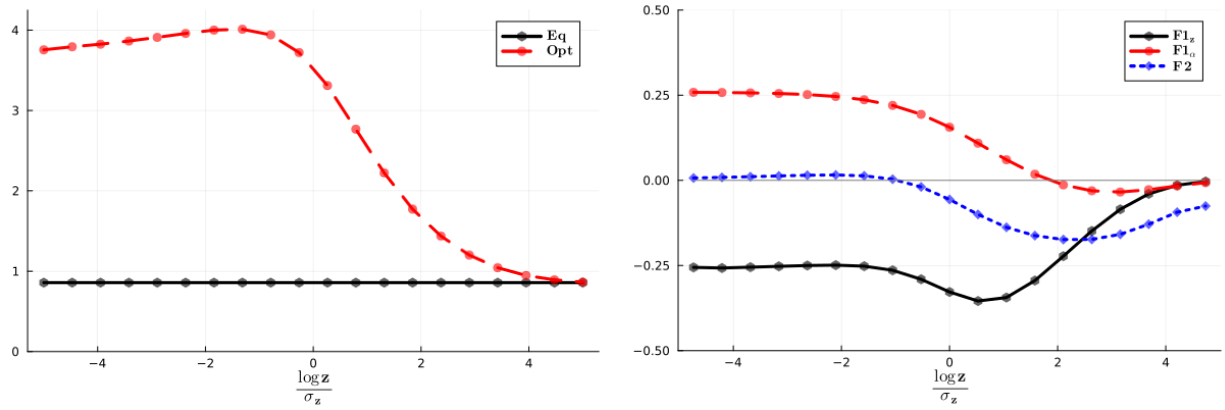
In the previous section, we have showed that the optimal $\mu_j(z_j)$-schedule can be non-monotone if the complementarity channel is strong enough. We now investigate this quantitatively by solving the constrained planner's problem with utilitarian social welfare ($\psi = 1$). Figure 4a plots the optimal markdown, μ_j, against the firm productivity. The planner increases the (employment-weighted) average markdown from 0.86 to 1.76 to increase the share of worker's income from 0.64 to 0.88.[37] The markdown, μ_j, generally decreases in firm productivity, but exhibits an inverted U-shape in the low producitivity region.

[34] In this benchmark, complementarity sorely generates positive sorting of workers into firms.

[35] $\kappa = 1.5$ is set to match the wage ratio of high- to low-skilled workers in a median productivity firm.

[36] When there are multiple types of workers, the equilibrium log wage is the sum of $\theta_j \log X_m$ and the firm pay premium φ_j. A worker with higher skills X_m earns higher wages, and the wage gains from working at productive firms are larger for more skilled workers. See Online Appendix D.1 for details of the calibration.

[37] In the presence of multiple skill types, the average level of μ_j (1.76) is higher than that in the single-type case (1.19) because of additional redistribution between skill types—see Online Appendix E. The level of μ_j can be lowered significantly if there is a more direct redistribution policy such as income tax.



(a) $\mathbb{E}[\mu|z]$ (b) Decomposition of $\frac{d\log\mu}{d\log z}$

Figure 4: Optimal Markdown

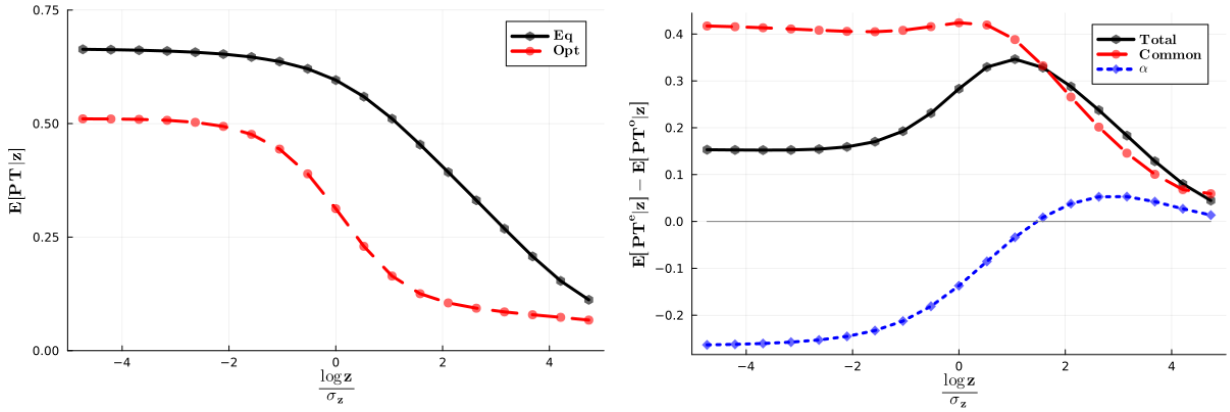
Figure 4b decomposes the slope of the optimal markdown as in Proposition 18. It confirms that negative terms $F1_z$ and $F2$ make the μ_j-schedule decrease in firm productivity in general. The planner compresses the wage schedule to redistribute across workers of different firms, and taxes skill-intensive firms more to redistribute from high- to low-skilled workers. It also confirms that $F1_\alpha$ changes its sign from positive to negative as productivity increases, and in the low productivity region, $F1_\alpha$ dominates, leading to non-monotone μ_j-schedule. When complementarity ($\theta > 0$) is strong enough, the average marginal product of labor is non-monotone, and the planner redistributes to workers at firms with lower average marginal products of labor.[38] In Online appendix D.9, we show that the exact shape of the μ_j-schedule depends on the importance of complementarity in the firm-worker sorting.[39]

6.3 Productivity Pass-through and Compensating Differentials

We now quantitatively investigate how complementarity ($\theta > 0$) affects the PT and CD schedule. Figure 5a plots the PT schedules for a median-skill worker ($m = 5$). To provide redistribution, the planner reduces the PT rates relative to those in equilibrium, generating

[38]Under the benchmark calibration, the schedule of $\overline{MR_j} = \sum_m \frac{L_{mj}}{\sum_{m'} L_{m'j}} MR_{mj}$ weakly decreases in the left tail and increases in the right tail of the firm productivity distribution.

[39]In the benchmark, we only allow demand-side sorting, where complementarity θ sorely generates positive sorting between firms and workers. In Online Appendix D.9, we also allow supply-side sorting. Then, a lower θ is enough to generate the sorting pattern in the data, which leads to weaker effects of $F1_\alpha$.



(a) $PT_{5j} = \mathbb{E}\left[\frac{d\log c_5}{d\log z}|z\right]$ (b) Decomposition of $PT^e_{5j} - PT^o_{5j}$

Figure 5: Productivity Pass-through

the efficiency cost. The planner adjusts the PT schedule to minimize this cost, but different from the single skill type case, Figure 5b shows that the gap of PT's under the constrained planner and the equilibrium is non-monotone in firm productivity as in Proposition 21.

Figure 5b decomposes the gap of PT's into common and skill-specific components. The shape of the common component shows that the planner reduces the efficiency cost of redistribution by setting higher PT rates at firms with a high average marginal product of labor. The gap is smallest in regions with high-productivity firms as in the single skill type case, but in the low productivity region, the gap is non-monotone due to the non-monotonicity of the average marginal product of labor. On the other hand, the skill-specific component inherits the shape of $F_{1\alpha}$, reflecting that the planner provides more redistribution towards workers working at firms with lower average marginal revenue products.

Figure 6 plots the CD schedule and shows that with a decreasing returns to scale production function, the equilibrium level of CD is positive. The constrained planner reduces CD to increase effective vertical amenities and makes the CD schedule increase in firm productivity to minimize efficiency costs. When there is no supply-side sorting with type-independent valuation of vertical amenities, the schedule of CD is identical regardless of worker skill type.

In Online Appendix D.8 we do the welfare gain decomposition andanalyze the case where labor income tax progressivity is fixed to the current U.S. level($\tau = 0.18$).

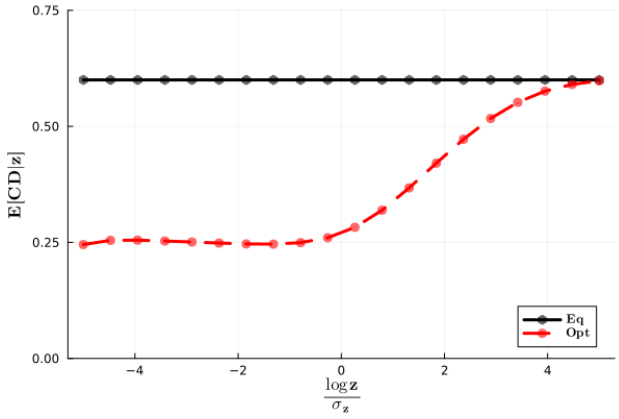


Figure 6: Compensating Differentials

7 Extensions

7.1 Optimal allocation in the presence of Labor Income Tax

So far, we have analyzed the planner's problem assuming no income tax. We show that the main results are robust to extension with redistributive labor income tax-transfer (see Online Appendix F for the detailed derivation).

In the baseline analysis without income tax, the NDC was: $\frac{c_{hj}}{MP_{hj}} = \frac{c_{\ell j}}{MP_{\ell j}} \quad \forall j$.[40] When we introduce the progressive labor income tax, however, the form of NDC should be changed accordingly. We focus on the so-called HSV functional form: $T(w) = w - \phi \cdot w^{1-\tau}$, where τ controls the progressivity and ϕ controls the level of tax. Without skill-dependent policies, the markdown $(\frac{w_{sj}}{MR_j})$ within the same firm should be the same for both skill types, which leads to the following NDC in terms of consumption: $\frac{c_{hj}}{MP_{hj}^{1-\tau}} = \frac{c_{\ell j}}{MP_{\ell j}^{1-\tau}}$. As the planner wants to redistribute from the high-skilled to the low-skilled, the NDC is essentially the following inequality: $\frac{c_{hj}}{MP_{hj}^{1-\tau}} \geq \frac{c_{\ell j}}{MP_{\ell j}^{1-\tau}}$, and increasing the progressivity τ can relax the NDC.

Thus, the introduction of the progressive labor income tax changes the optimal allocation through its effects on the NDC, which is the only channel in the model without intensive labor supply. In the planner's problem, who can directly choose consumption, the elasticity of labor supply with respect to worker's consumption, β, is unaffected by the labor income

[40] This specific form of the NDC can be derived under the linear proportional tax/transfers to the workers—including the zero tax case of the baseline.

tax. Note that if the planner can only use the HSV income tax (e.g. Berger, Herkenhoff, Mongey, and Mousavi (2024)), the key mechanism of optimal redistribution policy is the impact of progressivity on the elasticity of labor supply with respect to pre-tax wage.[41]

To analyze the role of income tax progressivity in a meaningful way, we need to consider the cost of progressivity more realistically. We do this by introducing an intensive margin of labor supply. In this section, we focus on the following special preferences for expositional simplicity, and we provide the analysis under general preferences in Online Appendix F.

Assumption 23. $u(c,h) = \log c - \frac{1}{1+\frac{1}{\epsilon_h}} h^{1+\frac{1}{\epsilon_h}}$

Under Assumption 23, worker's problem has a closed form solution for the working hours:

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

Then $\epsilon_s^P(w_{sj}) = \beta(1-\tau)$, and $\epsilon_s^h(w_{s,j}) = 0$, for all s, j.[42] Thus, even in the presence of intensive margin of labor supply, the NDC-τ constraint is simply $\frac{w_{hj}}{MP_{hj}} = \frac{w_{\ell j}}{MP_{\ell j}}$, and in terms of consumption, it is written as $\frac{c_{hj}}{MP_{hj}^{1-\tau}} = \frac{c_{\ell j}}{MP_{\ell j}^{1-\tau}}$. Then it is straightforward to solve the planner's problem (see Online Appendix F.2 for more details).

The progressivity of the income tax affects welfare 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.[43] The progressivity generates additional trade-offs through the intensive margin of the labor supply. It also affects the optimal balance of the trade-offs by modifying the optimal $\mu(z_j)$-schedule. In Online Appendix F, we show that the modification is mainly through two mechanisms: (i) less use of the markdown schedule for redistribution which is substituted by the progressivity, and (ii) stronger $F2$ channel due to relaxed NDC.

[41] Berger, Herkenhoff, Mongey, and Mousavi (2024) study the role of progressive taxes in the monopsonistic labor market, showing that 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 progressivity can be eliminated by an appropriate choice of markdown.

[42] The intensive elasticity is zero because the income effects and substitution effects exactly cancel out under the specific preferences.

[43] In Online Appendix F, we show that with general preferences, the progressivity can affect welfare through an additional channel—change in extensive elasticities.

7.2 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.

We restrict our attention to the simple case with only one type of worker (as in Section 3) and constant amenity ($g_j = 1$). Then the labor supply curve in firm j is given by $L_j(w_j) = \frac{w_j^\beta}{I} \times N$, where the wage index I is $I = \sum_k w_k^\beta$. When firms are not small and engage in strategic interactions, they take into account the effect of their wage decision w_j on the economy-wide wage index I. 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 the share of the firm j in the labor market. A higher market share implies lower elasticity. In an equilibrium without redistribution policies, firms with non-zero market share set the wage and markdown as follows: $w_j^e = \frac{\beta(1-s_j)}{1+\beta(1-s_j)} \times MR_j$, and $\mu_j^e = \frac{\beta(1-s_j)}{1+\beta(1-s_j)}$.

What is the implication of the nonzero market share on the optimal allocation? One might conjecture that a lower elasticity of labor supply of larger firms implies more redistribution with lower efficiency cost. However, it turns out that the nonzero market share leads to the opposite implication—more subsidy to larger firms. Here, we briefly explain the key effects of nonzero market share and provide more detailed analysis in Online Appendix G.

A market share s_j affects optimal consumption through two channels. First, a higher market share s_j implies a lower elasticity of labor supply and thus lower efficiency costs of wage compression. On the other hand, the planner provides redistribution to workers by assigning a higher transfer Δ_j to a firm with a higher market share s_j, implying a 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.

In decentralization, however, the nonzero market share has a clear implication on the optimal subsidy schedule regardless of the dominance between the two channels. The higher market share of firm j unambiguously implies more subsidy to that firm. This is because

both the firms and the planner face the same labor supply elasticities, and thus higher elasticity due to the market share 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 with nonzero market share implies a higher subsidy rate for the larger firms.

7.3 Economy with both Product and Labor Market Power

We now extend to a model with product market power to show that introducing the product market power magnifies the planner's intervention without changing the direction of the benchmark. Here, we focus on the single type case (see Online Appendix H for more analysis).

In this extension, entrepreneurs produce intermediate goods ($\{y_j\}$) that are then demanded by the final good producers. Final-good producers use the following technology to aggregate intermediate goods $y = \left(\sum_j y_j^{\frac{\delta-1}{\delta}}\right)^{\frac{\delta}{\delta-1}}$, where $\delta < \infty$ is the elasticity of substitution between different intermediate goods. Final-goods producers are competitive, so they take the prices of the intermediate goods $\{p_j\}$ as given. Then the product demand curve faced by the entrepreneur j is $p_j = p(y_j) = \left(\frac{y_j}{y}\right)^{-\frac{1}{\delta}}$, where we have normalized the final good price to 1.

In equilibrium, firms set wages with the double-markdown as in Kroft, Luo, Mogstad, and Setzler (2025): $\mu_j^e = \frac{w_j}{p_j z_j} = \frac{1-\frac{1}{\delta}}{1+\frac{1}{\beta}}$, where $1-\frac{1}{\delta}$ is the inverse markup and $1+\frac{1}{\beta}$ is the standard markdown. On the other hand, the planner can directly regulate production decision of firms by introducing a wedge between MRPL and MCL. The only change in the planner's problem due to the product market power is then the final goods production technology, which makes the firm's intermediate goods imperfectly substitutable ($\delta < \infty$).

The wedge between the equilibrium markdown μ_j^e and the optimal markdown μ_j^o is:

$$\mu_j^o = \mu_j^e \times \frac{1 \; + \; \frac{1}{\left(\frac{y_j}{y}\right)^{-\frac{1}{\delta}} z_j} \times \frac{1}{\beta} \times \left(1 - \beta\nu\frac{I}{N}\right)}{1 - \frac{1}{\delta}}. \tag{37}$$

Relative to the equilibrium markdown μ_j^e, the planner's markdown is adjusted by two factors. First, the planner corrects the inverse markup $1-\frac{1}{\delta}$. In the presence of product market power, 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 $\frac{1}{\left(\frac{y_j}{y}\right)^{-\frac{1}{\delta}} z_j} \cdot \frac{1}{\beta}\left(1-\beta\nu\frac{I}{N}\right)$ in the numerator of (37), which is the same factor as in the benchmark economy. The planner with redistributive concerns provides less subsidization for firms with higher productivity.[44]

The optimal allocation can be implemented with revenue-dependent subsidies to firms. The subsidy schedule $S_j(py)$ that implements the planner's allocation should satisfy $1+S_j'(p_j y_j) = \frac{\mu_j^o}{\mu_j^e} = \frac{1 + \frac{1}{\left(\frac{y_j}{y}\right)^{-\frac{1}{\delta}} z_j} \times \frac{1}{\beta} \times \left(1-\beta\nu\frac{I}{N}\right)}{1-\frac{1}{\delta}}$. The two adjustment factors in the planner's markdown exactly determine the optimal subsidy schedule. Introducing the product market power amplifies the subsidy rates without changing the direction of those without product market power.

8 Conclusion

We have analyzed optimal redistribution policy when firms have monopsonistic wage-setting power. We have the following key findings. First, a redistribution policy reduces the pass-through rate and compensating differentials, which improves consumption inequality and effective amenity at the cost of allocative inefficiency. The optimal redistribution policy can balance the trade-offs by using the refined schedules of PT and CD. Second, firm regulations can improve welfare in the presence of a redistributive income tax because they can be used to set the refined PT and CD schedules which minimize the efficiency cost of redistribution, but the effectiveness depends on the sources of inequality. Third, with multiple types of skills for workers, optimal allocation provides redistribution of both (i) across firms and (ii) between skill types, which makes the schedules of optimal markdown, PT, and CD non-monotone.

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, 983–1014.

[44] In Online Appendix H, we also discuss how the degree of wage compression responds to the introduction of the product market power.

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

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

Berger, D., T. Hasenzagl, K. Herkenhoff, S. Mongey, and E. A. Posner (2025): "Merger guidelines for the labor market," *Journal of Monetary Economics*, 153, 103785.

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

——— (2025): "Minimum Wages, Efficiency, and Welfare," *Econometrica*, 93, 265–301.

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. M. S. Oppenheimer (2025): "Labor Market Power, Tax Progressivity and Inequality," *Working Paper*.

Boar, C., and V. Midrigan (2025): "Markups and Inequality," *The Review of Economic Studies*, 92, 2828–2860.

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, 291–322, doi: 10.1086/720009.

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

Card, D. (2022): "Who Set Your Wage?," *American Economic Review*, 112, 1075–1090.

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, S13–S70, doi: 10.1086/694153.

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

Golosov, M., P. Maziero, and G. Menzio (2013): "Taxation and Redistribution of Residual Income Inequality," *Journal of Political Economy*, 121, 1160–1204, doi: 10.1086/674135.

Gottfries, A., and G. Jarosch (2023): "Dynamic Monopsony with Large Firms and Noncompetes," *NBER Working Paper*.

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

Jarosch, G., J. S. Nimczik, and I. Sorkin (2024): "Granular Search, Market Structure, and Wages," *The Review of Economic Studies*, 91, 3569–3607.

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

Kroft, K., Y. Luo, M. Mogstad, and B. Setzler (2025): "Imperfect competition and rents in labor and product markets: The case of the construction industry," *American Economic Review*, 115, 2926–2969.

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

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

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