CESifo
AREA
CONFERENCES 2020

Public Economics
Online, 8–9 December 2020

Monopsony Power, Income Taxation and Welfare

Albert Jan Hummel

CESifo

Monopsony power, income taxation and welfare[*]

Albert Jan Hummel[†]

October 28, 2020

Abstract

This paper studies the implications of monopsony power for optimal income taxation and welfare. Firms observe workers' abilities while the government does not and monopsony power determines what share of the labor market surplus is translated into profits. Monopsony power increases the tax incidence that falls on firms. This makes labor income taxes less (more) effective in redistributing labor income (profits). The optimal tax schedule is less progressive. Monopsony power alleviates the equity-efficiency trade-off that occurs because the government does not observe ability, but at the expense of exacerbating capital income inequality. I illustrate these findings for the US economy.

JEL classification: H21, H22, J42, J48
Keywords: monopsony, optimal taxation, tax incidence

---

[*] I would like to thank Ashley Craig, Katherine Cuff, Aart Gerritsen, Bas Jacobs, Marcelo Pedroni, Sander Renes, Dominik Sachs, Dirk Schindler, Kevin Spiritus, Christian Stoltenberg and seminar participants at NED 2019 and IIPF 2020 for many helpful comments and suggestions.

[†] University of Amsterdam. E-mail: a.j.hummel@uva.nl.

1 Introduction

There is growing concern among economists and policymakers that firms exercise monopsony power (or buyer power) in labor markets. Recently, the Council of Economic Advisers published an issue brief on labor market monopsony (CEA (2016)) and the topic was extensively discussed during hearings held by the Federal Trade Commission (FTC (2018a,b)) and the House of Representatives.[1] The report and hearings cite a growing body of evidence documenting that (i) labor markets are highly concentrated and (ii) labor market concentration is associated with significantly lower wages (see, e.g., Azar et al. (2018, 2019, 2020), Benmelech et al. (2018), Lipsius (2018), Rinz (2018), Hershbein et al. (2019), Qiu and Sojourner (2019), Arnold (2020), Schubert et al. (2020), Thoresson (2020)). In addition to the potentially adverse effects on employment, output and economic efficiency, many people have voiced concerns about the *distributional* implications of monopsony power.[2]

Are these concerns justified? How should policymakers take monopsony power into account when designing redistributive policies? I study these questions by extending the non-linear tax framework from Mirrlees (1971) with monopsony power. In my model, firms observe workers' abilities while the government does not. Monopsony power determines what share of the labor market surplus is translated into pure economic profits. After-tax profits flow back as capital income to individuals who differ in their ability and shareholdings. The government has a preference for redistribution and optimizes a non-linear tax on labor earnings. I study how monopsony power affects optimal labor income taxation and ultimately, welfare. Furthermore, I illustrate the findings by calibrating the model to the US economy.

The model generates two predictions that are of particular relevance to policymakers. First, monopsony power raises the incidence of labor income taxes that falls on firms and reduces the incidence that falls on workers. Intuitively, income taxes lower the joint firm-worker surplus and monopsony power determines what share of the surplus accrues to firms. As a result, income taxes reduce profits if firms have monopsony power. Second, monopsony power increases inequality in capital income but reduces inequality in labor market payoffs (i.e., after-tax labor earnings minus the disutility of working). This is because monopsony power raises aggregate profits and lowers the aggregate wage bill. As a result, any dispersion in labor (capital) income generated by differences in ability (shareholdings) is mitigated (exacerbated) if firms capture a larger share of the labor market surplus.

Turning to the optimal tax problem, I derive an intuitive expression for the marginal tax rate on labor earnings at each point in the income distribution. This formula demonstrates that taxes on labor earnings are not only used to redistribute labor income, but also to redistribute profits, i.e., capital income. The reason is that part of the incidence of labor income taxes falls on firms if they have monopsony power. Monopsony power thus makes taxes on

---

[1] The hearing on "Antitrust and Economic Opportunity: Competition in Labor Markets" was held on October 29, 2019. See `https://docs.house.gov/Committee/Calendar/ByEvent.aspx?EventId=110152`.

[2] For example, Alan Krueger noted in his address at the 2018 Fed conference in Jackson Hole:

> "... I would argue that the main effects of the increase in monopsony power and decline in worker bargaining power over the last few decades have been to shrink the slice of the pie going to workers and increase the slice going to employers, not to reduce the size of the pie overall." (Krueger (2018))

labor earnings less effective in redistributing labor income, but more effective in redistributing capital income. Whether monopsony power raises optimal marginal tax rates on labor earnings is *a priori* ambiguous and depends on the government's preferences for redistributing labor and capital income. I derive a condition which can be used to determine if monopsony power raises the optimal marginal tax rate at each point in the income distribution. In the typical case where the government wishes to redistribute both labor and capital income, this condition is more likely to be satisfied at lower earnings levels. In that sense, monopsony power makes the optimal tax schedule less progressive.

Monopsony power has an ambiguous effect on welfare. On the one hand, it increases inequality in capital income driven by differences in shareholdings. The associated impact on welfare is negative and proportional to the covariance between welfare weights and capital income. On the other hand, monopsony power decreases inequality in labor market payoffs driven by differences in ability. The associated impact on welfare is positive and proportional to the covariance between welfare weights and labor market payoffs. The reason why monopsony power can raise welfare is that firms observe ability, while the government does not. If firms have monopsony power, they reduce inequality in labor market payoffs generated by differences in ability. In the baseline, this reduction in inequality comes at zero efficiency costs, which can never be achieved with distortionary taxes on labor income. Monopsony power thus alleviates the equity-efficiency trade-off that occurs because the government does not observe ability. Put differently, monopsony power enables the government to exploit the informational advantage of firms, but at the expense of exacerbating inequality in capital income. Depending on the government's preferences for redistribution, it is optimal to have either perfect competition or full monopsony power. I derive conditions which can be used to determine if a marginal increase in monopsony power raises welfare and whether it is optimal to have perfect competition or full monopsony power.

In the baseline version of the model, workers with different abilities suffer to the same extent from monopsony power in the sense that with linear taxes on labor income, firms capture a constant (i.e., non ability-specific) share of the labor market surplus. I also analyze a version of the model where this share varies with ability, for example because individuals differ in their bargaining skills or the number of potential employers. If individuals with higher ability suffer less from monopsony, optimal marginal tax rates are higher and the welfare effect of raising monopsony power is lower than would be the case if monopsony power does not vary with ability. Intuitively, inequality driven by differences in ability is exacerbated if individuals with higher ability suffer less from monopsony.

Two critical assumptions in the analysis are that (i) monopsony power does not generate efficiency losses and (ii) profit taxes are non-distortionary. I relax the first of these by including an extensive (participation) margin and non-observable participation costs. Monopsony power then generates a classic distortion in employment, as individuals do not internalize the profits made by firms when making their participation decision. The optimal policy response is to lower taxes on labor earnings in order to stimulate labor participation. Moreover, monopsony power is less likely to raise welfare if it distorts labor participation. I relax the second of these assumptions by analyzing an extension where profit taxes lead firms to either reduce investment or to engage in costly profit shifting. Both extensions provide a micro-foundation for why the optimal profit tax is less than one, but they have different implications for optimal labor income taxation and the welfare effects of monopsony power. With investment distortions from profit taxes, optimal tax rates on labor income are reduced in order to stimulate labor effort, whereas the condition which can be used to determine if monopsony power raises welfare remains unaffected. By contrast, optimal tax rates on labor income are higher when firms engage in costly profit shifting, as they can be used to reduce aggregate profits and thereby aggregate shifting costs. Moreover, an increase in monopsony power is less likely to raise welfare if firms shift profits to tax havens.

To illustrate the quantitative implications of monopsony power for optimal income taxation and welfare, I calibrate the baseline version of the model to the US economy. The degree of monopsony power is used to target an estimate of the pure profit share from Barkai and Benzell (2018). I find that if the government wishes to redistribute both labor and capital income, monopsony power raises (lowers) optimal marginal tax rates at low (high) earnings levels. Hence, the optimal tax schedule is less progressive. Moreover, taking monopsony power into account when designing tax policy leads to modest welfare gains that range between 0.07% and 1.04% of GDP in the calibrated economy depending on the covariance between welfare weights and shareholdings. By contrast, changing the degree of monopsony power from its value in the calibrated economy to zero can have a large negative or positive impact on welfare (ranging between between –1.78% and +8.37% of GDP), again depending on the covariance between welfare weights and shareholdings. Finally, if the current tax system is optimal, an increase in monopsony power raises welfare only if the negative covariance between welfare weights and after-tax labor income is at least 2.85 times as large as the negative covariance between welfare weights and after-tax capital income.

Related literature. A few papers study optimal income taxation in an environment where firms have monopsony power. As I do, Hariton and Piaser (2007) and da Costa and Maestri (2019) analyze a model where labor supply responds on the intensive (hours, effort) margin, whereas Cahuc and Laroque (2014) focus on the extensive (participation) margin, which I add in an extension. These studies assume that firms – like the government – do *not* observe workers' abilities (Hariton and Piaser (2007) and da Costa and Maestri (2019)) or their reservation wages (Cahuc and Laroque (2014)). Monopsony power then leads to a downward distortion in employment, either in hours worked or the number of individuals employed. To partly off-set this distortion, the government finds it optimal to subsidize employment. This requires *negative* marginal (participation) tax rates if labor supply responds on the intensive (extensive) margin. By contrast, in my model firms observe ability and in the baseline there is no distortion in employment.[3] Optimal marginal tax rates only serve to redistribute income and are generally *positive*. Moreover, in my model monopsony power might raise welfare. This is not possible in Hariton and Piaser (2007), Cahuc and Laroque (2014) and da Costa and Maestri (2019), since firms do not have an informational advantage compared to the

---

[3] I analyze an extension where monopsony power distorts labor participation, because firms do not observe participation costs, as in Cahuc and Laroque (2014).

government about their workers' abilities.

This paper is also related to Kaplow (2019), who studies optimal income taxation in a model with multiple goods where firms sell their products at an exogenous, good-specific mark-up over labor costs. As in the classic model of monopoly, employment and output are inefficiently low. This calls for a downward adjustment in optimal tax rates on labor income. Without variation in mark-ups, such an adjustment would "undo the wrongs" of monopoly and market power has no impact on welfare.[4] The most important difference compared to Kaplow (2019) is that I assume firms offer workers a combination of earnings and labor effort instead of charging consumers a constant mark-up over labor costs. As a result, the outcome in the absence of taxation is efficient in the baseline version of the model. Tax policy is then exclusively aimed at redistribution – not to restore efficiency. Moreover, in my model tax policy cannot off-set the impact of monopsony power. Therefore, monopsony power affects welfare even if there is only one good and hence, no variation in mark-ups.

The model of labor market monopsony I analyze features important similarities and differences with the classic monopsony model from Robinson (1933) and the new monopsony models introduced in Manning (2003). The first similarity is that firms can exercise monopsony power because they face an upward-sloping labor supply curve. In Robinson (1933) and Manning (2003), this is because firms attract more workers if they pay higher wages. In my model, the number of workers available to each firm is fixed, but a firm can increase their labor effort by offering contracts that imply a higher wage per hour. Second, the mark-up of productivity over wages, the measure of "exploitation" due to Pigou (1920), is decreasing in the elasticity of labor supply. Third and in line with empirical evidence, the pass-through of productivity gains into wages is less than one-for-one.[5] The most important difference is that in Robinson (1933) and Manning (2003), monopsony power generates distortions. By contrast, in the baseline version of my model the equilibrium in the absence of taxation is efficient. The same is true in Sandmo (1994), who analyzes a setting where a monopsonist chooses a payment schedule that consists of a fixed income and a wage proportional to output. Sandmo (1994) discusses the distortionary effects and incidence of income taxes, but – like Robinson (1933) and Manning (2003) – he does not analyze how monopsony power affects optimal tax policy or welfare, which is the main goal of this paper. I separately analyze an extension where monopsony power generates distortions in labor participation by lowering the payoff from working.

Outline. The remainder of this paper is organized as follows. Section 2 presents the baseline version of the model. Section 3 analyzes how monopsony power affects optimal income taxation and welfare. Section 4 studies extensions where monopsony power generates efficiency losses and profit taxes are distortionary. Section 5 explores quantitatively the policy and welfare implications of monopsony power by calibrating the model to the US economy. Section 6 concludes. An appendix contains all proofs and additional details of the analysis.

---

[4] If mark-ups vary across goods, market power does affect welfare. Kaplow (2019) shows that optimal policy is aimed at reducing the *spread* in mark-ups.

[5] See, e.g., Kline et al. (2019) for recent evidence on the pass-through from productivity gains into wages.

2 A Mirrleesian model with monopsony power

The basic structure of the model follows Mirrlees (1971). There is a continuum of individuals who differ in their ability. They supply labor on the intensive margin to identical firms, which produce output using a linear technology with labor as the only input. The government has a preference for redistribution but – unlike firms – does not observe individuals' abilities. Instead it can only observe and hence, tax labor earnings. The main departure from the standard model is that I allow for the possibility that firms have monopsony power. Whenever this is the case, firms earn pure economic profits. These profits are taxed linearly and after-tax profits flow back as capital income to individuals according to their heterogeneous shareholdings. Consequently, the model features inequality in labor income generated by differences in ability and inequality in capital income generated by differences in shareholdings. Both types of inequality play an important role in what follows.

2.1 Individuals

There is a unit mass of individuals who differ in their ability $n \in [n_0, n_1]$ and shareholdings $\sigma \in [\sigma_0, \sigma_1]$ with $n_0 > 0$ and $\sigma_0 \geq 0$. Ability measures how much output an individual produces per unit of effort and shareholdings determine in what proportion aggregate profits flow back to individuals. Both ability and shareholdings are taken to be exogenous.[6] Let $H(n, \sigma)$ denote the joint distribution over ability and shareholdings and $h(n, \sigma)$ the corresponding density. The latter is assumed to be strictly positive on its entire support. Moreover, denote by $F(n)$ the marginal distribution of ability with density $f(n)$.

Individuals derive utility from consumption c and disutility from providing labor effort l. Their preferences are described by a quasi-linear utility function $u(c, l) = c - \phi(l)$, where $\phi(\cdot)$ is strictly increasing, strictly convex and satisfies $\phi(0) = \phi'(0) = 0$. The assumption of quasi-linearity is made for analytical convenience and ensures that all variables except capital income vary only with ability (and not with shareholdings).[7] I denote by $l(n) \geq 0$ the labor effort exerted by an individual with ability n. In exchange for her services, she receives labor income $z(n) \geq 0$, which is subject to a labor income tax $T(\cdot)$. Individuals also generate income from holding shares in a diversified portfolio. Each individual's capital income is therefore proportional to the economy's aggregate profits. Denote by $\pi(n) = nl(n) - z(n) \geq 0$ the profits firms generate from hiring a worker with ability n. Aggregate profits are given by

$$\bar{\pi} = \int_{n_0}^{n_1} \pi(n) f(n) dn. \tag{1}$$

Profits are taxed linearly at a rate $\tau \in [0, 1]$ and after-tax profits flow back as capital income to individuals according to how many shares they own. Normalizing aggregate shareholdings

---

[6] Hence, there is no human capital or wealth accumulation. I get back to this point in Section 6.

[7] This would also be the case with Greenwood-Hercowitz-Huffman (GHH) preferences, so that the utility function is of the form $u(c, l) = V(c - \phi(l))$, where $V(\cdot)$ is increasing. I briefly comment on this alternative specification when describing the welfare function below.

to one, the utility of an individual with ability n and shareholdings σ is

$$\mathcal{U}(n,\sigma) = \upsilon(n) + \sigma(1-\tau)\bar{\pi}. \tag{2}$$

Here, $\sigma(1-\tau)\bar{\pi}$ is after-tax capital income and $\upsilon(n) = z(n) - T(z(n)) - \phi(l(n))$ is the payoff from working, or labor market payoff.

2.2 Firms

Firms produce output using an identical, linear technology with labor as the only input. Each firm is matched exogenously with a number of workers.[8] As in Mirrlees (1971), I assume firms perfectly observe the ability of their workers while the government does not.[9] While admittedly a strong assumption, what turns out to be crucial for the results is that firms have an informational advantage about their workers' abilities compared to the government. There are at least two reasons to believe this is the case. First, firms spend significant resources to assess applicants and conduct performance evaluations once workers are hired. By contrast, the main proxy of an individual's ability the government uses for tax purposes is her labor income (which firms also observe). Second, individuals have an incentive to truthfully reveal their ability to firms. On the contrary, high-ability individuals would try to mimic low-ability individuals if the government attempts to tax ability.

To a (potential) employee with ability n, a firm offers a bundle (z,l) which specifies labor earnings $z \geq 0$ and effort (or hours) $l \geq 0$. The firm chooses the bundle to maximize profits, subject to the requirement that the employee's labor market payoff exceeds some threshold, or outside option $\underline{\upsilon}(n)$. The latter is taken as given by firms and weakly increases in ability.[10] As will be made clear below, the outside option determines how much monopsony power firms have and depends on the tax function $T(\cdot)$. If a firm is matched to a worker with ability n, it solves

$$\begin{aligned} \max_{l\geq 0, z\geq 0} \quad & \pi(n) = nl - z, \\ \text{s.t.} \quad & z - T(z) - \phi(l) \geq \underline{\upsilon}(n). \end{aligned} \tag{3}$$

I assume the tax function $T(\cdot)$ is such that the first-order conditions are both necessary and sufficient and denote the solution to the maximization problem (3) by $l(n)$ and $z(n)$.[11] At an interior solution, labor effort and earnings are related through

$$n = \frac{\phi'(l(n))}{1 - T'(z(n))}. \tag{4}$$

At the optimum, firms offer bundles which equate an individual's productivity (on the left-

---

[8] The role of firm heterogeneity is very limited. The only source of heterogeneity is that firms are matched exogenously to different workers.

[9] See Stantcheva (2014), Bastani et al. (2015) and Craig (2020) for an analysis of optimal income taxation if firms do not (perfectly) observe ability. Unlike Hariton and Piaser (2007) and da Costa and Maestri (2019), these studies assume that firms have no monopsony power and competition drives (expected) profits to zero.

[10] This explains why workers have an incentive to truthfully reveal their ability to firms.

[11] See Appendix III for additional details on the second-order conditions.

hand side) to her willingness to substitute between labor effort and earnings (on the right-hand side). The marginal tax rate $T'(z(n))$ distorts labor effort as it drives a wedge between the marginal rate of substitution and the marginal rate of transformation between consumption and labor effort. Without taxes on labor earnings, labor effort is not distorted. Importantly, this is true despite the fact that firms may have monopsony power, as embodied in low values of the outside option $\underline{v}(n)$. The reason why the equilibrium without taxation is efficient is that firms observe ability and take into account how labor earnings and effort affect the utility of its workers. As a result, there are no unexploited gains from trade. It is demonstrated in Section 4.1 that this is no longer true if individuals also supply labor on the extensive margin and firms do not observe participation costs. However, in the baseline version of the model without a participation margin, the equilibrium without taxation is efficient as workers and firms divide the full labor market surplus. How this is done depends on the degree of monopsony power.

2.3 Monopsony power

Monopsony power determines what share of the labor market surplus is translated into pure economic profits or, equivalently, how much utility must be promised to workers (i.e., their outside option). If labor markets are competitive as in Mirrlees (1971), the full labor market surplus accrues to workers as profits are driven to zero: $\pi(n) = 0$ and labor earnings satisfy $z(n) = nl(n)$. This equilibrium occurs if each individual's outside option is to work her preferred number of hours at an hourly wage equal to her productivity.[12] Conversely, if firms have full monopsony power, workers are put on their participation constraint and the entire labor market surplus is translated into profits. In that case, the outside option is $\underline{v}(n) = -T(0)$, where $-T(0)$ is the benefit an individual receives if she rejects the contract offered by firms. The Lagrangian associated with the firm's problem (3) is then

$$\mathcal{L}(n) = nl - z + \kappa_1 \Big[z - T(z) - \phi(l) + T(0) \Big] + \kappa_2 l + \kappa_3 z, \tag{5}$$

where the κ's are Lagrange multipliers. I assume the benefit $-T(0)$ is such that firms do not make profits from hiring the least productive workers: $\pi(n_0) = 0$.[13] To derive an expression for the profits $\pi(n)$ firms generate from hiring *any* worker if they have full monopsony power, differentiate the Lagrangian (5) with respect to ability n. By the envelope theorem, $\mathcal{L}'(n) = \pi'(n) = l(n)$, where $l(n)$ is the labor effort that is offered to an individual with ability n. Integrating this relationship and imposing the boundary condition $\pi(n_0) = 0$ gives an

---

[12] Formally, the outside option – which depends on the tax function $T(\cdot)$ – is then given by

$$\underline{v}(n) = \max_{l} \left\{ nl - T(nl) - \phi(l) \right\}.$$

[13] As is shown in Appendix V, from an optimal tax perspective the assumption that firms do not earn profits from hiring the least productive workers is without loss of generality.

expression for profits if firms have full monopsony power:

$$\pi(n) = \int_{n_0}^{n} l(m)dm. \tag{6}$$

For any intermediate degree of monopsony power, firms capture part of the labor market surplus. In order to study the welfare effects of monopsony power and to keep the optimal tax problem tractable, I choose a specific way to operationalize monopsony power. It is formally defined as follows.

Definition 1. ***Monopsony power*** *$\mu(n) \in [0,1]$ and the profits $\pi(n) = nl(n) - z(n)$ firms generate from hiring a worker with ability n are related through*

$$\pi(n) = \mu(n) \int_{n_0}^{n} l(m)dm. \tag{7}$$

Equation (7) relates firms' profits to the degree of monopsony power. Equivalently, it can be thought of as relating the outside option of workers to firms' monopsony power.[14] Clearly, profits are zero if labor markets are competitive, i.e., if the degree of monopsony power $\mu(n) = 0$. Conversely, if firms have full monopsony power, i.e., if $\mu(n) = 1$, equations (6) and (7) coincide. In this case, the full labor market surplus is translated into profits as workers are put on their participation constraint. At intermediate degrees of monopsony power $\mu(n) \in (0,1)$, firms capture part of the labor market surplus.

If taxes on labor income are linear, the degree of monopsony power $\mu(n) \in [0,1]$ equals the share of the labor market surplus that is translated into pure economic profits whenever a firm hires a worker with ability n. The payoffs for workers and firms then coincide with those obtained under the weighted Kalai-Smorodinsky bargaining solution introduced in Thomson (1994), where the payoff of each party is proportional to her ideal ('utopia') payoff.[15] The weights $\mu(n)$ and $1 - \mu(n)$ can therefore be interpreted as the bargaining power of firms and workers, respectively. Note that these weights may vary with ability, which captures that individuals with different abilities might suffer more or less from monopsony. Whether monopsony power is in-or decreasing in ability is *a priori* unclear. Individuals with higher ability may have better bargaining skills, but also fewer potential employers if they are highly specialized (see Caldwell and Danieli (2018) for empirical evidence). Throughout I assume individuals with higher ability do not suffer more from monopsony power to an extent they are actually worse off.[16] As shown in Appendix II, this assumption is always satisfied if monopsony power does not vary with ability (i.e., $\mu(n) = \mu \in [0,1]$ for all n), which is the case

---

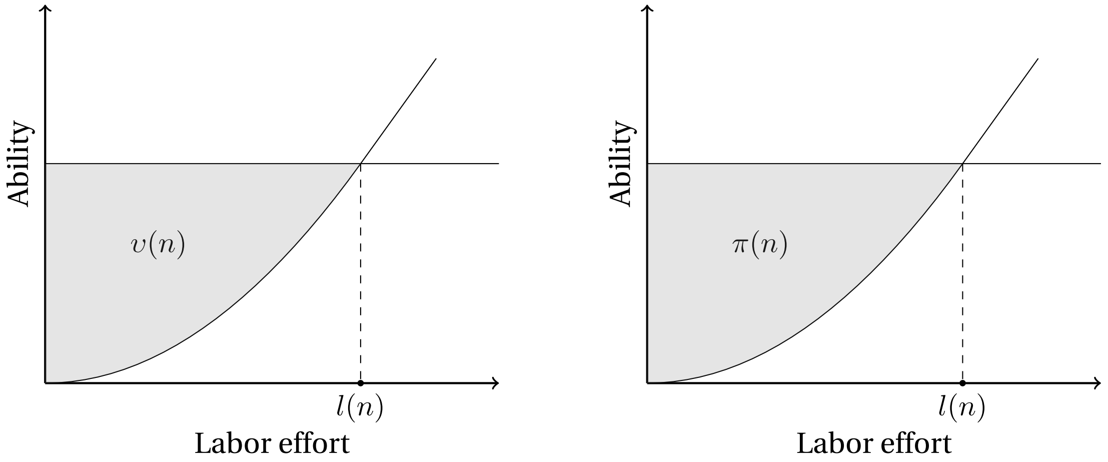
[14] To see this, note that the profits $\pi(n)$ firms generate from hiring a worker with ability n and this same worker's outside option $\underline{v}(n)$ are related through equation (3).

[15] Strictly speaking, the payoffs no longer necessarily coincide with those from the weighted Kalai-Smorodinsky solution if taxes on labor income are non-linear. The reason is that with non-linear taxes, labor effort generally depends on the degree of monopsony power as it is no longer pinned down solely by the first-order condition (4) (which would be the case if $T'(z(n))$ does vary with earnings, i.e., if the tax system is linear). As stated above, the reason for choosing to operationalize monopsony power in this specific way is to guarantee that the optimal tax problem remains tractable and to make it possible to study the welfare effects of monopsony power.

[16] Recall that the outside option (which in equilibrium coincides with the labor market payoff) weakly increases ability: $\underline{v}'(n) \geq 0$ and hence $v'(n) \geq 0$. Thus, individuals with higher ability are not worse off. Equation (53) from Appendix II demonstrates that this assumption implies $\mu'(n)$ is bounded from above.

I focus on in most of what follows.

Figure 1 graphically illustrates how monopsony power affects the payoffs of workers and firms. Here, I assume income taxes are absent: $T(z(n)) = 0$ if $z(n) > 0$. The horizontal line plots an individual's ability and corresponds to the labor demand schedule if labor markets are competitive. The upward-sloping line plots the relationship $\phi'(l) = n$, which – under perfect competition – corresponds to the labor supply schedule. The shaded area shows the labor market surplus. The latter is not affected by the degree of monopsony power. Put differently, monopsony power does not reduce the size of the pie (i.e., does not generate efficiency losses). I study an extension with an extensive margin and non-observable participation costs where monopsony power distorts labor supply in Section 4.1. Without a participation margin, monopsony power only affects how the labor market surplus is split between workers and firms. If labor markets are competitive, firms earn zero profits and the full surplus accrues to workers. The shaded area then corresponds to the individual's labor market payoff $v(n)$: see Figure 1a. Conversely, if labor markets are fully monopsonistic, all surplus accrues to firms. The shaded area then corresponds to profits $\pi(n)$: see Figure 1b.[17]



(a) Perfect competition: $\mu(n) = 0$ (b) Full monopsony power: $\mu(n) = 1$

Figure 1: Labor market equilibrium

2.4 Government

The government's preferences are described by the following social welfare function:

$$\mathcal{W} = \int_{\sigma_0}^{\sigma_1} \int_{n_0}^{n_1} \gamma(n,\sigma)\mathcal{U}(n,\sigma)h(n,\sigma)dnd\sigma. \tag{8}$$

Here, $\gamma(n,\sigma) \geq 0$ is the welfare weight (or Pareto weight) the government attaches to an individual with ability n and shareholdings σ. The average welfare weight is normalized to one. To make sure the government wishes to redistribute from individuals with high to individu-

---

[17] The equilibrium with full monopsony power also occurs if firms engage in first-degree price discrimination. In that case, firms pay workers their reservation wage *for every hour worked*. Hence, the hourly wage depends on the number of hours worked. Firms then continue to demand labor effort up to the point where the worker's productivity is high enough to compensate for the marginal disutility of working.

als with low capital income, I assume the average welfare weight of individuals with the same shareholdings $\mathbb{E}[\gamma(n,\sigma)|\sigma]$ is weakly decreasing in σ. Similarly, to generate a motive to redistribute from individuals with high to individuals with low labor income, I assume the average welfare weight of individuals with the same ability $g(n) = \mathbb{E}[\gamma(n,\sigma)|n]$ is weakly decreasing in n.[18] Using the welfare weights $g(n)$, it is instructive to write the welfare function as follows.

Lemma 1. *The welfare function (8) can be written as*

$$\mathcal{W} = \int_{n_0}^{n_1} \Big[g(n)\upsilon(n) + (1-\Sigma)(1-\tau)\pi(n) \Big] f(n)dn, \tag{9}$$

where $\Sigma = -\mathrm{Cov}[\boldsymbol{\sigma},\boldsymbol{\gamma}] \in [0,1]$ *is the negative covariance between shareholdings and welfare weights, which is bounded between zero and one.*

Proof. See Appendix I. ☐

Individuals derive utility from earning labor income and capital income. Welfare is therefore increasing in the labor market payoff and after-tax profits. Importantly, the extent to which after-tax profits contribute to welfare depends on the covariance between shareholdings and welfare weights. This is because the government wishes to redistribute from individuals with high to individuals with low capital income. A higher concentration of firm-ownership (captured by a higher Σ) therefore lowers the contribution of after-tax profits to welfare. It is worth pointing out that the covariance term Σ, which plays an important role in what follows, is exogenous and bounded between zero and one. It depends only on welfare weights and the distribution of ability and shareholdings. As such, it reflects properties of the joint distribution of capital and labor income and the government's desire to redistribute capital income. An increase in the government's desire to redistribute capital income raises Σ and thereby lowers the contribution of profits to welfare.

Turning to the instrument set, as in Mirrlees (1971) I assume the government does not observe individuals' abilities but only their labor earnings, which are subject to a non-linear tax $T(\cdot)$. In addition, the government observes aggregate profits, which are taxed linearly (either at the firm or the individual level) at an exogenous rate $\tau \in [0,1]$. The government's budget constraint is given by

$$\int_{n_0}^{n_1} \Big[T(z(n)) + \tau\pi(n) \Big] f(n)dn = G, \tag{10}$$

where G denotes an exogenous revenue requirement, which may be positive or negative. Because the government wishes to redistribute from individuals with high to individuals with low shareholdings and the profit tax is non-distortionary, it is optimal to levy a confiscatory tax on profits. One can therefore interpret the exogenous rate τ as the *maximum* share of pure economic profits that can be taxed. Without a restriction on profit taxation, $\tau = 1$.

---

[18] An alternative way to generate a motive for redistribution (without the need to specify exogenous Pareto weights) is to assume the individual utility function is of the GHH-form $u(c,l) = V(c-\phi(l))$, where $V(\cdot)$ is strictly increasing and strictly concave: see footnote 7. Doing so is slightly more complicated but does not generate additional, substantive insights. Another advantage of using exogenous Pareto weights is that in some cases it is possible to derive a closed-form solution for the optimal marginal tax rate, as will be made clear below.

Conversely, if profit taxation is restricted (e.g., due to political constraints or firm lobbying), $\tau < 1$. In Section 4.2, I analyze extensions of the model where profit taxes are distortionary because they induce firms to either reduce investment or to engage in costly profit shifting. Naturally, in those cases, the optimal profit tax is endogenously below one.

2.5 Equilibrium

An equilibrium with monopsony power is formally defined as follows.

Definition 2. *An* ***equilibrium with monopsony power*** *consists of levels of labor effort* $l(n) \geq 0$*, earnings* $z(n) \geq 0$ *and profits* $\pi(n) \geq 0$ *for all* n *such that, for given monopsony power* $\mu(n)$ *and given labor income taxes* $T(\cdot)$*, profit taxes* τ *and government spending* G*,*

(i) labor effort $l(n)$ *and earnings* $z(n)$ *are related through equation (4), or* $l(n) = z(n) = 0$*,*

(ii) profits are given by $\pi(n) = nl(n) - z(n)$ *and satisfy equation (7),*

(iii) the government runs a balanced budget cf. equation (10).

Definition 2 characterizes the equilibrium outcomes for a given set of tax instruments and a given degree of monopsony power. Two remarks are in order. First, given equilibrium effort and earnings, the labor market payoff can be calculated as $v(n) = z(n) - T(z(n)) - \phi(l(n))$, which in equilibrium coincides with the outside option $\underline{v}(n)$. Recall that the latter is taken as given by firms but not by the government, as it depends on the tax function $T(\cdot)$. Second, because of the specific way of modeling monopsony power, finding the equilibrium outcomes requires solving an integral equation if the tax function $T(\cdot)$ is non-linear.[19] As stated before, the main advantage of this modeling choice is that it keeps the optimal tax problem tractable and makes it possible to study the welfare effects of monopsony power. A disadvantage is that it is generally not possible to obtain sharp results when studying tax reforms or the impact of monopsony power on labor market outcomes. Keeping this caveat in mind, it is useful to highlight two implications of monopsony power.

First, monopsony power increases the incidence of labor income taxes that falls on firms and decreases the incidence that falls on workers. To see this, compare the equilibria with $\mu(n) = 0$ (perfect competition) and $\mu(n) = 1$ (full monopsony power) for all n. If labor markets are perfectly competitive, firms earn zero profits – irrespective of the level of taxation. The full incidence of labor income taxes then falls on workers. Conversely, if firms have full monopsony power, all workers are put on their participation constraint: $v(n) = -T(0)$ for all n. An increase in the tax burden $T(z(n))$ at $z(n) > 0$ must then be compensated one-for-one by higher labor earnings as otherwise workers prefer non-employment. In this case, the full incidence of labor income taxes falls on firms.

Second, monopsony power decreases inequality in labor market payoffs generated by differences in ability, but increases inequality in capital income generated by differences in shareholdings. This is because monopsony power increases the share of the labor market

---

[19] The integral equation is $\pi(n) = \mu(n)\int_{n_0}^{n} l(m)dm$, where $l(m)$ solves the first-order condition for profit maximization $m(1 - T'(ml(m) - \pi(m))) = \phi'(l(m))$ at an interior solution. See also footnote 15. A characterization of the equilibrium with a linear tax function can be found in Section 5.

surplus that accrues to firms. An increase in monopsony power thus raises aggregate profits and lowers the aggregate wage bill. This is demonstrated in Section 5 if taxes on labor income are linear and individuals have iso-elastic preferences. Under these assumptions, it is possible to obtain a closed-form characterization of the equilibrium. In the more general case where this is not possible, Appendix II demonstrates that if $\mu(n) = \mu$ for all n, an increase in monopsony power μ lowers inequality in labor market payoffs and raises inequality in capital income.[20] Hence, monopsony power mitigates inequality driven by differences in ability, but exacerbates inequality driven by differences in shareholdings.

I am not aware of any direct evidence either in favor or against these hypotheses. A key challenge is that one needs variation in monopsony power, which should then be linked to measures of tax incidence and inequality. Webber (2015) and Rinz (2018) attempt to do the latter. They find that a lower elasticity of labor supply at the firm level and a higher labor market concentration (the two most commonly used measures of monopsony power: see Azar et al. (2019)) are associated with higher inequality in labor earnings. At first sight, these findings appear inconsistent with the hypothesis that monopsony power reduces inequality in labor market payoffs. However, Section 5 illustrates that the model presented here does not make a clear-cut prediction on the impact of monopsony power on the measures of inequality used in these papers, i.e., the variance in log earnings and the P90/P10 earnings ratio. Moreover, the model can accommodate these findings if individuals with higher ability suffer less from monopsony (i.e., if $\mu'(n) < 0$). Regarding the impact of monopsony power on tax incidence, Saez et al. (2019) find that a payroll tax cut in Sweden raised profits without affecting net-of-tax wages. This result suggests firms have substantial monopsony power, but cannot be used to test if monopsony power increases the tax incidence borne by firms. Benmelech et al. (2018) find support for the closely related hypothesis that the pass-through from productivity gains into wages is lower if labor markets are more concentrated.

3 Optimal tax policy and the welfare effects of monopsony power

This Section analyzes how monopsony power affects optimal income taxation and welfare. For analytical convenience, I start by considering the case where monopsony power does not vary with ability: $\mu'(n) = 0$ for all n. Section 3.1 derives results for optimal income taxation and Section 3.2 analyzes how monopsony power affects welfare. Section 3.3 generalizes the main findings to the case where monopsony power varies with ability.

3.1 Optimal income taxation

The government's problem consists of choosing the non-linear tax function $T(\cdot)$ that maximizes welfare.[21] To solve this problem, I follow the approach pioneered by Mirrlees (1971) and characterize the allocation that maximizes welfare subject to resource and incentive

---

[20] From equations (52) and (53) it can be seen that if $\mu(n) = \mu$ for all n, an increase in the degree of monopsony power μ lowers $v'(n)$ (thereby reducing the dispersion in labor market payoffs) and raises $\pi'(n)$ (thereby raising aggregate profits and exacerbating the dispersion in capital income).

[21] The optimal linear tax problem is analyzed separately in Appendix XII.

constraints. The details can be found in Appendix II. Here, I directly state the first main result of this paper.

Proposition 1. *Suppose monopsony power does not vary with ability:* $\mu(n) = \mu \in [0,1]$ *for all* n*. At an interior solution, the optimal marginal tax rate on labor earnings* $z(n)$ *satisfies*

$$T'(z(n)) = \frac{1-F(n)}{nf(n)}\Big[\mu(1-\tau)\Sigma + (1-\mu)(1-T'(z(n)))\left(1+1/\varepsilon(n)\right)\left(1-\bar{g}(n)\right)\Big], \tag{11}$$

where $\bar{g}(n) \in [0,1]$ *is the average welfare weight of individuals with ability at least equal to* n *and* $\varepsilon(n) = \frac{\phi'(l(n))}{\phi''(l(n))l(n)} > 0$ *is the elasticity of labor supply.*

Proof. See Appendix V. ☐

Proposition 1 gives an expression for the optimal marginal tax rate at each point in the income distribution, which is generally positive and zero only at the top.[22] At the optimum, the marginal tax rate equals a weighted average between two components, where the weights depend on the degree of monopsony power. To understand this result, first consider the case where firms have full monopsony power: $\mu = 1$. The optimal marginal tax rate is then

$$T'(z(n)) = \frac{1-F(n)}{nf(n)}(1-\tau)\Sigma. \tag{12}$$

If firms have full monopsony power, taxes on labor earnings are used exclusively to redistribute capital income and not to redistribute labor income. This is because the full incidence of labor income taxes falls on firms as all workers are put on their participation constraint. An increase in the tax burden must then be compensated one-for-one by higher earnings as otherwise workers prefer non-employment. The purpose of the *marginal* tax rate at earnings level $z(n)$ is to raise the *tax burden* for all individuals with earnings at least equal to $z(n)$.[23] The mass of individuals for whom this is the case equals $1 - F(n)$, which shows up in the numerator of equation (12). Because labor earnings for these workers are increased one-for-one with an increase in the tax burden, the government indirectly taxes profits. This is valuable provided profit taxation is restricted and the negative covariance between welfare weights and shareholdings is positive: $\tau < 1$ and $\Sigma > 0$.[24] The benefits of indirectly taxing profits by raising the marginal tax rate $T'(z(n))$ should be weighed against the costs of distorting labor effort: see equation (4). The distortionary costs are proportional to ability n and the density $f(n)$, which determines for how many individuals labor effort is distorted. Both terms show up in the denominator of equation (12).

It is perhaps surprising that with full monopsony power, the optimal marginal tax rate (12) does not depend on the elasticity of labor supply. The reason is that, as stated above, the entire tax incidence falls on firms if they have full monopsony power. Following an increase

---

[22] Hence, the famous result from Seade (1977) that the optimal marginal tax rate equals zero at *both* end-points does not apply. As will be explained below, this is because the marginal tax rate at the bottom can be used to redistribute capital income by indirectly taxing profits.

[23] Note that individuals with different abilities do not earn the same labor income if firms have full monopsony power. This is because firms demand more labor effort from individuals with higher ability. To compensate them (i.e., to ensure the participation constraint holds), firms must pay higher labor earnings to these individuals.

[24] Section 4.2 presents two extensions where the optimal profit tax satisfies $\tau < 1$.

in the tax burden, firms must pay higher labor earnings as otherwise workers prefer non-employment. This is true *irrespective* of the utility function and hence, irrespective of the convexity in the disutility of labor $\phi(\cdot)$. The latter, in turn, determines the elasticity of labor supply. It follows that the elasticity of labor supply is not relevant for determining the optimal marginal tax rate on labor income if firms have full monopsony power.

The second component in the optimal tax formula (11) is as in the benchmark model without monopsony power. To see this, suppose labor markets are perfectly competitive: $\mu = 0$. The optimal tax formula can then be written as

$$\frac{T'(z(n))}{1-T'(z(n))} = \left(1+\frac{1}{\varepsilon(n)}\right)(1-\bar{g}(n))\left(\frac{1-F(n)}{nf(n)}\right). \tag{13}$$

This is the well-known ABC-formula from Diamond (1998), which Saez (2001) writes in terms of sufficient statistics (in particular, the income distribution and behavioral elasticities). Because profits are zero if labor markets are competitive, the sole purpose of the tax function is to redistribute labor income and not to redistribute profits, i.e., capital income. The optimal marginal tax rate trades off distributional benefits against distortionary costs. The former are captured by the term $1-\bar{g}(n)$, which summarizes how much the government values a transfer from individuals with earnings above $z(n)$ to the government budget. The distortionary costs of a higher marginal tax rate, in turn, are increasing in the elasticity of labor supply $\varepsilon(n)$. For a more detailed explanation of this formula, see Diamond (1998).

According to equation (11), the higher the degree of monopsony power, the more taxes on labor earnings are geared toward redistributing capital income and the less they are geared toward redistributing labor income. Intuitively, monopsony power increases the incidence of income taxes that falls on firms and decreases the incidence that falls on workers. Monopsony power therefore makes labor income taxes less (more) effective in redistributing labor (capital) income. Whether monopsony power raises or lowers optimal marginal tax rates is *a priori* ambiguous and depends crucially on the government's preferences for redistribution. This insight is formalized in the next Corollary.

Corollary 1. *Suppose the utility function is iso-elastic:* $\phi(l) = l^{1+1/\varepsilon}/(1+1/\varepsilon)$*, so that* $\varepsilon(n) = \varepsilon$ *for all* n*. At an interior solution, the optimal marginal tax rate is*

$$T'(z(n)) = \frac{\mu(1-\tau)\Sigma + (1-\mu)(1+1/\varepsilon)(1-\bar{g}(n))}{a(n) + (1-\mu)(1+1/\varepsilon)(1-\bar{g}(n))}, \tag{14}$$

where $a(n) = nf(n)/(1-F(n))$ *is the local Pareto parameter of the ability distribution. If* $(1-\tau)\Sigma > 0$ *and* $z(n_0) > 0$*, an increase in monopsony power raises the marginal tax rate at the bottom of the income distribution. Furthermore, at higher ability levels, an increase in monopsony power raises the marginal tax rate* $T'(z(n))$ *if and only if*

$$((1-\tau)\Sigma))^{-1} < ((1+1/\varepsilon)(1-\bar{g}(n)))^{-1} + a(n)^{-1}. \tag{15}$$

Proof. See Appendix VI. □

Equation (14) gives a closed-form solution for the optimal marginal tax rate in terms of exogenous variables. It follows directly from rearranging equation (11) and plays an important role when exploring the quantitative implications of monopsony power for tax policy in Section 5. Equation (15), in turn, gives a condition which can be used to determine if monopsony power raises or lowers the optimal marginal tax rate at each point in the income distribution. Because monopsony power makes income taxes more (less) effective in redistributing capital (labor) income, the impact of monopsony power on optimal tax rates is generally ambiguous. According to equation (15), the first (positive) effect dominates if profit taxation is severely restricted (i.e., if τ is low) and if the government has a strong preference for redistributing capital income (i.e., if Σ is high). Conversely, the second (negative) effect dominates if the government has a strong preference for redistributing labor income from individuals with high to individuals with low ability (i.e., if $\bar{g}(n)$ is low).[25]

The impact of monopsony power on optimal marginal tax rates varies along the income distribution depending on the behavior of $\bar{g}(n)$ and the local Pareto parameter $a(n)$. Because the average welfare weight of all individuals equals one (i.e., $\bar{g}(n_0) = 1$), condition (15) is always satisfied at the bottom of the income distribution provided $z(n_0) > 0$ (i.e., provided individuals with ability n_0 work). Intuitively, the marginal tax rate at the bottom only serves to indirectly tax profits as it does not help to redistribute labor income from individuals with high to individuals with low ability. This becomes more important if monopsony power increases. At higher levels of income, redistributing labor income from individuals above to individuals below that level becomes on average more valuable: $\bar{g}(n)$ is decreasing. Monopsony power makes income taxes less effective in redistributing labor income as part of the tax incidence falls on firms. *Ceteris paribus*, monopsony power therefore has a smaller positive or a larger negative impact on optimal tax rates at higher income levels.

If the government wishes to redistribute both labor and capital income (i.e., $\bar{g}(n)$ is decreasing in ability and $\Sigma > 0$), condition (15) is more likely to be satisfied at lower levels of income. Monopsony power thus makes the optimal tax schedule less progressive in the sense that it increases (decreases) *marginal* tax rates at lower (higher) levels of earnings. The reason is twofold. First, as explained above, monopsony power makes labor income taxes less effective in redistributing labor income. Hence, monopsony power dampens the "natural" force for increasing marginal tax rates, which is the government's desire to redistribute from individuals with high to individuals with low labor income. Second, in a typical calibration of the ability distribution, the local Pareto parameter $a(n)$ is small at the bottom and larger at middle and high levels of ability. The small value of the local Pareto parameter at the bottom implies that marginal tax rates at low earnings levels are a particularly effective tool to indirectly tax profits: see equation (12). The latter, in turn, becomes more important as monopsony power increases. These observations imply that condition (15) is more likely to be satisfied at lower levels of income and hence, that monopsony power makes the optimal tax schedule less progressive. I explore the quantitative implications of monopsonony power

---

[25] Monopsony power also lowers the optimal marginal tax rate if the local Pareto parameter $a(n)$ is high. The reason is quite mechanical. In the second component of equation (11), monopsony power affects optimal marginal tax rates through the term $T'(z(n))/(1 - T'(z(n)))$. The latter changes faster (and hence, implies a smaller change in the marginal tax rate), the higher is $T'(z(n))$. This is the case if the local Pareto parameter is low. Therefore, a lower Pareto parameter makes it easier for condition (15) to be satisfied.

for optimal marginal tax rates in Section 5.

It is worth pointing out that the optimal marginal tax rate according to equation (14) exceeds 100% if the local Pareto parameter $a(n) < \mu(1-\tau)\Sigma$. Clearly, this violates the first-order condition for profit maximization (4). In that case, the non-negativity constraint on labor effort $l(n) \geq 0$ in the government's optimization problem is binding: see Appendix V for details. Hence, some individuals may not work at the optimal allocation if firms have monopsony power. The reason why the government may find it optimal to set taxes in such a way that some individuals do not work (i.e., $l(n) = 0$ for some n) is that stimulating participation by lowering the tax liability raises aggregate profits if $\mu > 0$, which has a negative impact on welfare if $(1-\tau)\Sigma > 0$. Section 5 demonstrates that this issue is relevant only at the bottom of the ability distribution, where the local Pareto parameter $a(n)$ is low. At higher levels of ability, $a(n) \geq \mu(1-\tau)\Sigma$ and the optimal marginal tax rate is given by equation (14).

3.2 Welfare impact of raising monopsony power

I now turn to analyze how an increase in monopsony power affects welfare. The following Proposition states the second main result of this paper.

Proposition 2. *Suppose monopsony power does not vary with ability and the tax function $T(\cdot)$ is optimized. An increase in monopsony power μ raises welfare if and only if*

$$\mu\Sigma^{v} > (1-\mu)\Sigma^{k}, \tag{16}$$

where $\Sigma^{v} = -\mathrm{Cov}[\boldsymbol{v}, \boldsymbol{\gamma}] \geq 0$ is the negative covariance between labor market payoffs and welfare weights and $\Sigma^{k} = -\mathrm{Cov}[\boldsymbol{\sigma}(1-\tau)\bar{\pi}, \boldsymbol{\gamma}] = \Sigma(1-\tau)\bar{\pi} \geq 0$ is the negative covariance between capital income and welfare weights.

Proof. See Appendix VII. □

Monopsony power raises aggregate profits and lowers the aggregate wage bill. The associated impact on welfare is ambiguous. On the one hand, monopsony power reduces inequality in labor market payoffs generated by differences in ability. The positive welfare effect is captured by the left-hand side of equation (16). On the other hand, monopsony power increases inequality in capital income generated by differences in shareholdings. The negative welfare effect is captured by the right-hand side of equation (16).

To gain further intuition why monopsony power might raise welfare, recall that firms observe ability while the government does not. If labor markets are competitive, firms do not benefit from this information as profits are driven to zero. By contrast, profits are positive if firms have monopsony power. Moreover, the profits firms generate from hiring a worker are increasing in ability. An increase in monopsony power thus reduces inequality in labor market payoffs generated by differences in ability. Importantly, unlike with distortionary taxes on labor income, the reduction in inequality comes at zero efficiency costs. An increase in monopsony power thus alleviates the equity-efficiency trade-off that occurs because the government does not observe ability, cf. Mirrlees (1971). Put differently, monopsony power enables the government to exploit the informational advantage of firms about their workers' abilities. *Ceteris paribus*, the associated impact on welfare is positive.

Stantcheva (2014) also finds that a departure from perfect competition (in her model, adverse selection) can improve social welfare if the government has a preference for redistribution but does not observe ability. My result is similar to hers in the sense that with either adverse selection or monopsony power in the labor market, the benefits of having a higher ability are lower than would be the case if labor markets are competitive. Hence, both adverse selection and monopsony power reduce inequality generated by differences in ability, which has a positive impact on welfare. The mechanism, though, is very different. In the analysis of Stantcheva (2014), firms – like the government – do *not* observe workers' abilities and competitively screen them through non-linear compensation contracts. The use of working hours (or effort) as a screening device hurts high-ability workers relative to low-ability workers. By contrast, in my model firms – unlike the government – *do* observe ability and the reduction in inequality driven by differences in ability occurs because firms generate higher profits from hiring more productive workers. Hence, both adverse selection and monopsony power can improve welfare, but for very different reasons.[26]

The negative welfare effect of monopsony power that occurs because it exacerbates inequality in capital income depends critically on the extent to which pure economic profits can be taxed. If profits are taxed at a confiscatory rate (i.e., if $\tau = 1$), an increase in monopsony power unambiguously raises welfare. This is because monopsony power reduces inequality in labor market payoffs and there is no inequality in capital income that is exacerbated if monopsony power increases. However, in reality it is highly unlikely that profits can be taxed at a confiscatory rate or that doing so would be part of an optimal policy. This is because it is very difficult for policymakers to distinguish between normal returns and above-normal returns and because profit taxes generate distortions (which are introduced in Section 4.2). Hence, in the typical case where taxing profits at a confiscatory rate would be either unfeasible or undesirable, monopsony power has an ambiguous effect on welfare: it reduces inequality in labor market payoffs driven by differences in ability but exacerbates inequality in capital income driven by differences in shareholdings.

Welfare is highest if firms have full monopsony power and there is no restriction on profit taxation, i.e., if $\mu = \tau = 1$. Full monopsony power ensures there is no inequality in labor market payoffs as all workers are put on their identical participation constraint. A confiscatory tax on profits, in turn, guarantees there is no inequality in capital income either. Hence, all individuals are equally well off. The government can implement the first-best allocation by levying a confiscatory tax on profits to finance a universal basic income $-T(0)$. Importantly, the guaranteed income should not be taxed away if individuals earn labor income.[27] Doing so only distorts labor effort and does not generate any distributional benefits. Clearly, the

---

[26] Another way to understand how our results are linked is as follows. In Stantcheva (2014), the assumption that firms do not observe ability makes it less attractive for someone with a high ability to pretend she has a low ability by earning a lower income. Intuitively, misleading the tax authority means that individuals also have to mislead firms. Consequently, adverse selection relaxes the incentive constraints in the optimal tax problem. In a similar vein, monopsony power relaxes incentive constraints in my model as it lowers the benefits of having a higher ability: see equation (53).

[27] Put differently, optimal marginal tax rates are zero. To see this, substitute $\tau = \mu = 1$ in equation (11).

insight that welfare is maximized if firms have full monopsony power and profits are taxed at a confiscatory rate crucially depends on the assumption that (i) monopsony power does not generate efficiency losses and (ii) profit taxes are non-distortionary. I study several extensions in Section 4 where these assumptions are relaxed.

A few remarks are in order. First, equation (16) depends on capital income and labor market payoffs, which are both endogenous. As a result, one cannot conclude that the condition from Proposition 2 is always (never) satisfied if $\mu = 1$ ($\mu = 0$), because in that case $\Sigma^{v} = 0$ ($\Sigma^{k} = 0$) as well. I show in Appendix VII that the welfare effect of raising monopsony power can be written solely as a function of exogenous variables if the labor supply elasticity is constant (i.e., if $\phi(\cdot)$ is iso-elastic). Second, the result from Proposition 2 is derived assuming income taxes are optimized. Hence, condition (16) can only be used to assess the desirability of an increase in the degree of monopsony power at the *current* tax system under the additional assumption that the latter is optimized and hence, reflects the government's preferences for redistribution.[28] Third, labor market payoffs depend on the disutility of working, which is difficult to measure. It is also possible to derive a necessary condition for the desirability of raising monopsony power that depends on the covariance between welfare weights and after-tax labor income, as opposed to labor market payoffs.

Corollary 2. *Suppose monopsony power does not vary with ability and the tax function $T(\cdot)$ is optimized. If labor effort is weakly increasing in ability at the optimal allocation, i.e., $l'(n) \geq 0$, an increase in monopsony power μ raises welfare only if*

$$\mu\Sigma^{\ell} > (1-\mu)\Sigma^{k}, \tag{17}$$

where $\Sigma^{\ell} = -\mathrm{Cov}[\boldsymbol{z} - T(\boldsymbol{z}), \boldsymbol{\gamma}] > \Sigma^{v} \geq 0$ is the negative covariance between welfare weights and after-tax labor income.

Proof. See Appendix VII. ☐

If individuals with higher ability exert more effort, the negative covariance between welfare weights and after-tax labor income exceeds the negative covariance between welfare weights and labor market payoffs: see Appendix VII for details. Therefore, equation (17) gives a necessary condition which can be used to determine if an increase in monopsony power could raise welfare. The advantage compared to the necessary and sufficient condition from Proposition (2) is that condition (17) is easier to assess for policymakers, as it depends on after-tax labor income and not on the disutility of working.

The previous results can be used to assess if an increase in monopsony power raises or lowers welfare. It is also possible to determine the optimal degree of monopsony power.

Proposition 3. *The optimal degree of monopsony power is either $\mu^{*} = 0$ (perfect competition) or $\mu^{*} = 1$ (full monopsony power). Full monopsony power (perfect competition) is optimal if*

---

[28] The welfare weights that make the current tax system optimal can be calculated using the inverse optimal tax method: see Bourguignon and Spadaro (2012).

the following condition holds (does not hold):

$$\int_0^1 \left[\frac{\mu}{1-\mu}\Sigma^v - \Sigma^k\right] d\log\mu > 0. \tag{18}$$

Proof. See Appendix VIII. ☐

There exists no interior degree of monopsony power $\mu^* \in (0,1)$ that maximizes social welfare. As demonstrated formally in Appendix VIII, the welfare function is convex in monopsony power: $\mathcal{W}''(\mu) \geq 0$. This is because monopsony power has a larger positive or negative impact on welfare if individuals exert more labor effort. To see how that implies convexity, suppose monopsony power has a positive impact on welfare, for example because the government strongly dislikes inequality in labor market payoffs. In that case, monopsony power tends to reduce marginal tax rates (see Corollary 1), which in turn raises labor effort. Therefore, the higher the degree of monopsony power, the larger is the positive welfare impact of raising monopsony power. Conversely, if monopsony power has a negative impact on welfare (for example, because the government strongly dislikes inequality in capital income), it tends to raise marginal tax rates cf. Corollary 1, which lowers labor effort. In that case, the smaller the degree of monopsony power, the larger is the positive welfare impact of *reducing* monopsony power. This explains why, depending on the redistributive preferences, it is optimal to have either perfect competition or full monopsony power: $\mu^* = 0$ or $\mu^* = 1$. Equation (18) can be used to determine which of these is optimal. The left-hand side calculates the welfare difference between full monopsony power and perfect competition by integrating over the marginal welfare impact of raising monopsony power $\mathcal{W}'(\mu)$.[29] Full monopsony power is optimal if and only if this difference is positive.

3.3 Ability-specific monopsony power

The results from Propositions 1 and 2 are derived assuming all individuals suffer to the same extent from monopsony power. Hence, if labor income taxes are linear, firms capture a share of the labor market surplus that does not vary with ability: $\mu(n) = \mu$ for all n. I now generalize these results by allowing for the possibility that individuals with different abilities suffer more or less from monopsony. Throughout I maintain the assumption that $\mu'(n)$ is bounded from above in such a way that the labor market payoff is monotone in ability: $v'(n) \geq 0$. In words, individuals with higher ability do not suffer more from monopsony to an extent they are worse off than individuals whose ability is lower.[30]

Proposition 4. *Suppose monosony power $\mu(n)$ varies with ability. At an interior solution, the*

---

[29] To see how Propositions 2 and 3 are related, note that the term in brackets from equation (18) is positive if and only if condition (16) holds. Also, it is worth pointing out that equation (18) cannot be simplified further, because Σ^v and Σ^k are both endogenous and depend on the degree of monopsony power.

[30] In line with this assumption, the findings from Webber (2015) and Rinz (2018) suggest that individuals at lower parts of the earnings distribution suffer more from firms' ability to exercise monopsony power.

optimal marginal tax rate satisfies

$$T'(z(n)) = \frac{1-F(n)}{nf(n)}\Bigg[\bar{\mu}(n)(1-\tau)\Sigma + (1-\mu(n))(1-T'(z(n)))(1+1/\varepsilon(n))(1-\bar{g}(n)) \tag{19}$$
$$- \frac{\mu'(n)\pi(n)(1-T'(z(n)))(1-\bar{g}(n))}{\mu(n)\varepsilon(n)l(n)} - \frac{\int_n^{n_1}\mu'(m)(1-T'(z(m)))\left(\int_m^{n_1}(1-g(s))f(s)ds\right)dm}{1-F(n)}\Bigg],$$

where $\bar{\mu}(n)$ denotes the average monopsony power for individuals with ability at least equal to n. Furthermore, the welfare impact of a proportional increase in monopsony power from $\mu(n)$ to $\mu(n)(1+\alpha)$ is given by

$$\frac{\partial\mathcal{W}(\alpha)}{\partial\alpha} = \int_{n_0}^{n_1}\Bigg[(1-T'(z(n)))\int_n^{n_1}(1-g(m))f(m)dm - (1-\tau)\Sigma\int_n^{n_1}\frac{\mu(m)}{\mu(n)}f(m)dm$$
$$+ \int_n^{n_1}\frac{\mu'(m)}{\mu(n)}(1-T'(z(m)))\left(\int_m^{n_1}(1-g(s))f(s)ds\right)dm\Bigg]\mu(n)l(n)dn. \tag{20}$$

Proof. See Appendix V and VII. □

Compared to the result from Proposition 1, two additional effects show up in the optimal tax formula (19). To understand these, suppose individuals with higher ability suffer less from monopsony: $\mu'(n) < 0$. Inequality generated by differences in ability is then higher than would be the case if monopsony power does not vary with ability. This leads to a higher marginal tax rate for two reasons. First, a reduction in monopsony power *at a particular ability level* implies the labor market payoff increases more quickly in ability. Second, a reduction in monopsony power *at higher ability levels* lowers the profits firms generate from hiring more productive workers. Hence, individuals with higher ability manage to capture a larger share of the labor market surplus. Both effects raise the distributional benefits of income taxes and hence, raise the optimal marginal tax rate.

Equation (20) gives an expression for the welfare effect of raising monopsony power. If monopsony power does not vary with ability, the first (positive) term is proportional to Σ^v and the second (negative) term is proportional to Σ^k. Hence, one additional effect shows up in equation (20) compared to the result from Proposition 2. To understand this effect, consider again the case where individuals with higher ability suffer less from monopsony: $\mu'(n) < 0$. As stated before, individuals with higher ability then capture a larger share of the labor market surplus. This lowers the positive welfare effect of raising monopsony power that occurs because monopsony power mitigates inequality in labor market payoffs driven by differences in ability. Hence, if individuals with higher ability suffer less from monopsony, an increase in monopsony power has a smaller positive or a larger negative impact on welfare compared to the case where monopsony power does not vary with ability.

4 Extensions

This Section presents two types of extensions of the model. In the first (Section 4.1), monopsony power generates a classic distortion in employment by lowering the payoff from working. As a result, the *laissez-faire* equilibrium with monopsony power is no longer Pareto efficient. In the second, taxes on profits, i.e., capital income, are distortionary because they either reduce investment (Section 4.2.1) or induce firms to engage in costly profit shifting (Section 4.2.2). In both types of extensions, I derive optimal tax rules and analyze the welfare effects from monopsony power. For analytical convenience, I focus on the case where monopsony power does not vary with ability.

4.1 Efficiency losses from monopsony power

A critical feature of the model studied so far is that monopsony power does not harm economic efficiency. Put differently, monopsony power affects the way the pie is split, but not its size. The reason is that firms observe the ability of their workers and offer contracts which promise each worker a utility level corresponding to her outside option. As a result, monopsony power distorts neither employment nor hours worked. Naturally, the absence of distortions has implications for optimal income taxation and the welfare effects from monopsony power. I now investigate these implications by extending the model with an extensive (participation) margin and non-observable participation costs.

To model the extensive margin, I follow the standard approach in the literature (see, e.g., Diamond (1980), Choné and Laroque (2011), Jacquet et al. (2013)) and assume individuals also differ in their fixed costs of working, or participation costs $\varphi \in [\varphi_0, \varphi_1]$. Crucially, unlike ability, firms do ***not*** observe participation costs. As a result, the contracts that are offered to workers vary only with their ability and not with their participation costs. The government does not observe participation costs either. Instead, it observes the employment status of each individual and his or her labor income if employed. Consequently, in addition to a linear tax τ on aggregate profits, the government levies a non-linear tax $T(z(n))$ on labor earnings and pays a benefit b to individuals who are not employed.[31]

Because each firm observes the ability of its workers but not their participation costs, the profit maximization problem is the same as before. Whenever a firm is matched to a worker with ability n, it chooses labor effort $l(n)$ and earnings $z(n)$ to maximize profits $\pi(n) = nl(n) - z(n)$, subject to promising a labor market payoff $v(n) = z(n) - T(z(n)) - \phi(l(n))$ that exceeds some ability-specific threshold $\underline{v}(n)$. The latter is taken as given by firms, but not by the government as it depends on the tax function $T(\cdot)$. The threshold pins down the level of profits, which in turn is related to monopsony power according to Definition 1. Equilibrium labor effort, earnings and profits can again be found by solving equations (4) and (7) together with the relationship $\pi(n) = nl(n) - z(n)$. The utility of an individual with ability n, shareholdings σ and participation costs φ is then

$$\mathcal{U}(n, \sigma, \varphi) = \max\{v(n) - \varphi, b\} + \sigma(1 - \tau)\bar{\pi}. \tag{21}$$

Participation costs φ are subtracted from the labor market payoff $v(n)$ as they lower the utility from working. Equation (21) determines a participation threshold $\varphi(n) = v(n) - b$ at every

---

[31] It is useful to distinguish between a benefit b paid to non-participants and the transfer $-T(0)$ an individual receives if she rejects the contract offered by a firm. The latter does not occur in equilibrium, but $-T(0)$ can be used to guarantee that firms do not earn profits from hiring the least productive workers: $\pi(n_0) = 0$.

ability level. Hence, an individual with ability n and participation costs φ becomes employed if and only if $\varphi \leq \varphi(n)$. I denote the participation rate of individuals with ability n by

$$p(\varphi(n)) = \frac{\int_{\sigma_0}^{\sigma_1} \int_{\varphi_0}^{\varphi(n)} h(n, \sigma, \varphi) d\varphi d\sigma}{\int_{\sigma_0}^{\sigma_1} \int_{\varphi_0}^{\varphi_1} h(n, \sigma, \varphi) d\varphi d\sigma}, \tag{22}$$

where $h(\cdot)$ is the density associated with the cumulative distribution $H(\cdot)$ of types.

As before, monopsony power does not distort labor supply on the intensive (effort) margin. However, it does lead to distortions in labor supply on the extensive (participation) margin by lowering the labor market payoff $\upsilon(n)$ that accrues to workers. To see this, suppose there are no taxes and benefits: $T(z(n)) = b = 0$. An allocation is Pareto efficient if a worker becomes employed whenever the *joint* firm-worker surplus exceeds the participation costs: $\upsilon(n) + \pi(n) \geq \varphi$. However, individuals become employed only if their *individual* labor market payoff exceeds the participation costs: $\upsilon(n) \geq \varphi$. Individuals do not internalize the profits made by firms when making their participation decision. Consequently, labor participation is distorted downwards whenever firms make positive profits, i.e., whenever firms have monopsony power. The combination of monopsony power and non-observable (hence, non-contractible) participation costs leads to a hold-up problem as not all workers are willing to "invest" their participation costs if part of the benefits accrue to firms.

The optimal tax problem with an extensive margin and efficiency losses from monopsony power is formally defined in Appendix IX. The next Proposition characterizes optimal tax policy and the welfare impact of monopsony power.

Proposition 5. *Suppose monopsony power does not vary with ability and individuals supply labor on the extensive margin according to equation (21). At an interior solution, the optimal marginal tax rate satisfies*

$$\begin{aligned} T'(z(n)) = \frac{1-F_p(n)}{nf_p(n)} \Big[&\mu(1-\tau)\Sigma + (1-\mu)(1-T'(z(n)))\,(1+1/\varepsilon(n)) \\ &\times \mathbb{E}\left[1 - g_p(m) - \hat{p}(m)\pi(m)(1-(1-\tau)\Sigma) - \hat{p}(m)(T(z(m))+b) | m \geq n\right] \Big], \end{aligned} \tag{23}$$

where the conditional expectation $\mathbb{E}[\cdot]$ is taken using the distribution of employed individuals $F_p(n) = 1 - \int_n^{n_1} p(\varphi(m))f(m)dm$ with density $f_p(n) = p(\varphi(n))f(n)$, $g_p(n)$ is the average welfare weight of participants with ability n and $\hat{p}(n) = p'(\varphi(n))/p(\varphi(n))$ is the semi-elasticity of the participation rate with respect to the participation threshold $\varphi(n) = \upsilon(n) - b$. Furthermore, an increase in monopsony power μ raises welfare if and only if

$$\begin{aligned} &\mu \int_{n_0}^{n_1} \upsilon(n)(1-g_p(n))f_p(n)dn > (1-\mu)\Sigma^k \\ &+ \mu \int_{n_0}^{n_1} \upsilon(n)\hat{p}(n)\pi(n)(1-(1-\tau)\Sigma)f_p(n)dn + \mu \int_{n_0}^{n_1} \upsilon(n)\hat{p}(n)(T(z(n))+b)f_p(n)dn. \end{aligned} \tag{24}$$

Proof. See Appendix IX. □

The optimal marginal tax rate (23) differs in two important ways from the expression stated in

Proposition 1. First, monopsony power generates a downward distortion in labor participation. This is the case whenever the participation response is positive and firms make profits, i.e., whenever $\hat{p}(m)\pi(m) > 0$ for some m. This term shows up on the second line of equation (23). The tax system is used to partly off-set these distortions. This is achieved by reducing the marginal tax rate $T'(z(n))$, which lowers the tax burden for all employed individuals with ability $m \geq n$. Provided part of the tax incidence falls on workers, i.e., provided $\mu < 1$, the reduction in the tax liability raises labor participation. This generates a positive externality, as individuals do not take into account the profits made by firms when deciding whether or not to participate. The negative impact on the optimal marginal tax rate is scaled down by a factor $1 - (1 - \tau)\Sigma \in [0, 1]$, which reflects that the government dislikes inequality in capital income. Provided the government values profits to some extent, i.e., provided $(1 - \tau)\Sigma < 1$, the distortions from monopsony power reduce optimal marginal tax rates.

Second, changes in the participation rate affect government finances. A higher marginal tax rate $T'(z(n))$ raises the tax burden for individuals with earnings at least equal to $z(n)$. Provided the tax incidence falls partly on workers, i.e., provided $\mu < 1$, the increase in the tax burden lowers the participation rate for individuals with ability $m \geq n$. The change in the participation rate affects the government budget through the participation tax $T(z(m)) + b$, which also shows up in the second line of equation (23). The participation tax is positive for most values of m if the government wishes to redistribute on average from employed to non-employed individuals. The optimal marginal tax rate is then lower than would be the case if individuals only supply labor on the intensive margin. This modification of the optimal tax formula is also present if firms do not have monopsony power. See, e.g., Saez (2002), Jacquet et al. (2013), Jacobs et al. (2017) and Hansen (2019), who study optimal taxation with labor supply responses on both the intensive and extensive margin in the context of competitive labor markets. Substituting $\mu = 0$ and $\pi(m) = 0$ for all $m \geq n$ in equation (23) gives an optimal tax formula that is very similar to the ones derived in these papers.[32]

Equation (24) generalizes the result from Proposition 2. Without a participation margin, $g_p(n) = g(n)$, $f_p(n) = f(n)$ and $\hat{p}(n) = 0$ for all n, the left-hand side simplifies to $\mu\Sigma^{\upsilon} = -\mu\text{Cov}[\boldsymbol{\upsilon}, \boldsymbol{\gamma}]$ and both terms on the second line cancel. Compared to equation (16), the desirability condition is modified in two substantive ways. First, monopsony power generates efficiency losses if $\hat{p}(n)\pi(n) > 0$, as captured by the first term on the second line. Provided the government values profits to some extent, i.e., provided $(1 - \tau)\Sigma < 1$, the fact that monopsony power distorts labor participation lowers the welfare impact of raising monopsony power. Put differently, distortions from monopsony power make it less likely that an increase in monopsony power raises welfare. Second, changes in the participation rate generate a fiscal externality that is proportional to the participation tax $T(z(n))+b$, which shows up on the second line as well. According to equation (24), the welfare impact of raising monopsony power is lower if the participation tax is positive for most values of n. By lowering labor participation, monopsony power has a negative impact on government finances. Again, this makes it less likely that an increase in monopsony power raises welfare.

---

[32] There are slight differences when it comes to presentation. For example, Saez (2002), Jacquet et al. (2013) and Jacobs et al. (2017) write the optimal tax formula in terms of sufficient statistics and Saez (2002) and Hansen (2019) consider a model with a discrete set of ability levels.

Recall that without a participation margin, monopsony power unambiguously raises welfare if profits are taxed at a confiscatory rate: see Proposition 2 and note that $\Sigma^k = 0$ if $\tau = 1$. This is because monopsony power lowers inequality in labor market payoffs and there is no inequality in capital income that is exacerbated if monopsony power increases. However, if individuals also supply labor on the participation margin, monopsony power has an ambiguous effect on welfare even if there is no restriction on profit taxation. The reason is that monopsony power not only reduces inequality in labor market payoffs driven by differences in ability (which has a positive impact on welfare); it also distorts labor participation (which has a negative impact on welfare). Hence, the reduction in inequality generated by differences in ability does not come at zero efficiency costs if monopsony power distorts labor participation. This explains why monopsony power has an ambiguous effect on welfare even if profit taxation is unrestricted: it alleviates the equity-efficiency trade-off that occurs because the government does not observe ability, but at the expense of distorting labor participation. The first (second) of these effects is more likely to dominate if labor supply responses on the intensive (extensive) margin are important.

If there are no efficiency losses from monopsony power (as in Section 3) and the government is only concerned with efficiency (i.e., does not value redistribution), optimal marginal tax rates are zero and an increase in monopsony power does not affect welfare. This is no longer the case if monopsony power distorts labor participation.

Proposition 6. *Suppose monopsony power does not vary with ability and individuals supply labor on the extensive margin according to equation (21). If the government is only concerned with efficiency (i.e., if it attaches the same welfare weight to all individuals), the optimal marginal tax rate is non-positive:* $T'(z(n)) \leq 0$ *with a strict inequality for* $z(n) \in (z(n_0), z(n_1))$ *if* $\mu \in (0,1)$. *The optimal tax formula (23) then simplifies to*

$$\frac{T'(z(n))}{1-T'(z(n))}\frac{\varepsilon(n)}{1+\varepsilon(n)}nf_p(n) = -(1-\mu)\int_n^{n_1} \hat{p}(m)\big(T(z(m)) + b + \pi(m)\big)f_p(m)dm. \tag{25}$$

Furthermore, monopsony power has a non-positive impact on welfare: $\partial\mathcal{W}/\partial\mu \leq 0$ *with a strict inequality if* $\mu > 0$.

Proof. See Appendix IX. □

Proposition 6 characterizes optimal tax policy if the government is solely concerned with efficiency and does not value redistribution. In that case, optimal marginal tax rates are non-positive. This is because the optimal tax system balances distortions on the intensive and extensive margin, as illustrated by equation (25). To fully off-set distortions from monopsony power on labor participation, the tax system should satisfy $T(z(n)) + b = -\pi(n)$.[33] In words, participation should be subsidized to make sure workers, when making their participation decision, internalize the profits made by firms. Because earnings and profits are increasing in ability, this tax system features *negative* marginal tax rates. But non-zero marginal tax rates

---

[33] An individual participates whenever $v(n) - \varphi \geq b$ or, equivalently, whenever $z(n) - T(z(n)) - \phi(l(n)) - b \geq \varphi$. If $T(z(n)) + b = -\pi(n)$, this condition can be written as $z(n) + \pi(n) - \phi(l(n)) = nl(n) - \phi(l(n)) \geq \varphi$. This tax system off-sets distortions from monopsony power on labor participation, because individuals participate whenever the total labor market surplus is positive: $nl(n) - \phi(l(n)) - \varphi \geq 0$.

generate distortions in labor effort: see equation (4). Therefore, fully off-setting distortions in labor participation from monopsony power is not part of an optimal policy. Instead, the government trades off lower labor supply distortions on the extensive margin (on the right-hand side) against lower labor supply distortions on the intensive margin (on the left-hand side). This is achieved by setting negative marginal tax rates.

There are two cases where optimal marginal tax rates are zero at all earnings levels if the government is only concerned with efficiency. First, if labor markets are competitive, i.e., if $\mu = 0$, firms do not make profits and there are no distortions in labor participation from monopsony power. The optimal tax system then satisfies $T'(z(n)) = T(z(n)) + b = 0$ for all n. This tax system implements the first-best allocation, as there are neither distortions on the intensive margin nor on the extensive margin.[34] Second, optimal marginal tax rates are zero as well if firms have full monopsony power. To see this, substitute $\mu = 1$ in equation (25). Recall that with full monopsony power, the entire tax incidence falls on firms. In that case, a reduction in marginal tax rates raises profits without stimulating labor participation. As a result, marginal tax rates are completely ineffective in alleviating distortions from monopsony power on labor participation. Because they do generate distortions in labor effort cf. equation (4), it follows that optimal marginal tax rates are zero as well if firms have full monopsony power. This is true despite the fact that monopsony power distorts labor participation.

The final result from Proposition 6 states that an increase in monopsony power lowers welfare if the government does not value redistribution. This is because monopsony power distorts labor participation and as a result, the allocation with monopsony power is not Pareto efficient. A further increase in monopsony power is undesirable from the perspective of a government that is only concerned with efficiency.

4.2 Distortionary profit taxes

The results up to this point have been derived assuming the profit tax is exogenous and does not affect economic decisions. Given there are distributional benefits associated with taxing profits (provided $\Sigma > 0$), it immediately follows that the optimal profit tax equals $\tau = 1$. In reality, a tax on profits or, more generally, capital income distorts many decisions, as it affects incentives to save and invest, to engage in profit shifting, to finance with debt or equity, etc. How the main results regarding optimal income taxation and the welfare impact of monopsony power should be modified in the presence of distortionary profit taxes turns out to depend on the type of distortions profit taxes generate. I illustrate this point by presenting an extension of the model where profit taxes lead firms to either reduce investment (Section 4.2.1) or to shift profits to a tax haven (Section 4.2.2). Both extensions provide a micro-foundation for why the optimal profit tax is less than one, but they have different implications for optimal income taxation and the welfare effects from monopsony power.

---

[34] Note that $T(z(n)) + b = 0$, but $T(z(n))$ or b separately may be different from zero depending on the revenue requirement G. A non-negativity constraint on individual consumption would then require $G \leq 0$.

4.2.1 Investment

My model abstracts from productive capital and all profits made by firms are pure economic rents. As is well known, a tax on pure economic rents is non-distortionary: the choices which maximize before-tax profits also maximize after-tax profits. This is no longer the case if the profits made by firms are quasi-rents generated by prior investments and not all investment costs are tax deductible. The latter could reflect that some costs are difficult to verify (e.g., those related to entrepreneurial effort) or that previously incurred losses can be carried forward for tax purposes only for a limited number of years.

I introduce investment distortions from profit taxes in the simplest way possible. Suppose firms can invest a fraction $I \in [0, 1]$ of their output to generate productivity growth of $A(I)$ percent. The function $A(\cdot)$ is strictly increasing, strictly concave and satisfies $A(0) = 0$. Importantly, investment costs are not tax deductible.[35] Therefore, the after-tax profits a firm generates from hiring a worker with ability n are

$$\hat{\pi}(n) = \max_{I\in[0,1]} \big\{((1 + A(I))nl(n) - z(n))(1 - \tau) - Inl(n)\big\}, \tag{26}$$

where $l(n)$ and $z(n)$ reflect the optimal choice of labor effort and earnings (see Appendix X). The first-order condition with respect to the investment rate I can be rearranged to find

$$A'(I) = \frac{1}{1 - \tau}. \tag{27}$$

The investment rate depends negatively on the profit tax: the higher is the profit tax, the lower is the share of output firms devote to investment. This is because the benefits of investment are taxed, but the costs are not fully tax deductible. See Djankov et al. (2010) and references therein for empirical evidence on the adverse impact of corporate taxes on investment. In the current framework, the combination of profit taxes and non-deductible investment costs leads to a downward distortion in investment. To see this, suppose the government does not tax profits or that all investment costs are tax deductible. In either case, firms continue to invest until the social marginal benefits of a higher investment rate are equal to the social marginal costs: $A'(I) = 1$. The equilibrium investment rate according to equation (27) is lower, i.e., distorted downward, whenever $\tau > 0$.

Appendix X characterizes equilibrium, sets up and solves the optimal tax problem and derives an expression for the welfare impact of monopsony power. The next Proposition summarizes the main results.

Proposition 7. *Suppose monopsony power does not vary with ability and the investment rate* $I(\tau) \in [0, 1]$ *is determined by equation (27). At an interior solution, the optimal profit tax* τ *satisfies*

$$\frac{\tau}{1 - \tau} = \frac{\Sigma^k}{\mathcal{I}\epsilon_{I,r}}, \tag{28}$$

---

[35] It is straightforward to allow for the possibility that a fraction of investment costs are tax deductible. The results are qualitatively the same as long as this fraction is below one.

where $\mathcal{I} = \int_{n_0}^{n_1} I(\tau)nl(n)f(n)dn$ *denotes aggregate investment and* $\epsilon_{I,r} = \frac{\partial I}{\partial r}\frac{r}{I} > 0$ *measures the percentage increase in the investment rate if, following a reduction in the profit tax* τ*, there is a one percent increase in the investment-retention rate* $r(\tau) = (1 + A(I(\tau)))(1 - \tau) - I(\tau)$*. The optimal marginal tax rate, in turn, satisfies*

$$T'(z(n)) = \frac{1 - F(n)}{nf(n)}\left[\mu\Sigma(1-\tau) + (1-\mu)(1 - T'(z(n)))(1 + 1/\varepsilon(n))(1 - \bar{g}(n))\right] - \frac{\tau I(\tau)}{r(\tau)}. \quad (29)$$

Furthermore, an increase in monopsony power μ *raises welfare if and only if*

$$\mu\Sigma^{\upsilon} > (1-\mu)\Sigma^{k}. \quad (30)$$

Proof. See Appendix X. ☐

Equation (28) gives an inverse elasticity rule for the optimal profit tax. At the optimum, the government balances distributional benefits from taxing profits (in the numerator) against the distortionary costs (in the denominator). An increase in the government's desire to redistribute capital income, i.e., an increase in Σ^k, raises the optimal profit tax. If the government has no preference for redistributing capital income, i.e., if welfare weights do not vary with shareholdings, the optimal profit tax equals zero. The distortionary costs of taxing profits, in turn, are increasing in aggregate investment $\mathcal{I}$ and the responsiveness of the investment rate to changes in the profit tax. The behavioral response $\epsilon_{I,r} > 0$ measures the percentage increase in the investment rate if, following a reduction in the profit tax τ, there is a one percent increase in the investment-retention rate $r(\tau) = (1 + A(I(\tau)))(1 - \tau) - I(\tau)$. The latter is maximized by the investment rate $I(\tau)$ and captures what fraction of output can be paid out as dividends after labor costs are subtracted. According to equation (28), the larger are the investment distortions, the lower is the optimal profit tax.

There is one important difference in the expression for the optimal marginal tax rate on labor earnings (29) compared to the result from Proposition 1, where it was assumed that the the profit is exogenous and does not affect economic decisions. The difference is captured by the last term on the right-hand side of equation (29). With non-deductible investment costs, the profit tax not only distorts investment downward cf. equation (27), but also labor effort (see Appendix X for details). Intuitively, a larger share of the additional output that is generated by an extra unit of labor effort accrues to the government if investment costs are not tax deductible. A higher profit tax therefore not only leads firms to reduce investment, but also to offer their employees bundles with lower labor effort. The government can partly alleviate the downward distortion in labor effort from profit taxes by lowering the marginal tax rate on labor earnings. This explains why *ceteris paribus* the optimal marginal tax rate is lower than would be the case with a non-distortionary profit tax.[36]

The condition which can be used to determine if an increase in monopsony power raises welfare is the same as before: compare equations (30) and (16). As is the case without investment distortions from profit taxes, monopsony power decreases inequality generated by

---

[36] It is worth pointing out that for a given profit tax, the last term on the right-hand side of equation (29) also shows up if labor markets are competitive, i.e., if $\mu = 0$. However, in that case the optimal profit tax is zero, as there are no aggregate profits and hence, welfare weights do not covary with capital income: $\Sigma^k = 0$.

differences in ability but increases inequality generated by differences in shareholdings. An increase in monopsony power raises welfare if and only if the first, positive effect (on the left-hand side) outweighs the second, negative effect (on the right-hand side).

4.2.2 Profit shifting

Another source of distortions from profit taxes is that they induce firms to engage in activities to prevent paying these taxes. For example, in a recent paper Tørsløv et al. (2020) estimate that approximately 40% of multinational profits are shifted to tax havens. Unlike investment, these activities are not generally considered to be productive in the sense that they shift outward the production possibility frontier. On the contrary, there could be sizable (opportunity) costs associated with profit shifting.[37]

This Section presents a simple model of costly profit shifting that is similar in spirit to Hines and Rice (1994). Firms can choose to shift a fraction $s \in [0,1]$ of pretax profits to a tax haven at a cost of $\tilde{\rho}(s) \in [0,1]$ per dollar shifted. Hence, as with iceberg transport costs, only a fraction $1-\tilde{\rho}(s)$ of the shifted profits reaches its final destination (i.e., returns to shareholders). The function $\tilde{\rho}(s)$ is increasing, weakly convex and satisfies $\tilde{\rho}(0)=0$. It captures in a reduced-form way the costs associated with shifting profits, often referred to as "concealment" costs (see, e.g., Haufler and Schjelderup (2000)). Firms choose the share s to maximize after-tax payments to shareholders:

$$\Pi(\bar{\pi},\tau) = \max_{s\in[0,1]} \left\{ s\bar{\pi}(1-\tilde{\rho}(s)) + (1-s)\bar{\pi}(1-\tau) \right\} = \max_{s\in[0,1]} \left\{ (1-(1-s)\tau - \rho(s))\bar{\pi} \right\}. \tag{31}$$

Here, $\rho(s) = s\tilde{\rho}(s)$ denotes the total shifting costs per unit of pretax profits $\bar{\pi}$. At an interior solution, the first-order necessary and sufficient condition is

$$\rho'(s) = \tau. \tag{32}$$

Firms continue to shift profits until the marginal costs of doing so (on the left-hand side) are equated the marginal benefits in the form of tax savings (on the right-hand side). The assumptions on $\tilde{\rho}(\cdot)$ guarantee that the share of profits shifted to tax havens is increasing in the profit tax and that this share is positive whenever profits are taxed, i.e., whenever $\tau > 0$. Despite being privately optimal, from a social perspective profit shifting is costly because it reduces the total amount of resources available for consumption.

I analyze the optimal tax problem and the welfare impact of raising monopsony power in Appendix XI. The main findings are summarized below.

Proposition 8. *Suppose monopsony power does not vary with ability and firms shift a fraction $s(\tau) \in [0,1]$ of pretax profits to a tax haven according to equation (32). At an interior solution, the optimal profit tax τ satisfies*

$$\frac{\tau}{1-\tau} = \frac{\Sigma}{\epsilon_{1-s,1-\tau}}, \tag{33}$$

---

[37] See, e.g., Huizinga and Laeven (2008) for an attempt to quantify these costs.

where $\epsilon_{1-s,1-\tau} = \frac{\partial(1-s)}{\partial(1-\tau)}\frac{1-\tau}{1-s} > 0$ *is the elasticity of the share of profits not shifted with respect to the net-of-tax rate. The optimal marginal tax rate, in turn, satisfies*

$$T'(z(n)) = \frac{1-F(n)}{nf(n)}\Bigg[\mu\Big(\big[1-(1-s(\tau))\tau-\rho(s(\tau))\big]\Sigma+\rho(s(\tau))\Big) \\ +(1-\mu)(1-T'(z(n)))(1+1/\varepsilon(n))(1-\bar{g}(n))\Bigg]. \tag{34}$$

Furthermore, an increase in monopsony power μ *raises welfare if and only if*

$$\mu\Sigma^{v} > (1-\mu)\Sigma^{k} + (1-\mu)R, \tag{35}$$

where $R = \rho(s(\tau))\int_{n_0}^{n_1}\pi(n)f(n)dn$ *denotes the total costs of profit shifting.*

Proof. See Appendix XI. ☐

Equation (33) gives an expression for the optimal profit tax that is similar to the first result from Proposition 7. The optimal profit tax is higher, the larger are the distributional benefits (in the numerator) and the smaller are the distortionary costs (in the denominator). The elasticity $\epsilon_{1-s,1-\tau}$ measures the responsiveness of profit shifting activities to changes in the profit tax. Clearly, the optimal profit tax is zero if there are no distributional benefits associated with taxing profits: $\Sigma = 0$ implies $\tau = 0$. Conversely, if the government has a preference for redistributing capital income, i.e., if $\Sigma > 0$, the optimal profit tax is positive. According to equation (33), the distributional benefits of taxing profits should be weighed against the distortionary costs of increased profit shifting.

The expression for the optimal marginal tax rate (34) is almost the same as before, see Proposition 1. The only difference is that the term $(1-\tau)\Sigma$ is replaced by $(1-(1-s)\tau-\rho(s))\Sigma+\rho(s)$. There are two, distinct reasons why the optimal marginal tax rate on labor income is higher when firms engage in costly profit shifting. First, the distributional benefits of using taxes on labor income to indirectly tax profits are larger if firms can shift a fraction of their profits to tax havens: $(1-(1-s)\tau-\rho(s))\Sigma \geq (1-\tau)\Sigma$.[38] Second, profit shifting is costly because it reduces the amount of resources available for consumption: $\rho(s) > 0$ whenever $s > 0$. If firms have monopsony power, i.e., if $\mu > 0$, the government can use taxes on labor income to reduce aggregate profits. A reduction in profits, in turn, lowers the aggregate resource costs associated with shifting profits, which is socially desirable. Taxes on labor income can thus be used to lower the costs of profit shifting.

The final result from Proposition 8 states that the existence of profit shifting opportunities makes it less likely that an increase in monopsony power raises welfare: compare equations (35) and (16). Intuitively, an increase in monopsony power raises aggregate profits and thereby exacerbates inequality in capital income. If the government has a preference for redistributing capital income, it levies a positive tax on profits. The profit tax, in turn, leads firms to shift profits to tax havens. Doing so is costly from a social perspective, because it reduces the resources available for consumption by an amount equal to R. Therefore, the

---

[38] To see this, note that $s \in [0,1]$ is chosen to maximize $\Lambda(s) = 1-(1-s)\tau-\rho(s)$, the term multiplied by Σ, and that $\Lambda(0) = 1-\tau$ because $\rho(0) = 0$. Together, these observations imply $1-(1-s)\tau-\rho(s) \geq 1-\tau$.

welfare impact of raising monopsony power is lower if firms engage in costly profit shifting.

5 Numerical illustration

This Section quantitatively explores the implications of monopsony power in the baseline version of the model where monopsony power does not vary with ability. After presenting the calibration (Section 5.1) and the welfare function (Section 5.2), I analyze how monopsony power affects optimal income taxation (Section 5.3) and welfare (Section 5.4).

5.1 Calibration

5.1.1 Data

I calibrate the model on the basis of US data. The primary data source is the March release of the 2018 Current Population Survey (CPS), which provides detailed information on income and taxes for a large sample of individuals. For each individual I observe taxable income, the tax liability (computed as the sum of federal and state taxes) and income from wage and salary payments. In the remainder the latter is referred to as labor income, or labor earnings. In the analysis I include individuals between 25 and 65 years who derive strictly positive labor income and whose hourly wage is at least half the federal minimum wage of $7.25. For individuals whose labor income is top-coded I multiply the reported income with a factor 2.67, consistent with an estimate of the Pareto parameter of 1.6 for the distribution of labor income at the top obtained by Saez and Stantcheva (2018).[39]

5.1.2 Functional forms

To calibrate the model I require a specification of the utility function and the current tax schedule. The utility function is assumed to be of the iso-elastic form

$$u(c,l) = c - \frac{l^{1+1/\varepsilon}}{1+1/\varepsilon}, \tag{36}$$

where ε is the constant elasticity of labor supply. The latter is set at a value $\varepsilon = 0.33$, as suggested by Chetty (2012). I approximate the current tax schedule using a linear specification

$$T(z(n)) = -q + tz(n). \tag{37}$$

Values for the lump-sum transfer q and the constant marginal tax rate t are obtained by regressing the tax liability on taxable income, see, e.g., Saez (2001). This gives $q = \$4,590$ and $t = 33.1\%$ with an R^2 of approximately 0.94. Figure 4 in Appendix XIII plots the actual and fitted values for incomes up to $500,000. The question how q and t should be optimized is taken up in Appendix XII, which sets up and solves the optimal linear tax problem.

---

[39] If labor income at the top follows a Pareto distribution with tail parameter $\tilde{a}$, the expected value of income above a certain amount z' equals $\mathbb{E}[z|z \geq z'] = \left(\frac{\tilde{a}}{\tilde{a}-1}\right) z'$.

5.1.3 Equilibrium

If the utility function is iso-elastic and the tax function is linear, it is straightforward to derive the equilibrium (cf. Definition 2). Labor effort follows from equation (4):

$$l(n) = (1-t)^{\varepsilon} n^{\varepsilon}. \tag{38}$$

Labor earnings, in turn, are obtained by substituting labor effort in equation (7) and using the definition $\pi(n) = nl(n) - z(n)$. This gives

$$z(n) = \left(1 - \frac{\mu}{1+\varepsilon}\right)(1-t)^{\varepsilon} n^{1+\varepsilon} + \left(\frac{\mu}{1+\varepsilon}\right) z(n_0). \tag{39}$$

An individual's labor income equals a weighted average of the output she produces (first term) and the labor income of the individuals with the lowest ability (second term).[40] The profits $\pi(n) = nl(n) - z(n)$ firms generate from hiring a worker with ability n are given by

$$\pi(n) = \left(\frac{\mu}{1+\varepsilon-\mu}\right)(z(n) - z(n_0)). \tag{40}$$

Equations (39) and (40) give a mapping from (observable) labor income to (unobservable) ability and pure economic profits, respectively.

With this closed-form characterization of the equilibrium, a few remarks are in place. First, as in the classic and new monopsony models introduced in Robinson (1933) and Manning (2003), the mark-up of productivity over wages (or output over earnings) is decreasing in the elasticity of labor supply. To see this, denote by $w(n) = z(n)/l(n)$ the hourly wage of an individual with ability n and assume $z(n_0)$ is very small, as in the data. Using equation (39), the mark-up, i.e., the measure of "exploitation" introduced by Pigou (1920), is

$$\frac{n - w(n)}{w(n)} = \frac{\mu}{1+\varepsilon-\mu}. \tag{41}$$

Clearly, the latter is increasing in monopsony power μ and decreasing in the elasticity of labor supply ε. Second, equation (39) implies that if firms have monposony power, productivity gains (captured by an increase in ability n) are not translated one-for-one into higher wages. This is a standard prediction from models where firms have monopsony power that is supported by empirical evidence (see Kline et al. (2019) for a recent example). Third, from equation (39) it is clear that monopsony power mitigates inequality in labor earnings driven by differences in ability. Despite this, monposony power has no impact on typical measures of inequality in labor earnings, such as the Gini coefficient, the variance in log earnings or the P90/P10 earnings ratio. The reason is that monopsony power simply scales down labor earnings for this particular choice of the utility and tax function. In the more general case

---

[40] The reason why the lowest income shows up in equation (39) is that, by assumption, firms make no profits from hiring individuals with the lowest ability: $\pi(n_0) = 0$. This can only be the case for *any* degree of monposony power if individuals with ability n_0 are indifferent between working and not working. Therefore, the lowest income level is informative about the outside option of non-employment. Note that the value of non-employment generally differs from the lump-sum transfer g, for example because non-employed individuals are entitled to an additional benefit or because of (non-modeled) utility costs or benefits of having a job.

where monopsony power, the marginal tax rate or the elasticity of labor supply vary with ability, the model does not make a clear-cut prediction on the impact of monopsony power on these measures of inequality.[41]

5.1.4 Monopsony power

Monopsony power μ determines how much pure economic, or above-normal profits firms make. In recent work, Barkai and Benzell (2018) and Barkai (2020) decompose US output into a labor share, a capital share and a profit share. The labor share is calculated as total compensation to employees as a fraction of gross value added. The capital share, in turn, is calculated as the product of the capital stock and the required (or normal) rate of return, again as a fraction of gross value added. The remainder, i.e., the profit share, is a measure of pure economic profits. Because my model abstracts from productive capital, I calibrate monopsony power μ to target the ratio of aggregate profits to aggregate labor income, or the ratio of the *profit share* to the *labor share*. For the most recent year 2015, Barkai and Benzell (2018) calculate that the ratio of aggregate profits to aggregate wages is approximately 24.2%. Using their estimate, the value for monopsony power μ can be calculated by integrating equation (40) over the ability distribution and dividing by aggregate labor income $\bar{z} = \int_{n_0}^{n_1} z(n)f(n)dn$. This gives

$$\left(\frac{\bar{\pi}}{\bar{z}}\right) = \left(\frac{\mu}{1+\varepsilon-\mu}\right)\left(1 - \left(\frac{z(n_0)}{\bar{z}}\right)\right) \quad \Leftrightarrow \quad \mu = (1+\varepsilon)\left[\frac{(\bar{\pi}/\bar{z})}{1+(\bar{\pi}/\bar{z})-(z(n_0)/\bar{z})}\right]. \tag{42}$$

Substituting out for the elasticity of labor supply and the ratio of profits to wages gives a value for monopsony power of approximately $\mu = 0.26$.[42]

5.1.5 Ability distribution

As in Saez (2001), I calibrate the ability distribution to match the empirical income distribution. To do so, I use equation (39) and calculate the ability n for each individual with positive labor earnings. This gives an empirical counterpart of the ability distribution $F(n)$. I subsequently smooth this distribution by estimating a kernel density. The empirical distribution and the kernel density are plotted in the top panel of Figure 5 in Appendix XIII. The bottom panel plots the distribution of labor earnings and the implied kernel density.

I make one adjustment to the density as plotted in the top panel of Figure 5. In particular, I append a right Pareto tail starting at an ability level associated with $350,000 in annual earnings. The reason for doing so is that individuals with very high labor earnings are significantly under-represented in the CPS data. I choose the tail parameter of the *ability* distribution to be consistent with a tail parameter of 1.6 of the *labor income* distribution at the top.[43] This

---

[41] This could also explain why Webber (2015) and Rinz (2018) find a positive association between measures of monopsony power and the variance in log earnings or the P90/P10 earnings ratio, respectively.

[42] In the CPS data, the lowest earnings level is very small compared to average earnings. Hence, the choice of $z(n_0)/\bar{z}$ only has a small effect on the calibrated value of μ.

[43] Let $\tilde{F}(z(n))$ denote the labor income distribution with density $\tilde{f}(z(n))$. Monotonicity of labor earnings implies $F(n) = \tilde{F}(z(n))$ for all n where $z(n) > 0$ and hence, $f(n) = \tilde{f}(z(n))z'(n)$. The local Pareto parameter of the ability distribution $a(n) = nf(n)/(1-F(n))$ and income distribution $\tilde{a}(z(n)) = z(n)\tilde{f}(z(n))/(1-\tilde{F}(z(n)))$

is the estimate obtained by Saez and Stantcheva (2018) using tax returns data. The scale parameter of the Pareto distribution is set to ensure there is no jump in the density at the point where the Pareto tail is pasted.

5.1.6 Profit taxation and revenue requirement

In the model, there is no productive capital and τ is the rate at which pure economic, or above-normal profits are taxed. The current tax system does not distinguish between normal and above-normal returns. I therefore assume all capital income is taxed at a rate $\tau = 36\%$, taken from Trabandt and Uhlig (2011). This figure is very similar to the one that is obtained if the government levies a corporate tax rate of 21% at the firm level and a capital gains tax rate of 20% at the individual level. For a given value of τ, the government's budget constraint (10) can be used to calculate the revenue requirement. This gives $G = \$22,049$, which in the calibrated economy corresponds to approximately 28.6% of aggregate output. Table 1 summarizes the calibration strategy.

Variable	Target	Source	Value
μ	Aggregate profits over wages	Barkai and Benzell (2018)	0.26
ε	Elasticity of labor supply	Chetty (2012)	0.33
τ	Tax rate on capital income	Trabandt and Uhlig (2011)	0.36
G	Government budget constraint	Equilibrium condition	\$22,049
$T(z)$	Tax liability	CPS 2018	Figure 4
$F(n)$	Income distribution	CPS 2018	Figure 5

Table 1: Calibration

5.2 Welfare function

The welfare function (9) depends on the average welfare weights $g(n)$ of individuals with the same ability and the negative covariance $\Sigma \in [0, 1]$ between welfare weights and shareholdings. The first (second) determines how much the government values reducing inequality generated by differences in ability (shareholdings). In the remainder, I let Σ vary between zero and one. If $\Sigma = 0$, the government does not value redistributing capital income. Conversely, if $\Sigma = 1$, the government cares a lot about redistributing capital income as all shares are held by individuals with a welfare weight of zero. Regarding the average welfare weights of individuals with the same ability, I use the following specification:

$$g(n) = \zeta n^{-\beta}. \tag{43}$$

Here, $\zeta > 0$ is a scaling parameter and $\beta \geq 0$ governs how much the government wishes to redistribute from individuals with high to individuals with low ability. If $\beta = 0$, the government

---

are related through $a(n) = \tilde{a}(z(n))e_{zn}$, where $e_{zn} = z'(n)n/z(n)$ is the elasticity of labor earnings with respect to ability. The latter equals approximately $1 + \varepsilon$ at high levels of labor earnings: see equation (39).

attaches the same average weight to individuals of all ability levels. Conversely, if $\beta \to \infty$, the government only cares about individuals with the lowest ability.

Before selecting a value for ζ and β, I make one adjustment to the welfare function (9). In particular, I assume there is a mass of $\nu = 0.05$ non-participants, who earn zero labor and capital income and whose welfare weight equals twice the average welfare weight of all other individuals. The government optimizes a benefit for the non-participants, subject to the requirement that their utility does not exceed the labor market payoff of individuals with the lowest ability. Under these conditions, the optimal marginal tax rate at the bottom of the income distribution is positive even if labor markets are competitive or there is no (desire to reduce) capital income inequality: $(1-\tau)\Sigma = 0$. This avoids technical difficulties associated with steeply increasing marginal tax rates at very low earnings.[44] The parameter ζ is set to make sure the average welfare weight of all individuals (including the non-participants) equals one. Moreover, I choose the value of β such that the average marginal tax rate at the optimal tax system with competitive labor markets equals the current rate $t = 33.1\%$.

5.3 Optimal marginal tax rates

Figure 2 plots optimal marginal tax rates for different assumptions on the degree of monopsony power μ and the negative covariance between welfare weights and shareholdings Σ. To facilitate the comparison, the horizontal axis shows *current* labor earnings. The red, solid line plots the marginal tax rates a "naive" government would set that acts *as if* labor markets are competitive. The tax rates are calculated by substituting $\mu = 0$ in equation (14). Consistent with the calibrated value of β, the average marginal tax rate equals 33.1%. The conventional U-shape pattern (see, e.g., Diamond (1998) and Saez (2001)) follows from the behavior of the local Pareto parameter $a(n)$: see Figure 6 in Appendix XIII.

The blue, dashed line in Figure 2 plots the optimal marginal tax rates if the degree of monopsony power is as in the calibrated economy and the government does not value redistributing capital income: $\mu = 0.26$ and $\Sigma = 0$. Compared to the case with competitive labor markets, optimal marginal tax rates are lower, cf. Corollary 1. This is because monopsony power makes labor income taxes less effective in redistributing labor income as part of the incidence falls on firms. The average reduction in optimal marginal tax rates brought about by monopsony power is 5.6 percentage points.

The black, dotted line plots the optimal marginal tax rates if the government has a very strong preference for redistributing capital income: $\Sigma = 1$. Naturally, tax rates are higher compared to the case with $\Sigma = 0$. The average increase brought about by a change in the covariance between welfare weights and shareholdings is 13.2 percentage points. Compared to the case with competitive labor markets, optimal marginal tax rates are higher (lower) for individuals whose current labor earnings are below (above) approximately $122,000. On average, the optimal marginal tax rate with monopsony power is 7.7 percentage points higher. The increase is driven mostly by substantially higher marginal tax rates at low earnings lev-

---

[44] These difficulties arise because a low value of the local Pareto parameter $a(n)$ at the bottom implies the optimal marginal tax rate jumps from $T'(z(n_0)) = 0$ immediately to a high value. Such a jump often leads to a violation of the monotonicity condition: see Appendix III.

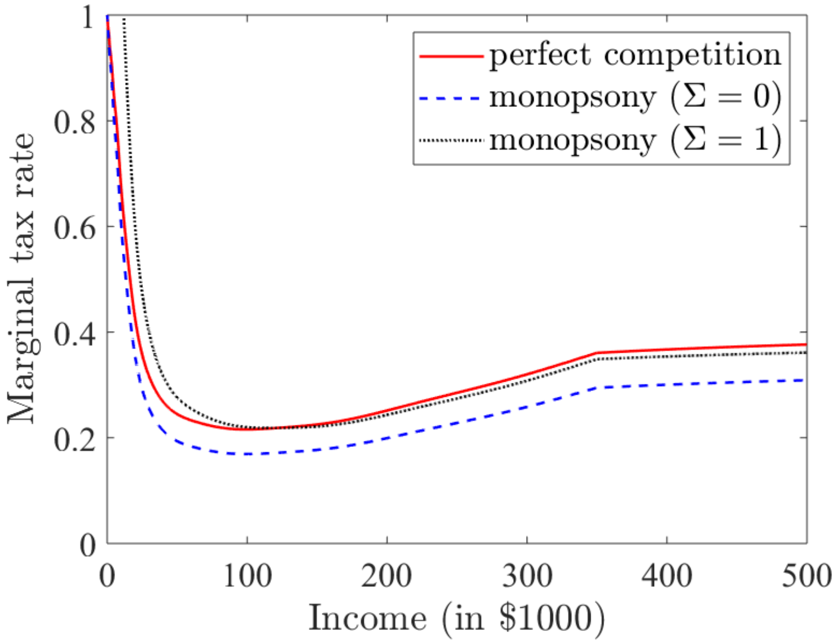


Figure 2: Optimal marginal tax rates

els, where the local Pareto parameter $a(n)$ is low: see Corollary 1 and Figure 6. The low Pareto parameter at the bottom also implies that some individuals do not work at the optimal allocation, as the constraint $l(n) \geq 0$ is binding. This is the case for individuals whose current labor earnings are below approximately $12,000.

According to Corollary 1, the impact of monopsony power on optimal tax rates is generally ambiguous. The analysis here suggests that if the government wishes to reduce inequality generated by differences in *both* ability and shareholdings, monopsony power tends to increase optimal marginal tax rates at lower earnings levels and to decrease optimal marginal tax rates at higher earnings levels. At what earnings level the impact changes from positive to negative depends critically on the covariance between welfare weights and shareholdings.

5.4 Implications for welfare

To assess the quantitative implications of monopsony power for welfare in the calibrated economy, I conduct two exercises. First, I calculate the welfare costs of ignoring monopsony power when designing tax policy. To do so, I compare the allocation that is obtained if the government sets income taxes optimally (cf. the dashed and dotted lines in Figure 2) with the one that is obtained if a "naive" government wrongfully sets tax policy *as if* labor markets are competitive (cf. the solid line in Figure 2). Second, I calculate how much the government is willing to pay for changing the degree of monopsony power to zero. The first exercise gives an indication of the welfare benefits of taking a *given* degree of monopsony power into account when designing tax policy, whereas the second exercise is informative about the costs or benefits of *changing* the degree of monopsony power.

Figure 3 shows the results of both exercises for different values of the covariance between welfare weights and shareholdings. The left axis plots the welfare costs of ignoring monopsony power when designing tax policy (i.e., the costs of "misoptimization"). The right axis plots the welfare effect of changing the degree of monopsony power from its value in the calibrated economy to zero. In both cases, the welfare impact is expressed in consumption equivalents as a percentage of current GDP in the calibrated economy. Regarding the first exercise, the welfare costs of ignoring monopsony power when designing tax policy range between \$57 and \$802 in consumption equivalents, or between 0.07% and 1.04% of GDP. These costs are small for low values of the negative covariance between welfare weights and shareholdings and largest if the government has a strong preference for redistributing capital income. To illustrate, moving from the solid to the dashed tax code plotted in Figure 2 generates a welfare gain equivalent to increasing all individuals' net income by \$70, or 0.09% of GDP. By contrast, moving from the solid to the dotted tax code plotted in Figure 2 generates a welfare gain equivalent to increasing all individuals' net income by \$802, or 1.04% of GDP.

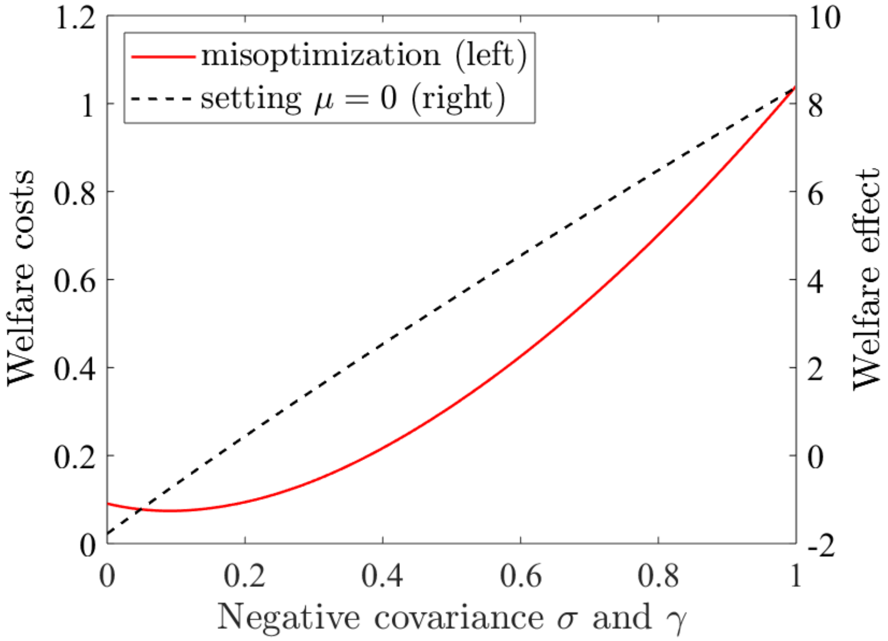


Figure 3: Welfare impact in consumption equivalents (% of GDP)

Regarding the second exercise, changing the degree of monopsony power from its value in the calibrated economy to zero can have a negative or positive impact on welfare depending on the covariance between shareholdings and welfare weights. If $\Sigma = 0$, getting rid of monopsony power leads to a welfare loss of \$1,370 in consumption equivalents, or 1.78% of GDP. This loss occurs because a reduction in monopsony power exacerbates labor income inequality and the government does not value the associated reduction in capital income inequality. By contrast, the welfare impact is positive if the government cares about redistributing capital income. In the calibrated economy, this happens whenever $\Sigma \geq 0.17$. If $\Sigma = 1$, the welfare gain of firms losing monopsony power is large and equals \$6,453 in consumption equivalents, or 8.37% of GDP.

The previous exercise illustrates that changing the degree of monopsony power from its value in the calibrated economy to zero and simultaneously re-optimizing the tax code can have a large negative or positive impact on welfare. It is also possible to analyze the welfare effect of a marginal increase in monopsony power at the current tax system provided the latter reflects the government's preferences for redistribution. Because labor effort is increasing in ability (see equation (38)), the result from Corollary 2 applies. Hence, an increase in monopsony power raises welfare only if

$$\left(\frac{\Sigma^{\ell}}{\Sigma^{k}}\right) > \left(\frac{1-\mu}{\mu}\right). \tag{44}$$

In the calibrated economy, the right-hand side equals approximately 2.85. Hence, if the current tax system is optimal, an increase in monopsony power raises welfare only if the negative covariance between welfare weights and after-tax labor income exceeds the negative covariance between welfare weights and after-tax capital income by a factor of at least this amount. If the preferences for redistribution are such that this condition is not satisfied at the current tax system, an increase in monopsony power lowers welfare.

To summarize, correcting the sub-optimal tax code by taking monopsony power into account leads to welfare gains that vary between 0.07% and 1.04% of current GDP in the calibrated economy. Moreover, changing the degree of monopsony power to zero has a welfare impact that ranges between –1.78% to +8.37% of GDP depending on the covariance between welfare weights and shareholdings. Finally, if the current tax system is optimal, an increase in monopsony power raises welfare only if the negative covariance between welfare weights and labor income exceeds the negative covariance between welfare weights and capital income by a factor of at least 2.85.

6 Conclusion

This paper extends the non-linear tax framework of Mirrlees (1971) with monopsony power and studies the implications for optimal income taxation and welfare.

Monopsony power makes labor income taxes less effective in redistributing labor income, but more effective in redistributing profits, i.e., capital income. This is because monopsony power raises the tax incidence that falls on firms and lowers the tax incidence that falls on workers. The impact of monopsony power on optimal marginal tax rates is ambiguous and depends on the government's preference for redistribution. In the typical case where the government wishes to redistribute both labor and capital income, optimal marginal tax rates are higher (lower) at low (high) levels of labor earnings. In that sense, monopsony power makes the optimal tax system less progressive. A calibration exercise to the US economy suggests that the welfare costs of ignoring monopsony power range between 0.07% and 1.04% of GDP depending on the covariance between welfare weights and shareholdings.

Monopsony power has an ambiguous effect on welfare, as it reduces inequality in labor market payoffs (i.e., after-tax labor earnings minus the disbuility of working) but increases inequality in capital income. The reason why monopsony power might raise welfare is that firms observe ability, while the government does not. Monopsony power reduces inequality generated by differences in ability. This alleviates the trade-off between equity and efficiency that occurs because the government does not observe ability, but at the expense of increasing capital income inequality. In the calibrated economy, eliminating monopsony power has a welfare effect that ranges between –1.78% and +8.37% of GDP depending on the covariance between welfare weights and shareholdings. Moreover, if the current tax system is optimal, an increase in monopsony power raises welfare only if the negative covariance between welfare weights and after-tax labor income is at least 2.85 times as high as the negative covariance between welfare weights and after-tax capital income.

If monopsony power generates distortions in employment by lowering the payoff from working, it becomes less likely that an increase in monopsony power raises welfare. Optimal marginal tax rates are then reduced in order to partly alleviate the distortions from monopsony power and stimulate labor participation. Furthermore, the welfare impact of raising monopsony power is lower if firms engage in costly profit shifting.

In order to study the implications of monopsony power for optimal income taxation and welfare in a tractable way, this paper has abstracted from a number of dimensions. First, there is no productive capital or wealth accumulation and all capital income consists of pure economic rents. In reality, capital can raise labor productivity and part of the income it generates are normal returns (e.g., the return for postponing consumption). Adding these features most likely reduces the optimal tax on capital income (cf. the findings from Atkinson and Stiglitz (1976), Judd (1985) and Chamley (1986)), but does not fundamentally alter the insight that (i) taxes on labor earnings can be used to indirectly tax profits if firms have monopsony power and (ii) monopsony power has an ambiguous impact on welfare. Second, there is no human capital accumulation: ability is exogenous. If, realistically, individuals can invest in their skills (e.g., through formal education), monopsony power adversely affects the incentives to do so. As with a distortion in labor participation, a distortion in human capital accumulation could call for lower tax rates on labor earnings and reduce the welfare impact of raising monopsony power.[45] Third, there is no meaningful role for firm heterogeneity: firms operate an identical technology and are matched exogenously with heterogeneous workers (who may suffer more or less from monopsony). Firm heterogeneity plays an important role in explaining wage differences (see, e.g., Abowd et al. (1999) and Song et al. (2019)) and firm size could be a source of monopsony power. It would be challenging, but very interesting to incorporate firm heterogeneity and endogenize the degree of monopsony power firms have. I leave this as an extension for future research.

References

Abowd, J., F. Kramarz, and D. Margolis (1999). High wage workers and high wage firms. *Econometrica 67*(2), 251–333.

Arnold, D. (2020). Mergers and acquisitions, local labor market concentration, and worker outcomes. mimeo, Princeton University.

---

[45] These effects would be mitigated if monopsony power leads firms to invest *more* in their workers' skills, e.g., through job training programs.

Atkinson, A. and J. Stiglitz (1976). The design of tax structure: direct versus indirect taxation. *Journal of Public Economics 6*(1-2), 55–75.

Azar, J., I. Marinescu, and M. Steinbaum (2019). Measuring labor market power two ways. *AEA Papers and Proceedings 109*, 317–321.

Azar, J., I. Marinescu, and M. Steinbaum (2020). Labor market concentration. *Journal of Human Resources*. Forthcoming.

Azar, J., I. Marinescu, M. Steinbaum, and B. Taska (2018). Concentration in US labor markets: evidence from online vacancy data. NBER Working Paper No. 24395, Cambridge-MA: NBER.

Barkai, S. (2020). Declining labor and capital shares. *Journal of Finance*. Forthcoming.

Barkai, S. and S. Benzell (2018). 70 years of US corporate profits. Stigler Center Working Paper No. 22, Chicago: Stigler Center.

Bastani, S., T. Blumkin, and L. Micheletto (2015). Optimal wage redistribution in the presence of adverse selection in the labor market. *Journal of Public Economics 131*, 41–57.

Benmelech, E., N. Bergman, and H. Kim (2018). Strong employers and weak employees: how does employer concentration affect wages? NBER Working Paper No. 24307, Cambridge-MA: NBER.

Bourguignon, F. and A. Spadaro (2012). Tax–benefit revealed social preferences. *Journal of Economic Inequality 10*(1), 75–108.

Cahuc, P. and G. Laroque (2014). Optimal taxation and monopsonistic labor market: does monopsony justify the minimum wage? *Journal of Public Economic Theory 16*(2), 259–273.

Caldwell, S. and O. Danieli (2018). Outside options in the labor market. mimeo, Harvard University.

CEA (2016). Labor market monopsony: trends, consequences, and policy responses. Issue Brief, White House Council of Economic Advisers, Washington, DC.

Chamley, C. (1986). Optimal taxation of capital income in general equilibrium with infinite lives. *Econometrica 53*(3), 607–622.

Chetty, R. (2009). Sufficient statistics for welfare analysis: a bridge between structural and reduced-form methods. *Annual Review of Economics 1*(1), 451–488.

Chetty, R. (2012). Bounds on elasticities with optimization frictions: a synthesis of micro and macro evidence on labor supply. *Econometrica 80*(3), 969–1018.

Choné, P. and G. Laroque (2011). Optimal taxation in the extensive model. *Journal of Economic Theory 146*(2), 425–453.

Craig, A. (2020). Optimal income taxation with spillovers from employer learning. mimeo, University of Michigan.

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

Diamond, P. (1980). Income taxation with fixed hours of work. *Journal of Public Economics 13*(1), 101–110.

Diamond, P. (1998). Optimal income taxation: an example with a U-shaped pattern of optimal marginal tax rates. *American Economic Review 88*(1), 83–95.

Djankov, S., T. Ganser, C. McLiesh, R. Ramalho, and A. Shleifer (2010). The effect of corporate taxes on investment and entrepreneurship. *American Economic Journal: Macroeconomics 2*(3), 31–64.

FTC (2018a). Hearings on competition and consumer protection in the 21st century, no. 2: monopsony and the state of US antitrust law. Federal Trade Commission, Washington, DC.

FTC (2018b). Hearings on competition and consumer protection in the 21st century, no. 3: multi-sided platforms, labor markets and potential competition. Federal Trade Commission, Washington, DC.

Golosov, M., A. Tsyvisnki, and N. Werquin (2014). A variational approach to the analysis of tax systems. NBER Working Paper No. 20780, Cambridge-MA: NBER.

Hansen, E. (2019). Optimal income taxation with labor supply responses at two margins: when is an Earned Income Tax Credit optimal? mimeo, University of Cologne.

Hariton, C. and G. Piaser (2007). When redistribution leads to regressive taxation. *Journal of Public Economic Theory 9*(4), 589–606.

Haufler, A. and G. Schjelderup (2000). Corporate tax systems and cross country profit shifting. *Oxford Economic Papers 52*(2), 306–325.

Hershbein, B., C. Macaluso, and C. Yeh (2019). Concentration in local US labor markets: evidence from vacancy and employment data. mimeo, W.E. Upjohn Institute.

Hines, J. and E. Rice (1994). Fiscal paradise: Foreign tax havens and American business. *Quarterly Journal of Economics 109*(1), 149–182.

Huizinga, H. and L. Laeven (2008). International profit shifting within multinationals: A multi-country perspective. *Journal of Public Economics 92*(5-6), 1164–1182.

Jacobs, B. (2018). The marginal cost of public funds is one at the optimal tax system. *International Tax and Public Finance 25*(4), 883–912.

Jacobs, B., E. Jongen, and F. Zoutman (2017). Revealed social preferences of Dutch political parties. *Journal of Public Economics 156*, 81–100.

Jacquet, L., E. Lehmann, and B. Van der Linden (2013). Optimal redistributive taxation with both extensive and intensive responses. *Journal of Economic Theory 148*(5), 1770–1805.

Judd, K. (1985). Redistributive taxation in a simple perfect foresight model. *Journal of Public Economics 28*(1), 59–83.

Kaplow, L. (2019). Market power and income taxation. NBER Working Paper No. 25578, Cambridge-MA: NBER.

Kline, P., N. Petkova, H. Williams, and O. Zidar (2019). Who profits from patents? Rent-sharing at innovative firms. *Quarterly Journal of Economics 134*(3), 1343–1404.

Krueger, A. (2018). Reflections on dwindling worker bargaining power and monetary policy. mimeo, Luncheon Address at the Jackson Hole Economic Symposium.

Lipsius, B. (2018). Labor market concentration does not explain the falling labor share. mimeo, University of Michigan.

Manning, A. (2003). *Monopsony in motion: imperfect competition in labor markets.* Princeton: Princeton University Press.

Mirrlees, J. (1971). An exploration in the theory of optimum income taxation. *Review of Economic Studies 38*(114), 175–208.

Pigou, A. (1920). *The economics of welfare.* London: Palgrave Macmillan UK.

Qiu, Y. and A. Sojourner (2019). Labor-market concentration and labor compensation. IZA Discussion Paper No. 12089, Bonn: IZA.

Rinz, K. (2018). Labor market concentration, earnings inequality and earnings mobility. CARRA Working Paper No. 2018-10, Washington, DC: US Census Bureau.

Robinson, J. (1933). *The economics of imperfect competition.* London: Palgrave Macmillan UK.

Saez, E. (2001). Using elasticities to derive optimal income tax rates. *Review of Economic Studies 68*(1), 205–229.

Saez, E. (2002). Optimal income transfer programs: intensive versus extensive labor supply responses. *Quarterly Journal of Economics 117*(3), 1039–1073.

Saez, E., B. Schoefer, and D. Seim (2019). Payroll taxes, firm behavior, and rent sharing: evidence from a young workers' tax cut in Sweden. *American Economic Review 109*(5), 1717–63.

Saez, E. and S. Stantcheva (2018). A simpler theory of optimal capital taxation. *Journal of Public Economics 162*, 120–142.

Sandmo, A. (1994). Monopsonistic wage discrimination, incentives and efficiency. *Labour Economics 1*(2), 151–170.

Schubert, G., A. Stansbury, and B. Taska (2020). Employer concentration and outside options. mimeo, Harvard University.

Seade, J. (1977). On the shape of optimal tax schedules. *Journal of Public Economics 7*(2), 203–235.

Song, J., D. Price, F. Guvenen, N. Bloom, and T. Von Wachter (2019). Firming up inequality. *Quarterly Journal of Economics 134*(1), 1–50.

Stantcheva, S. (2014). Optimal income taxation with adverse selection in the labour market. *Review of Economic Studies 81*(3), 1296–1329.

Thomson, W. (1994). Cooperative models of bargaining. In R. Aumann and S. Hart (Eds.), *Handbook of Game Theory with Economic Applications*, Volume 2, Chapter 35, pp. 1237–1284. Amsterdam: Elsevier.

Thoresson, A. (2020). Employer concentration and wages for specialised workers. mimeo, Uppsala University.

Tørsløv, T., L. Wier, and G. Zucman (2020). The missing profits of nations. NBER Working Paper No. 24701, Cambridge-MA: NBER.

Trabandt, M. and H. Uhlig (2011). The Laffer curve revisited. *Journal of Monetary Economics 58*(4), 305–327.

Webber, D. (2015). Firm market power and the earnings distribution. *Labour Economics 35*, 123–134.

Appendix

I Rewriting the welfare function

The result from Lemma 1 can be obtained as follows. Substitute the utility function (2) in the welfare function (8) and rewrite the resulting expression in a number of steps:

$$\begin{aligned}
\mathcal{W} &= \int_{\sigma_0}^{\sigma_1}\int_{n_0}^{n_1} \gamma(n,\sigma)\mathcal{U}(n,\sigma)h(n,\sigma)dnd\sigma \\
&= \int_{\sigma_0}^{\sigma_1}\int_{n_0}^{n_1} \gamma(n,\sigma)\Big[\upsilon(n)+\sigma(1-\tau)\bar{\pi}\Big]h(n,\sigma)dnd\sigma \\
&= \int_{n_0}^{n_1} \upsilon(n)\underbrace{\left(\int_{\sigma_0}^{\sigma_1}\gamma(n,\sigma)h(n,\sigma)d\sigma\right)}_{=g(n)f(n)}dn + (1-\tau)\bar{\pi}\int_{\sigma_0}^{\sigma_1}\int_{n_0}^{n_1}\sigma\gamma(n,\sigma)h(n,\sigma)dnd\sigma \\
&= \int_{n_0}^{n_1} g(n)\upsilon(n)f(n)dn + (1-\tau)\bar{\pi}\Bigg(1+\underbrace{\int_{\sigma_0}^{\sigma_1}\int_{n_0}^{n_1}(\sigma-1)(\gamma(n,\sigma)-1)h(n,\sigma)dnd\sigma}_{=\mathrm{Cov}[\boldsymbol{\sigma},\boldsymbol{\gamma}]=-\Sigma}\Bigg) \\
&= \int_{n_0}^{n_1} g(n)\upsilon(n)f(n)dn + (1-\tau)(1-\Sigma)\int_{n_0}^{n_1}\pi(n)f(n)dn, \qquad (45)
\end{aligned}$$

which corresponds to equation (9). To show that $\Sigma \in [0,1]$, write

$$\Sigma = -\int_{\sigma_0}^{\sigma_1}\int_{n_0}^{n_1}(\sigma-1)(\gamma(n,\sigma)-1)h(n,\sigma)dnd\sigma = 1-\int_{\sigma_0}^{\sigma_1}\int_{n_0}^{n_1}\sigma\gamma(n,\sigma)h(n,\sigma)dnd\sigma. \tag{46}$$

Given that $\sigma \geq 0$ and $\gamma(n,\sigma)\geq 0$, it follows that $\Sigma \leq 1$. Next, write the covariance as

$$\begin{aligned}\Sigma &= \int_{\sigma_0}^{\sigma_1}\int_{n_0}^{n_1}(1-\sigma)\gamma(n,\sigma)h(n,\sigma)dnd\sigma = \int_{\sigma_0}^{\sigma_1}(1-\sigma)\int_{n_0}^{n_1}\gamma(n,\sigma)h(n,\sigma)dnd\sigma \\ &= \int_{\sigma_0}^{\sigma_1}(1-\sigma)\underbrace{\left(\frac{\int_{n_0}^{n_1}\gamma(n,\sigma)h(n,\sigma)dn}{\int_{n_0}^{n_1}h(n,\sigma)dn}\right)}_{=\mathbb{E}[\gamma(n,\sigma)|\sigma]}\underbrace{\left(\int_{n_0}^{n_1}h(n,\sigma)dn\right)}_{=\hat{h}(\sigma)}d\sigma.\end{aligned} \tag{47}$$

By assumption, $\mathbb{E}[\gamma(n,\sigma)|\sigma]$ is non-increasing and averages to one. Therefore,

$$\Sigma = \int_{\sigma_0}^{\sigma_1}(1-\sigma)\mathbb{E}[\gamma(n,\sigma)|\sigma]\hat{h}(\sigma)d\sigma \geq \int_{\sigma_0}^{\sigma_1}(1-\sigma)\hat{h}(\sigma)d\sigma = 0. \tag{48}$$

II Optimal tax problem

To solve the optimal tax problem, I follow the approach from Mirrlees (1971) and let the government choose the allocation variables to maximize welfare (9) subject to resource and incentive constraints. The allocation variables are labor effort $l(n)$, the labor market payoff $v(n)$ and the profits $\pi(n)$ firms make from hiring a worker with ability n. To derive the resource constraint in terms of the allocation variables, substitute $T(z(n)) = z(n) - v(n) - \phi(l(n)) = nl(n) - \pi(n) - v(n) - \phi(l(n))$ in the government's budget constraint (10) and rearrange to find

$$\int_{n_0}^{n_1} nl(n)f(n)dn = \int_{n_0}^{n_1}\Big[v(n)+\phi(l(n))+(1-\tau)\pi(n)\Big]f(n)dn + G. \tag{49}$$

In words, aggregate output equals the sum of private consumption (first term) and public consumption (second term).

In addition to the resource constraint, the allocation must also satisfy incentive constraints. To derive the first of these, differentiate the labor market payoff $v(n) = z(n) - T(z(n)) - \phi(l(n))$ with respect to ability to find

$$v'(n) = (1-T'(z(n)))z'(n) - \phi'(l(n))l'(n). \tag{50}$$

Next, use the first-order condition from the profit maximization problem (4) and the relationship $\pi(n) = nl(n) - z(n)$. Condition (50) can then be written as

$$v'(n) = \frac{\phi'(l(n))}{n}\Big[l(n) - \pi'(n)\Big]. \tag{51}$$

This condition differs from the incentive constraint in the Mirrlees (1971) problem through the occurrence of the term $\pi'(n)$, which is zero if labor markets are competitive. The labor market payoff increases less quickly in ability if firms generate more profits from hiring individuals with higher ability.

To derive the second incentive constraint, differentiate the condition for profits (7) with respect to ability to find

$$\pi'(n) = \mu(n)l(n) + \frac{\mu'(n)}{\mu(n)}\pi(n). \tag{52}$$

Intuitively, profits increase more rapidly in ability the higher is monopsony power and labor effort. Profits increase less quickly in ability if individuals with higher ability suffer less from monopsony (i.e., if $\mu'(n) < 0$). Combining equations (51) and (52) gives

$$\upsilon'(n) = \frac{\phi'(l(n))}{n}\left[(1-\mu(n))l(n) - \frac{\mu'(n)}{\mu(n)}\pi(n)\right]. \tag{53}$$

As stated in the main text, I assume $\mu'(n)$ is bounded from above in such a way that the labor market payoff is weakly increasing in ability: $\upsilon'(n) \geq 0$.[46] The labor market payoff does not vary with ability if firms have full monopsony power (i.e., if $\mu(n) = 1$ for all n). In that case, all individuals are put on their identical participation constraint and hence, $\upsilon'(n) = 0$.

The government's problem consists of choosing the allocation variables $\upsilon(n)$, $\pi(n)$ and $l(n)$ at each ability level n to maximize welfare (9), subject to the resource constraint (49) and the incentive constraints (52) – (53). As it turns out, it is important to take the non-negativity constraint $l(n) \geq 0$ explicitly into account.[47] The final restriction we need to impose is that the profits from hiring the least productive workers are non-negative: $\pi(n_0) \geq 0$. This condition guarantees that firms are willing to hire individuals of all ability levels.[48] It is shown in Appendix V that this constraint is always binding, which *ex post* validates the assumption that $\pi(n_0) = 0$ in the description of the equilibrium: see Definition 2 and equation (7). The optimal tax problem can now be formulated as a standard optimal control problem where $\upsilon(n)$ and $\pi(n)$ are the state variables and $l(n)$ is the control variable. The corresponding Lagrangian and first-order conditions can be found in Appendix IV.

To make sure that the optimal allocation (as implicitly characterized in Appendix IV) can be decentralized using a tax on profits τ and a non-linear tax on labor income $T(z(n))$, I assume that earnings $z(n) = nl(n) - \pi(n)$ are increasing in ability whenever the non-negativity constraint on labor effort is not binding: $z'(n) > 0$ if $l(n) > 0$. This condition serves two purposes. First, it guarantees that individuals with different abilities do not earn the same income and hence, are not required to face the same marginal tax rate. Second, the monotonicity condition also ensures that the second-order condition for profit maximization is satisfied – see Appendix III for details.

---

[46] From equation (53), it follows that this is the case if $\mu'(n) \leq \mu(n)(1-\mu(n))l(n)/\pi(n)$ provided $\pi(n) > 0$. This condition always holds if monopsony power does not vary with ability.

[47] To ensure consumption is non-negative, one could also include the constraint $\upsilon(n) + \phi(l(n)) \geq 0$ for all n. I assume the revenue requirement G and preferences for redistribution are such that this constraint never binds.

[48] To see why, note that the general solution to the differential equation (52) is

$$\pi(n) = \mu(n)\left[\frac{\pi(n_0)}{\mu(n_0)} + \int_{n_0}^{n} l(m)dm\right],$$

which simplifies to equation (7) if $\pi(n_0) = 0$. Because labor effort is non-negative, it follows that $\pi(n_0) \geq 0$ implies $\pi(n) \geq 0$ for all n.

III Monotonicity condition

This Appendix demonstrates the equivalence between the monotonicity condition $z'(n) > 0$ and the requirement that the second order-condition for the profit maximization problem (3) is satisfied. To do so, note that the constraint in the firm's maximization problem (3) is always binding. If not, firms can raise profits by increasing labor effort. Invert the constraint with respect to labor effort to write $l = \hat{l}(z, \upsilon(n))$, where $\upsilon(n) = \underline{\upsilon}(n)$ for all n. The profit maximization problem is

$$\max_{z \geq 0} \quad n\hat{l}(z, \upsilon(n)) - z. \tag{54}$$

By the implicit function theorem, $\hat{l}_z = (1 - T')/\phi'$, where I ignore function arguments to simplify notation. At an interior solution, the first-order condition is given by

$$\frac{n(1 - T'(z))}{\phi'(\hat{l}(z, \upsilon(n)))} - 1 = 0. \tag{55}$$

The second-order condition is strictly satisfied if the left-hand side of equation (55) is strictly decreasing in earnings z. The latter is true if and only if

$$-\phi''(l) - n^2 T''(z) < 0, \tag{56}$$

where I used the first-order condition (55) and substituted out for $\hat{l}(z, \upsilon(n)) = l$. Because $\phi(\cdot)$ is strictly convex, condition (56) is satisfied as long as the tax function is not too concave.

To determine how earnings z vary with ability, rewrite equation (55) and define

$$L(z, n) \equiv n(1 - T'(z)) - \phi'(\hat{l}(z, \upsilon(n))) = 0. \tag{57}$$

Next, apply the implicit function theorem and use the first-order condition (55) and the property $\hat{l}_\upsilon = -1/\phi'$ to find

$$z'(n) = -\frac{L_n(z, n)}{L_z(z, n)} = \frac{\phi'(l) + \frac{\phi''(l)}{\phi'(l)} n\upsilon'(n)}{\phi''(l) + n^2 T''(z)}. \tag{58}$$

From the incentive constraint (53), $\upsilon'(n) \geq 0$ as long as monopsony power is not too quickly increasing in ability (which is assumed throughout). The numerator in (58) is therefore unambiguously positive. Hence, $z'(n) > 0$ if and only if the denominator is positive as well. This is the case if and only if the second-order condition (56) is satisfied. Therefore, if the allocation satisfies the monotonicity condition $z'(n) > 0$, it follows that the first-order condition for profit maximization (55) is both necessary and sufficient.

IV Lagrangian and first-order conditions

Written in terms of the allocation variables, the optimal tax problem is

$$\max_{[\upsilon(n),\pi(n),l(n)]_{n_0}^{n_1}} \mathcal{W} = \int_{n_0}^{n_1} \Big[g(n)\upsilon(n) + (1-\Sigma)(1-\tau)\pi(n)\Big] f(n)dn, \tag{59}$$

$$\begin{aligned}
\text{s.t.} \quad & \int_{n_0}^{n_1} \Big[nl(n) - \upsilon(n) - \phi(l(n)) - (1-\tau)\pi(n)\Big] f(n)dn = G, \\
\forall n: \quad & \upsilon'(n) = \frac{\phi'(l(n))}{n}\Big[(1-\mu(n))l(n) - \frac{\mu'(n)}{\mu(n)}\pi(n)\Big], \\
\forall n: \quad & \pi'(n) = \mu(n)l(n) + \frac{\mu'(n)}{\mu(n)}\pi(n), \\
\forall n: \quad & l(n) \geq 0, \\
& \pi(n_0) \geq 0.
\end{aligned}$$

The corresponding Lagrangian is given by

$$\begin{aligned}
&\mathscr{L} = \qquad (60) \\
&\int_{n_0}^{n_1} \Bigg[\Bigg(g(n)\upsilon(n) + (1-\Sigma)(1-\tau)\pi(n) + \eta\Big(nl(n) - \upsilon(n) - \phi(l(n)) - (1-\tau)\pi(n) - G\Big)\Bigg) f(n) \\
&+ \chi(n)\frac{\phi'(l(n))}{n}\left((1-\mu(n))l(n) - \frac{\mu'(n)}{\mu(n)}\pi(n)\right) + \chi'(n)\upsilon(n) + \lambda(n)\left(\mu(n)l(n) + \frac{\mu'(n)}{\mu(n)}\pi(n)\right) \\
&+ \lambda'(n)\pi(n) + \psi(n)l(n)\Bigg] dn + \chi(n_0)\upsilon(n_0) - \chi(n_1)\upsilon(n_1) + \lambda(n_0)\pi(n_0) - \lambda(n_1)\pi(n_1) + \xi\pi(n_0).
\end{aligned}$$

Suppressing the function argument of $\phi'(\cdot)$ and $\phi''(\cdot)$ to simplify notation, the first-order conditions are given by

$$\upsilon(n): \quad (g(n) - \eta)\, f(n) + \chi'(n) = 0, \tag{61}$$

$$\pi(n): \quad (1-\tau)(1-\Sigma-\eta)f(n) - \frac{\mu'(n)}{\mu(n)}\left(\chi(n)\frac{\phi'}{n} - \lambda(n)\right) + \lambda'(n) = 0, \tag{62}$$

$$\begin{aligned}
l(n): \quad & \eta\left(n - \phi'\right) f(n) + \frac{\chi(n)}{n}\left((1-\mu(n))(\phi' + \phi'' l(n)) - \phi''\frac{\mu'(n)}{\mu(n)}\pi(n)\right) \\
& + \lambda(n)\mu(n) + \psi(n) = 0,
\end{aligned} \tag{63}$$

$$\chi(n): \quad \frac{\phi'}{n}\left((1-\mu(n))l(n) - \frac{\mu'(n)}{\mu(n)}\pi(n)\right) - \upsilon'(n) = 0, \tag{64}$$

$$\lambda(n): \quad \mu(n)l(n) + \frac{\mu'(n)}{\mu(n)}\pi(n) - \pi'(n) = 0, \tag{65}$$

$$\eta: \quad \int_{n_0}^{n_1} (nl(n) - \upsilon(n) - \phi(l(n)) - (1-\tau)\pi(n) - G) f(n)dn = 0, \tag{66}$$

$$\upsilon(n_0): \quad \chi(n_0) = 0, \tag{67}$$

$$\upsilon(n_1): \quad -\chi(n_1) = 0, \tag{68}$$

$$\pi(n_0): \quad \lambda(n_0) + \xi = 0, \tag{69}$$

$$\pi(n_1): \quad -\lambda(n_1) = 0, \tag{70}$$

$$\psi(n): \quad \psi(n)l(n)=0, \qquad \psi(n)\geq 0 \ \text{ and } \ l(n)\geq 0, \tag{71}$$

$$\xi: \quad \xi\pi(n_0)=0, \qquad \xi\geq 0 \ \text{ and } \ \pi(n_0)\geq 0. \tag{72}$$

I assume the second-order conditions for the welfare maximization problem are satisfied and that earnings $z(n)=nl(n)-\pi(n)$ satisfy the monotonicity condition $z'(n)>0$ if $l(n)>0$.

V Derivation of the optimal marginal tax rate

This Appendix derives the optimal marginal tax rate in the general case where monopsony power $\mu(n)$ varies with ability. To that end, it is useful to first derive an expression for the multipliers $\chi(n)$ and $\lambda(n)$. Combining equations (61) and (68) gives

$$\chi(n)=\chi(n_1)-\int_n^{n_1}\chi'(m)dm=-\int_n^{n_1}(\eta-g(m))\,f(m)dm. \tag{73}$$

Evaluate equation (73) at $n=n_0$ and use the transversality condition (67) and the normalization $\int_{n_0}^{n_1}g(n)f(n)dn=1$ to find

$$\int_{n_0}^{n_1}(\eta-g(n))\,f(n)dn=\eta-1=0. \tag{74}$$

This is a standard result in optimal tax theory. When the tax system is optimized, the marginal cost of public funds equals one: see Jacobs (2018). Next, define by

$$\bar{g}(n)=\frac{\int_n^{n_1}g(m)f(m)dm}{1-F(n)} \tag{75}$$

the average welfare weight of individuals with ability at least equal to n, so that $\chi(n)=-(1-\bar{g}(n))(1-F(n))$. Because $\bar{g}(n_0)=1$ and $g(n)$ is non-increasing in ability it follows that $\bar{g}(n)\leq 1$ and hence, $\chi(n)\leq 0$. To derive an expression for $\lambda(n)$, rewrite equation (62):

$$\lambda'(n)+\frac{\mu'(n)}{\mu(n)}\lambda(n)=(1-\tau)\Sigma f(n)-\frac{\mu'(n)}{\mu(n)}\frac{\phi'(l(n))}{n}\int_n^{n_1}(1-g(m))f(m)dm, \tag{76}$$

where I used equation (73) to substitute out for $\chi(n)$. Equation (76) is a linear differential equation in $\lambda(n)$. Using the transversality condition (70), the solution is

$$\lambda(n)=-\frac{\bar{\mu}(n)}{\mu(n)}(1-\tau)\Sigma(1-F(n))+\int_n^{n_1}\frac{\mu'(m)}{\mu(n)}\frac{\phi'(l(m))}{m}\int_m^{n_1}(1-g(s))f(s)dsdm, \tag{77}$$

where $\bar{\mu}(n)$ is the average monopsony power of individuals with ability at least equal to n. To sign $\lambda(n)$, note that $\phi'\geq 0$. If monopsony power is not too quickly increasing in ability (as assumed throughout), $\lambda(n)\leq 0$. Equations (69) and (72) then imply $\xi\geq 0$. The assumption that firms do not earn profits from hiring the least productive workers (i.e., $\pi(n_0)=0$) is therefore without loss of generality.

To derive an expression for the marginal tax rate, consider the first-order condition for labor effort (63). Because $\phi'=0$ and $\pi(n)=0$ if $l(n)=0$, the non-negativity constraint on labor effort is binding (i.e., $\psi(n) > 0$) if

$$nf(n) - \bar{\mu}(n)(1-\tau)\Sigma(1-F(n)) + \int_n^{n_1} \mu'(m)\frac{\phi'(l(m))}{m}\int_m^{n_1}(1-g(s))f(s)dsdm < 0, \tag{78}$$

where I imposed $\eta = 1$ and substituted out for $\lambda(n)$ using equation (77). The latter is true if the local Pareto parameter

$$a(n) = \frac{nf(n)}{1-F(n)} < \bar{\mu}(n)(1-\tau)\Sigma - \frac{\int_n^{n_1}\mu'(m)\frac{\phi'(l(m))}{m}\int_m^{n_1}(1-g(s))f(s)dsdm}{1-F(n)}. \tag{79}$$

Hence, at ability levels where condition (79) holds, optimal labor effort and earnings are zero: $l(n) = 0$ and $z(n) = nl(n) - \pi(n) = 0$. If monopsony power does not vary with ability (i.e., if $\mu(n) = \mu$), the right-hand side simplifies to $\mu(1-\tau)\Sigma$. At ability levels where condition (79) does not hold, labor effort and earnings are positive. Substituting $\psi(n) = 0$, $\eta = 1$ and the first-order condition for profit maximization $n(1-T') = \phi'$ in equation (63) gives

$$T'(z(n))nf(n) = -\frac{\chi(n)}{n}\left((1-\mu(n))(\phi' + \phi''l(n)) - \phi''\frac{\mu'(n)}{\mu(n)}\pi(n)\right) - \mu(n)\lambda(n). \tag{80}$$

Substituting $\chi(n)$ and $\lambda(n)$ from equations (73) and (77), equation (80) can be written as

$$\begin{aligned} T'(z(n))nf(n) &= (1-\bar{g}(n))(1-F(n))\frac{\phi'}{n}\left[(1-\mu(n))\left(1+\frac{\phi''l(n)}{\phi'}\right) - \pi(n)\frac{\phi''}{\phi'}\frac{\mu'(n)}{\mu(n)}\right] \\ &+ \bar{\mu}(n)(1-\tau)\Sigma(1-F(n)) - \int_n^{n_1}\mu'(m)\frac{\phi'(l(m))}{m}\left(\int_m^{n_1}(1-g(s))f(s)ds\right)dm. \end{aligned} \tag{81}$$

Next, use the condition $n(1-T') = \phi'$ and denote by $\varepsilon(n) = \frac{\phi'}{\phi''l(n)}$ the elasticity of labor supply. Upon dividing equation (81) by $nf(n)$ and rearranging, we obtain equation (19) from Proposition 4:

$$\begin{aligned} T'(z(n)) = \frac{1-F(n)}{nf(n)}\Bigg[&\bar{\mu}(n)(1-\tau)\Sigma + (1-\mu(n))(1-T'(z(n)))(1+1/\varepsilon(n))(1-\bar{g}(n)) \\ &- \frac{\mu'(n)\pi(n)(1-T'(z(n)))(1-\bar{g}(n))}{\mu(n)\varepsilon(n)l(n)} - \frac{\int_n^{n_1}\mu'(m)(1-T'(z(m)))\left(\int_m^{n_1}(1-g(s))f(s)ds\right)dm}{1-F(n)}\Bigg]. \end{aligned} \tag{82}$$

If monopsony power does not vary with ability (i.e., $\mu'(n) = 0$), the last two terms cancel. Substituting $\mu(n) = \bar{\mu}(n) = \mu$ gives equation (11) from Proposition 1.

From equation (82) it follows immediately that the optimal marginal tax rate is zero at the top: $T'(z(n_1)) = 0$. To show that the optimal marginal tax rate is generally positive, note that monopsony power is not too quickly increasing in ability (as assumed throughout): $\mu'(n)$ is bounded from above. Moreover, $\bar{g}(n) \leq 1$ and from the profit-maximization condition (4) it follows that the marginal tax rate cannot exceed one at an interior solution. It follows that the optimal marginal tax rate is generally positive.

VI Impact of monopsony power on optimal marginal tax rates

To derive an expression for the optimal marginal tax rate if monopsony power does not vary with ability and the utility function is iso-elastic (i.e., $\phi(l) = l^{1+1/\varepsilon}/(1+1/\varepsilon)$), substitute $\varepsilon(n) = \varepsilon$ in equation (11) and use the definition of $a(n)$. Rearranging gives the result from Corollary 1:

$$T'(z(n)) = \frac{\mu(1-\tau)\Sigma + (1-\mu)(1+1/\varepsilon)(1-\bar{g}(n))}{a(n) + (1-\mu)(1+1/\varepsilon)(1-\bar{g}(n))}. \tag{83}$$

This is a closed-form solution for the optimal marginal tax rate. To determine how the latter varies with monopsony power, differentiate equation (83) with respect to μ to find

$$\frac{\partial T'(z(n))}{\partial \mu} = \frac{a(n)((1-\tau)\Sigma - (1+1/\varepsilon)(1-\bar{g}(n))) + (1-\tau)\Sigma(1+1/\varepsilon)(1-\bar{g}(n))}{(a(n) + (1-\mu)(1+1/\varepsilon)(1-\bar{g}(n)))^2}. \tag{84}$$

Equation (84) is positive if and only if the numerator is positive. Because $\bar{g}(n_0) = 1$, this is always the case at the bottom of the income distribution if $(1-\tau)\Sigma > 0$. At higher ability levels, the impact of monposony power on optimal tax rates is generally ambiguous. To see why, note that $\bar{g}(n) < 1$ for all $n > n_0$ if the government wishes to reduce inequality generated by differences in ability. To derive the result from the corollary, divide the numerator in equation (84) by $a(n)(1-\tau)\Sigma(1-\bar{g}(n)) > 0$. The resulting expression is positive if and only if

$$((1-\tau)\Sigma))^{-1} < ((1+1/\varepsilon)(1-\bar{g}(n)))^{-1} + a(n)^{-1}. \tag{85}$$

VII Welfare effect of raising monopsony power

This Appendix analyzes the welfare effect of a proportional increase in monopsony power by α percent, starting from a situation where monopsony power might vary with ability. Hence, after the increase monopsony power is $\hat{\mu}(n) = \mu(n)(1+\alpha)$. Welfare is then given by

$$\begin{aligned}\mathscr{L}(\alpha) = \int_{n_0}^{n_1} \Bigg[\Big(&(g(n)-\eta)v(n) + (1-\Sigma-\eta)(1-\tau)\pi(n) + \eta(nl(n) - \phi(l(n)) - G) \Big) f(n) \\ &+ \chi'(n)v(n) + \chi(n)\frac{\phi'(l(n))}{n}\left((1-\mu(n)(1+\alpha))l(n) - \frac{\mu'(n)}{\mu(n)}\pi(n) \right) + \lambda'(n)\pi(n) \\ &+ \lambda(n)\left(\mu(n)(1+\alpha)l(n) + \frac{\mu'(n)}{\mu(n)}\pi(n) \right) + \psi(n)l(n) \Bigg] dn + \chi(n_0)v(n_0) - \chi(n_1)v(n_1) \\ &+ \lambda(n_0)\pi(n_0) - \lambda(n_1)\pi(n_1) + \xi\pi(n_0),\end{aligned} \tag{86}$$

which is the optimized Lagrangian (60) evaluated at $\hat{\mu}(n) = \mu(n)(1+\alpha)$. Here I used the fact that the increase in monopsony power is proportional, which implies

$$\frac{\hat{\mu}'(n)}{\hat{\mu}(n)} = \frac{\mu'(n)(1+\alpha)}{\mu(n)(1+\alpha)} = \frac{\mu'(n)}{\mu(n)}. \tag{87}$$

By the envelope theorem, the welfare effect is

$$\frac{\partial \mathcal{W}(\alpha)}{\partial \alpha} = \frac{\partial \mathscr{L}(\alpha)}{\partial \alpha} = \int_{n_0}^{n_1} \left(-\chi(n)\frac{\phi'}{n} + \lambda(n) \right) \mu(n) l(n) dn. \tag{88}$$

Next, use equations (73) and (77) to substitute out for $\chi(n)$ and $\lambda(n)$. This leads to

$$\begin{aligned} \frac{\partial \mathcal{W}(\alpha)}{\partial \alpha} = \int_{n_0}^{n_1} \Bigg[\frac{\phi'(l(n))}{n} \int_n^{n_1} (1-g(m)) f(m) dm - (1-\tau)\Sigma \int_n^{n_1} \frac{\mu(m)}{\mu(n)} f(m) dm \\ + \int_n^{n_1} \frac{\mu'(m)}{\mu(n)} \frac{\phi'(l(m))}{m} \left(\int_m^{n_1} (1-g(s)) f(s) ds \right) dm \Bigg] \mu(n) l(n) dn, \end{aligned} \tag{89}$$

which coincides with equation (20) from Proposition 4 after imposing $n(1-T') = \phi'$.

The above expression simplifies considerably if monopsony power does not vary with ability. The term in the second line of equation (89) cancels. Substituting $\mu(n) = \mu$ gives

$$\begin{aligned} &\frac{\partial \mathcal{W}(\alpha)}{\partial \alpha} = \\ &\int_{n_0}^{n_1} \Bigg[\frac{\mu}{1-\mu} \underbrace{(1-\mu)\frac{\phi'(l(n))l(n)}{n}}_{=\upsilon'(n)} \int_n^{n_1} (1-g(m)) f(m) dm - (1-\tau)\Sigma \underbrace{\mu l(n)}_{=\pi'(n)} \int_n^{n_1} f(m) dm \Bigg] dn. \end{aligned} \tag{90}$$

Apply integration by parts with boundary conditions $\bar{g}(n_0) = 1$ and $\pi(n_0) = 0$:

$$\frac{\partial \mathcal{W}(\alpha)}{\partial \alpha} = \int_{n_0}^{n_1} \left[\frac{\mu}{1-\mu}(1-g(n))\upsilon(n) - (1-\tau)\Sigma\pi(n) \right] f(n) dn. \tag{91}$$

The latter can be simplified further after defining

$$\Sigma^{\upsilon} = -\mathrm{Cov}[\boldsymbol{\upsilon}, \boldsymbol{\gamma}] \geq 0, \tag{92}$$

$$\Sigma^{k} = -\mathrm{Cov}[\boldsymbol{\sigma}(1-\tau)\bar{\pi}, \boldsymbol{\gamma}] = \Sigma(1-\tau)\bar{\pi} \geq 0. \tag{93}$$

The first measures the negative covariance between labor market payoffs $\upsilon(n)$ and welfare weights $\gamma(n, \sigma)$. The second measures the negative covariance between welfare weights and capital income $\sigma(1-\tau)\bar{\pi}$. It is proportional to the covariance between shareholdings and welfare weights introduced before. Substituting these terms in equation (91) gives

$$\frac{\partial \mathcal{W}(\alpha)}{\partial \alpha} = \frac{\mu}{1-\mu}\Sigma^{\upsilon} - \Sigma^{k}. \tag{94}$$

From this relationship, it immediately follows that if the tax system is optimized, an increase in monopsony power raises welfare if and only if (cf. Proposition 2)

$$\mu\Sigma^{\upsilon} > (1-\mu)\Sigma^{k}. \tag{95}$$

A closed-form expression for the welfare impact of monopsony power

As stated in the main text, it is possible to derive an expression for the welfare effect of raising monopsony power in terms of exogenous variables if the utility function is iso-elastic: $\phi(l) =$

$l^{1+1/\varepsilon}/(1+1/\varepsilon)$. To see this, recall that Corollary 1 gives a closed-form expression for the marginal tax rate:

$$T'(z(n)) = \frac{\mu(1-\tau)\Sigma + (1-\mu)(1+1/\varepsilon)(1-\bar{g}(n))}{a(n) + (1-\mu)(1+1/\varepsilon)(1-\bar{g}(n))}, \tag{96}$$

provided $a(n) \geq \mu(1-\tau)\Sigma$. Labor effort can then be determined from equation (4):

$$l(n) = n^{\varepsilon}\left(\frac{a(n) - \mu(1-\tau)\Sigma}{a(n) + (1-\mu)(1+1/\varepsilon)(1-\bar{g}(n))}\right)^{\varepsilon} \tag{97}$$

and $l(n) = 0$ if $a(n) < \mu(1-\tau)\Sigma$. Denote by $n' \geq n_0$ the highest ability level where the non-negativity constraint on labor effort $l(n) \geq 0$ binds. Substituting the above in equation (89) and setting $\mu(n) = \mu$ and hence, $\mu'(n) = 0$ gives

$$\begin{aligned} \frac{\partial \mathcal{W}(\alpha)}{\partial \alpha} = \int_{n'}^{n_1} \mu \Bigg[\left(\frac{a(n) - \mu(1-\tau)\Sigma}{a(n) + (1-\mu)(1+1/\varepsilon)(1-\bar{g}(n))}\right)(1-\bar{g}(n)) - (1-\tau)\Sigma \Bigg] \\ \times (1-F(n))n^{\varepsilon}\left(\frac{a(n) - \mu(1-\tau)\Sigma}{a(n) + (1-\mu)(1+1/\varepsilon)(1-\bar{g}(n))}\right)^{\varepsilon} dn, \end{aligned} \tag{98}$$

which is expressed solely in terms of exogenous variables.

Proof Corollary 2

To derive the result from Corollary 2, note that equation (95) gives a necessary and sufficient condition to determine if an increase in monopsony power raises welfare. Next, write

$$\begin{aligned} \Sigma^{v} &= \int_{n_0}^{n_1} (1-g(n))v(n)f(n)dn = \int_{n_0}^{n_1} (1-g(n))(z(n) - T(z(n)) - \phi(l(n)))f(n)dn \\ &= \int_{n_0}^{n_1} (1-g(n))(z(n) - T(z(n)))f(n)dn - \int_{n_0}^{n_1} (1-g(n))\phi(l(n))f(n)dn \\ &= -\mathrm{Cov}[\boldsymbol{z} - T(\boldsymbol{z}), \boldsymbol{\gamma}] - \int_{n_0}^{n_1} (1-g(n))\phi(l(n))f(n)dn \\ &= \Sigma^{\ell} - \int_{n_0}^{n_1} (1-g(n))\phi(l(n))f(n)dn. \end{aligned} \tag{99}$$

Because $g(n)$ is weakly decreasing in ability and averages to one, the second term on the last line of equation (99) is non-negative if labor effort is weakly increasing in ability. Therefore, $\Sigma^{\ell} \geq \Sigma^{v}$ if $l'(n) \geq 0$. In that case, an increase in monopsony power raises welfare only if

$$\mu\Sigma^{\ell} > (1-\mu)\Sigma^{k}. \tag{100}$$

Unlike equation (95), this condition is necessary but not sufficient.

VIII Optimal degree of monopsony power

Suppose monopsony power does not vary with ability: $\mu(n) = \mu$ for all n. Then, the welfare impact of raising monopsony power is (see equation (88)):

$$\frac{\partial \mathcal{W}(\mu)}{\partial \mu} = \int_{n_0}^{n_1} \left(-\chi(n) \frac{\phi'(l(n))}{n} + \lambda(n) \right) l(n) dn. \tag{101}$$

The solutions for $\chi(n)$ and $\lambda(n)$ do not depend on the degree of monopsony power and are given by (cf. equations (73) and (77)):

$$\chi(n) = -(1 - \bar{g}(n))(1 - F(n)), \qquad \lambda(n) = -(1 - \tau)\Sigma(1 - F(n)). \tag{102}$$

The solution for $l(n)$, in turn, is determined implicitly by the first-order condition (63):

$$\Gamma(l(n), \mu) = (n - \phi'(l(n)))f(n) + (1 - \mu)\frac{\chi(n)}{n}(\phi'(l(n)) + \phi''(l(n))l(n)) + \mu\lambda(n) = 0 \tag{103}$$

or $l(n) = 0$ if $\psi(n) > 0$, where I imposed $\mu(n) = \mu$, $\mu'(n) = 0$ and used $\eta = 1$. After substituting the solution for $\chi(n)$ and $\lambda(n)$, this equation pins down optimal labor effort as a function of exogenous variables only.

To determine if the welfare function is concave or convex in the degree of monopsony power, differentiate equation (101) again with respect to μ:

$$\frac{\partial^2 \mathcal{W}(\mu)}{\partial \mu^2} = \int_{n_0}^{n_1} \frac{\partial l(n)}{\partial \mu} \left[-\frac{\chi(n)}{n}(\phi'(l(n)) + \phi''(l(n))l(n)) + \lambda(n) \right] dn. \tag{104}$$

The impact of monopsony power on labor effort $\partial l(n)/\partial \mu$, in turn, can be found by applying the implicit function theorem on equation (103):

$$\frac{\partial l(n)}{\partial \mu} = -\frac{\partial \Gamma(l(n), \mu)/\partial \mu}{\partial \Gamma(l(n), \mu)/\partial l(n)}. \tag{105}$$

Because the solution for labor effort $l(n)$ maximizes the Lagrangian (60), it must be that $\Gamma(\cdot)$ is decreasing in labor effort: $\partial \Gamma(l(n), \mu)/\partial l(n) < 0$. Therefore, $\partial l(n)/\partial \mu > 0$ if and only if $\partial \Gamma(l(n), \mu)/\partial \mu > 0$, i.e., if and only if

$$\frac{\partial \Gamma(l(n), \mu)}{\partial \mu} = -\frac{\chi(n)}{n}(\phi'(l(n)) + \phi''(l(n))l(n)) + \lambda(n) > 0. \tag{106}$$

The right-hand side is exactly the term that is multiplied by $\partial l(n)/\partial \mu$ below the integral sign in equation (104). It follows that the welfare function is convex:

$$\frac{\partial^2 \mathcal{W}(\mu)}{\partial \mu^2} = \int_{n_0}^{n_1} \frac{-1}{\partial \Gamma(l(n), \mu)/\partial l(n)} \left[-\frac{\chi(n)}{n}(\phi'(l(n)) + \phi''(l(n))l(n)) + \lambda(n) \right]^2 dn \geq 0. \tag{107}$$

Because the welfare function $\mathcal{W}(\mu)$ is convex in μ, the degree of monopsony power that maximizes social welfare is either $\mu^* = 0$ (perfect competition) or $\mu^* = 1$ (full monopsony power). To determine which of these is optimal, compute the welfare difference

$$\Delta\mathcal{W} = \mathcal{W}(1) - \mathcal{W}(0) = \int_0^1 \mathcal{W}'(\mu)d\mu. \tag{108}$$

The marginal welfare impact of raising monopsony power $\mathcal{W}'(\mu)$ can be obtained directly from the relationship $\partial\mathcal{W}/\partial\alpha = (\partial\mathcal{W}/\partial\mu) \times \mu$, where $\partial\mathcal{W}/\partial\alpha$ follows from equation (94):

$$\mathcal{W}'(\mu) = \frac{1}{\mu}\left[\frac{\mu}{1-\mu}\Sigma^v - \Sigma^k\right]. \tag{109}$$

As a last step, combine equations (108) and (109) and use the property $d\log\mu = d\mu/\mu$. Full monopsony power is optimal if and only if $\Delta\mathcal{W} > 0$, i.e., if and only if

$$\int_0^1 \left[\frac{\mu}{1-\mu}\Sigma^v - \Sigma^k\right] d\log\mu > 0. \tag{110}$$

IX Participation margin

Setting up the optimal tax problem

The optimal tax problem is similar as in the model without a participation margin, see Appendix II. The government chooses $l(n)$, $v(n)$ and $\pi(n)$ for all n to maximize social welfare subject to resource and incentive constraints. The main differences are that the government also chooses a uniform benefit b paid to non-participants and it has to take into account that changes in the participation threshold $\varphi(n) = v(n) - b$ induce labor supply responses on the extensive margin, cf. equation (21).

To derive the welfare function, denote by $H(n,\sigma,\varphi)$ the joint distribution of types with density $h(n,\sigma,\varphi)$ and let $\gamma(n,\sigma,\varphi) \geq 0$ denote the welfare weight the government attaches to an individual of type (n,σ,φ). The average welfare weight is normalized to one. The welfare function can be derived using similar steps as in Appendix I:

$$\begin{aligned}
\mathcal{W} &= \int_{n_0}^{n_1}\int_{\sigma_0}^{\sigma_1}\int_{\varphi_0}^{\varphi_1} \gamma(n,\sigma,\varphi)\mathcal{U}(n,\sigma,\varphi)h(n,\sigma,\varphi)d\varphi d\sigma dn \\
&= \int_{n_0}^{n_1}\int_{\sigma_0}^{\sigma_1}\int_{\varphi_0}^{\varphi_1} \gamma(n,\sigma,\varphi)\Big[\max\{v(n)-\varphi, b\} + \sigma(1-\tau)\bar{\pi}\Big]h(n,\sigma,\varphi)d\varphi d\sigma dn \\
&= \int_{n_0}^{n_1}\int_{\sigma_0}^{\sigma_1}\left[\int_{\varphi_0}^{\varphi(n)} \gamma(n,\sigma,\varphi)(v(n)-\varphi)h(n,\sigma,\varphi)d\varphi + \int_{\varphi(n)}^{\varphi_1}\gamma(n,\sigma,\varphi)bh(n,\sigma,\varphi)d\varphi\right]d\sigma dn \\
&\quad + (1-\tau)\bar{\pi}\int_{n_0}^{n_1}\int_{\sigma_0}^{\sigma_1}\int_{\varphi_0}^{\varphi_1}\gamma(n,\sigma,\varphi)\sigma h(n,\sigma,\varphi)d\varphi d\sigma dn \\
&= \int_{n_0}^{n_1}\left[\int_{\varphi_0}^{\varphi(n)} g(n,\varphi)(v(n)-\varphi)k(n,\varphi)d\varphi + \int_{\varphi(n)}^{\varphi_1} g(n,\varphi)bk(n,\varphi)d\varphi\right]dn \\
&\quad + (1-\tau)\int_{n_0}^{n_1}\int_{\varphi_0}^{\varphi(n)}\pi(n)k(n,\varphi)d\varphi dn \int_{n_0}^{n_1}\int_{\sigma_0}^{\sigma_1}\int_{\varphi_0}^{\varphi_1}\gamma(n,\sigma,\varphi)\sigma h(n,\sigma,\varphi)d\varphi d\sigma dn \\
&= \int_{n_0}^{n_1}\int_{\varphi_0}^{\varphi(n)}\Big[g(n,\varphi)(v(n)-\varphi) + (1-\tau)(1-\Sigma)\pi(n)\Big]k(n,\varphi)d\varphi dn
\end{aligned}$$

$$+\int_{n_0}^{n_1}\int_{\varphi(n)}^{\varphi_1} g(n,\varphi)bk(n,\varphi)d\varphi dn. \tag{111}$$

As before, Σ is the negative covariance between shareholdings and welfare weights. In addition, $g(n,\varphi)$ and $k(n,\varphi)$ denote the welfare weight and density of individuals with ability n and participation costs φ, averaged over shareholdings σ:

$$g(n,\varphi) = \frac{\int_{\sigma_0}^{\sigma_1}\gamma(n,\sigma,\varphi)h(n,\sigma,\varphi)d\sigma}{\int_{\sigma_0}^{\sigma_1}h(n,\sigma,\varphi)d\sigma}, \quad k(n,\varphi) = \int_{\sigma_0}^{\sigma_1}h(n,\sigma,\varphi)d\sigma. \tag{112}$$

To derive the aggregate resource constraint, note that an individual with ability n participates if and only if her participation costs $\varphi \leq \varphi(n) = v(n) - b$. Therefore, the government's budget constraint is

$$\int_{n_0}^{n_1}\int_{\varphi_0}^{\varphi(n)}\big(T(z(n)) + \tau\pi(n)\big)k(n,\varphi)d\varphi dn = \int_{n_0}^{n_1}\int_{\varphi(n)}^{\varphi_1} bk(n,\varphi)d\varphi dn + G. \tag{113}$$

In words, the government collects income taxes $T(z(n))$ and profit taxes τ to finance a benefit b for non-participants and an exogenous revenue requirement G. To write the final result in terms of the allocation variables, substitute out for labor income taxes $T(z(n)) = nl(n) - \pi(n) - v(n) - \phi(l(n))$:

$$\begin{aligned}&\int_{n_0}^{n_1}\int_{\varphi_0}^{\varphi(n)}\big(nl(n) - v(n) - \phi(l(n)) - (1-\tau)\pi(n)\big)k(n,\varphi)d\varphi dn\\ =&\int_{n_0}^{n_1}\int_{\varphi(n)}^{\varphi_1} bk(n,\varphi)d\varphi dn + G.\end{aligned} \tag{114}$$

Using the definition for the participation rate (22), the final equation can be written as

$$\int_{n_0}^{n_1}\Big[p(\varphi(n))(nl(n) - v(n) - \phi(l(n)) - (1-\tau)\pi(n)) - (1-p(\varphi(n)))b\Big]f(n)dn = G. \tag{115}$$

where $f(n)$ is the density associated with the marginal distribution $F(n)$ of ability:

$$f(n) = \int_{\varphi_0}^{\varphi_1} k(n,\varphi)d\varphi. \tag{116}$$

The incentive constraints are the same as before and given by equations (52)–(53). This is because firms cannot observe participation costs. Hence, as in the model without a participation margin, equilibrium labor effort $l(n)$, earnings $z(n)$ and profits $\pi(n)$ can again be found by solving equations (4) and (7) together with the relationship $\pi(n) = nl(n) - z(n)$. As a result, the incentive constraints are unaffected.

The government chooses $l(n)$, $v(n)$ and $\pi(n)$ for all n and a benefit b to maximize social welfare (111) subject to the resource constraint (114), incentive constraints (52)–(53) and the requirements $\pi(n_0) \geq 0$ and $l(n) \geq 0$ for all n. The corresponding Lagrangian is

$$\mathscr{L} = \int_{n_0}^{n_1}\Bigg[\int_{\varphi_0}^{v(n)-b}\Big(g(n,\varphi)(b + (v(n) - b) - \varphi) + (1-\tau)(1-\Sigma)\pi(n) \tag{117}$$

$$
\begin{aligned}
&+ \eta\Big[nl(n) - (\upsilon(n) - b) - b - \phi(l(n)) - (1-\tau)\pi(n) - G\Big]\Big)k(n,\varphi)d\varphi \\
&+ \int_{\upsilon(n)-b}^{\varphi_1} \Big(g(n,\varphi)b - \eta\big(b + G\big)\Big)k(n,\varphi)d\varphi \\
&+ \chi(n)\frac{\phi'(l(n))}{n}\left[(1-\mu(n))l(n) - \frac{\mu'(n)}{\mu(n)}\pi(n)\right] + \chi'(n)(b + (\upsilon(n) - b)) \\
&+ \lambda(n)\left[\mu(n)l(n) + \frac{\mu'(n)}{\mu(n)}\pi(n)\right] + \lambda'(n)\pi(n) + \psi(n)l(n)\Bigg]dn \\
&+ \chi(n_0)(b + (\upsilon(n_0) - b)) - \chi(n_1)(b + (\upsilon(n_1) - b)) + \lambda(n_0)\pi(n_0) - \lambda(n_1)\pi(n_1) + \xi\pi(n_0),
\end{aligned}
$$

where I substituted out for $\varphi(n) = \upsilon(n) - b$ and wrote $\upsilon(n) = (\upsilon(n) - b) + b$ to make it easier to differentiate with respect to b and $\upsilon(n) - b$ directly (instead of b and $\upsilon(n)$).

Derivation of equation (23)

The first-order conditions are very similar to those in Appendix IV. Here I only state the ones that are new or different. The first-order condition with respect to the benefit b is

$$
\int_{n_0}^{n_1}\int_{\varphi_0}^{\varphi_1}(g(n,\varphi) - \eta)k(n,\varphi)d\varphi dn + \int_{n_0}^{n_1}\chi'(n)dn + \chi(n_0) - \chi(n_1) = 0. \tag{118}
$$

The final terms on the left-hand side cancel out. Because the average welfare weight is normalized to one, the above condition immediately implies $\eta = 1$, as before. The first-order condition with respect to η gives the aggregate resource constraint (115):

$$
\int_{n_0}^{n_1}\Big[p(nl(n) - \upsilon(n) - \phi(l(n)) - (1-\tau)\pi(n)) - (1-p)b - G\Big]f(n)dn = 0. \tag{119}
$$

Here and in the remainder, the function argument of $p(\varphi(n))$ is suppressed to save on notation. The first-order conditions with respect to profits $\pi(n)$ and labor effort $l(n)$ are

$$
\pi(n):\quad p(1-\tau)(1-\Sigma-\eta)f(n) - \frac{\mu'(n)}{\mu(n)}\left(\chi(n)\frac{\phi'}{n} - \lambda(n)\right) + \lambda'(n) = 0, \tag{120}
$$

$$
\begin{aligned}
l(n):\quad & p\eta\left(n - \phi'\right)f(n) + \frac{\chi(n)}{n}\left((1-\mu(n))(\phi' + \phi'' l(n)) - \phi''\frac{\mu'(n)}{\mu(n)}\pi(n)\right) \\
& + \lambda(n)\mu(n) + \psi(n) = 0.
\end{aligned} \tag{121}
$$

These equations differ from the ones in Appendix IV only through the multiplication of the first term on the left-hand side by the participation rate $p(\varphi(n))$.

The most significant difference compared to the model without a participation margin is in the first-order condition with respect to $\upsilon(n)$ (or, equivalently, $\upsilon(n) - b$):

$$
\begin{aligned}
& p\left[g_p(n) - \eta + \frac{p'}{p}\Big((1-\tau)(1-\Sigma)\pi(n) + \eta(nl(n) - \upsilon(n) - \phi(l(n)) + b - (1-\tau)\pi(n))\Big)\right]f(n) \\
& + \chi'(n) = 0.
\end{aligned} \tag{122}
$$

Here, $p' = p'(\varphi(n))$ captures the increase in the participation rate if the participation threshold $\varphi(n) = \upsilon(n) - b$ increases. Moreover, $g_p(n)$ is the average welfare weight of individuals with ability n who are employed, i.e., for whom $\varphi \leq \upsilon(n) - b$:

$$g_p(n) = \frac{\int_{\varphi_0}^{\upsilon(n)-b} g(n,\varphi)k(n,\varphi)d\varphi}{\int_{\varphi_0}^{\upsilon(n)-b} k(n,\varphi)d\varphi}. \tag{123}$$

The first-order condition with respect to labor effort (121) can be used to derive an expression for the optimal marginal tax rate $T'(z(n))$. If the non-negativity constraint on labor effort is not binding (i.e., $\psi(n) = 0$), the latter can be written as

$$pT'(z(n))nf(n) = -\frac{\chi(n)}{n}\left((1-\mu(n))(\phi' + \phi''l(n)) - \phi''\frac{\mu'(n)}{\mu(n)}\pi(n)\right) - \mu(n)\lambda(n). \tag{124}$$

where I substituted out for $\eta = 1$ and used $n(1 - T'(z(n))) = \phi'(l(n))$. The expression for $\chi(n)$ can be obtained as follows. Substitute $\eta = 1$ and $T(z(n)) = nl(n) - \pi(n) - \upsilon(n) - \phi(l(n))$ in equation (122) and use the transversality condition $\chi(n_1) = 0$:

$$\chi(n) = -\int_n^{n_1} p\left[1 - g_p(m) - \frac{p'}{p}\big(T(z(m)) + b + (1-(1-\tau)\Sigma)\pi(m)\big)\right] f(m)dm. \tag{125}$$

To derive an expression for $\lambda(n)$, substitute $\eta = 1$ in equation (120) and rearrange to find the following linear differential equation in $\lambda(n)$:

$$\lambda'(n) + \frac{\mu'(n)}{\mu(n)}\lambda(n) = p(1-\tau)\Sigma f(n) + \frac{\mu'(n)}{\mu(n)}\frac{\phi'}{n}\chi(n). \tag{126}$$

Using the transversality condition $\lambda(n_1) = 0$, the solution is

$$\begin{aligned}\lambda(n) &= -\int_n^{n_1}\left[(1-\tau)\Sigma p\frac{\mu(m)}{\mu(n)}f(m) + \frac{\mu'(m)}{\mu(n)}\frac{\phi'}{m}\chi(m)\right]dm \\ &= -(1-\tau)\Sigma\int_n^{n_1} p\frac{\mu(m)}{\mu(n)}f(m)dm \\ &+ \int_n^{n_1}\frac{\mu'(m)}{\mu(n)}\frac{\phi'}{m}\left[\int_m^{n_1} p\left[1 - g_p(s) - \frac{p'}{p}\big(T(z(s)) + b + (1-(1-\tau)\Sigma)\pi(s)\big)\right]f(s)ds\right]dm.\end{aligned} \tag{127}$$

Substituting the solutions for $\chi(n)$ and $\lambda(n)$ in equation (124) gives, after rearranging,

$$\begin{aligned}pT'(z(n))nf(n) &= (1-\tau)\Sigma\int_n^{n_1} p\mu(m)f(m)dm + (1-\mu(n))(1-T'(z(n)))(1+1/\varepsilon(n)) \\ &\times \int_n^{n_1} p\left[1 - g_p(m) - \frac{p'}{p}\big(T(z(m)) + b + (1-(1-\tau)\Sigma)\pi(m)\big)\right]f(m)dm \\ &- \frac{\mu'(n)\pi(n)\phi'}{\mu(n)\varepsilon(n)l(n)n}\int_n^{n_1} p\left[1 - g_p(m) - \frac{p'}{p}\big(T(z(m)) + b + (1-(1-\tau)\Sigma)\pi(m)\big)\right]f(m)dm \\ &- \int_n^{n_1}\mu'(m)\frac{\phi'}{m}\left[\int_m^{n_1} p\left[1 - g_p(s) - \frac{p'}{p}\big(T(z(s)) + b + (1-(1-\tau)\Sigma)\pi(s)\big)\right]f(s)ds\right]dm.\end{aligned} \tag{128}$$

If monopsony power does not vary with ability (i.e. $\mu(n) = \mu$), this condition simplifies to

$$p(\varphi(n))T'(z(n))nf(n) = \tag{129}$$

$$
\begin{aligned}
&\mu(1-\tau)\Sigma \int_n^{n_1} p(\varphi(m))f(m)dm + (1-\mu)(1-T'(z(n)))(1+1/\varepsilon(n)) \\
&\times \int_n^{n_1} p(\varphi(m))\left[1-g_p(m)-\frac{p'(\varphi(m))}{p(\varphi(m))}\big(T(z(m))+b+(1-(1-\tau)\Sigma)\pi(m)\big)\right] f(m)dm.
\end{aligned}
$$

As a final step, denote by $F_p(n) = 1 - \int_n^{n_1} p(\varphi(m))f(m)dm \; g_p(n)$ the distribution of employed individuals with density $f_p(n) = p(\varphi(n))f(n)$ and by $\hat{p}(n) = p'(\varphi(n))/p(\varphi(n))$ the semi-elasticity of the participation rate with respect to the threshold $\varphi(n)$. Substituting these in the final equation and rearranging gives the result from Proposition 5:

$$
\begin{aligned}
T'(z(n)) = \frac{1-F_p(n)}{nf_p(n)}\Big[&\mu(1-\tau)\Sigma + (1-\mu)(1-T'(z(n)))\,(1+1/\varepsilon(n)) \\
&\times \mathbb{E}\left[1-g_p(m)-\hat{p}(m)(T(z(m))+b)-\hat{p}(m)\pi(m)(1-(1-\tau)\Sigma)\middle| m \geq n\right]\Big],
\end{aligned}
\tag{130}
$$

where the expectation is taken using the distribution function $F_p(n)$.

Derivation of equation (24)

To derive the welfare impact of a proportional increase in monopsony power from $\mu(n)$ to $\mu(n)(1+\alpha)$, modify the Lagrangian (117) to

$$
\begin{aligned}
\mathscr{L}(\alpha) = &\int_{n_0}^{n_1}\Bigg[\int_{\varphi_0}^{\upsilon(n)-b}\Big(g(n,\varphi)(b+(\upsilon(n)-b)-\varphi)+(1-\tau)(1-\Sigma)\pi(n) \\
&+\eta\Big[nl(n)-(\upsilon(n)-b)-b-\phi(l(n))-(1-\tau)\pi(n)-G\Big]\Big)k(n,\varphi)d\varphi \\
&+\int_{\upsilon(n)-b}^{\varphi_1}\Big(g(n,\varphi)b-\eta\big(b+G\big)\Big)k(n,\varphi)d\varphi \\
&+\chi(n)\frac{\phi'(l(n))}{n}\left[(1-\mu(n)(1+\alpha))l(n)-\frac{\mu'(n)}{\mu(n)}\pi(n)\right]+\chi'(n)(b+(\upsilon(n)-b)) \\
&+\lambda(n)\left[\mu(n)(1+\alpha)l(n)+\frac{\mu'(n)}{\mu(n)}\pi(n)\right]+\lambda'(n)\pi(n)+\psi(n)l(n)\Bigg]dn \\
&+\chi(n_0)(b+(\upsilon(n_0)-b))-\chi(n_1)(b+(\upsilon(n_1)-b))+\lambda(n_0)\pi(n_0)-\lambda(n_1)\pi(n_1)+\xi\pi(n_0).
\end{aligned}
\tag{131}
$$

By the envelope theorem,

$$
\frac{\partial \mathcal{W}(\alpha)}{\partial \alpha} = \frac{\partial \mathscr{L}(\alpha)}{\partial \alpha} = \int_{n_0}^{n_1}\left(-\chi(n)\frac{\phi'}{n}+\lambda(n)\right)\mu(n)l(n)dn.
\tag{132}
$$

Substituting in the solution for $\chi(n)$ and $\lambda(n)$ from equations (125) and (127):

$$
\begin{aligned}
\frac{\partial \mathcal{W}(\alpha)}{\partial \alpha} = \int_{n_0}^{n_1}\mu(n)l(n)\Bigg[&-(1-\tau)\Sigma\int_n^{n_1} p\frac{\mu(m)}{\mu(n)}f(m)dm+\int_n^{n_1}\frac{\mu'(m)}{\mu(n)}\frac{\phi'}{m} \\
&\times\left(\int_m^{n_1} p\left[1-g_p(s)-\frac{p'}{p}\big(T(z(s))+b+(1-(1-\tau)\Sigma)\pi(s)\big)\right]f(s)ds\right)dm \\
&+\frac{\phi'}{n}\int_n^{n_1} p\left[1-g_p(m)-\frac{p'}{p}\big(T(z(m))+b+(1-(1-\tau)\Sigma)\pi(m)\big)\right]f(m)dm\Bigg]dn.
\end{aligned}
\tag{133}
$$

If monopsony power does not vary with ability, the final equation simplifies to

$$\begin{aligned}\frac{\partial \mathcal{W}(\alpha)}{\partial \alpha} = \int_{n_0}^{n_1} \mu l(n)\Bigg[&-(1-\tau)\Sigma \int_{n}^{n_1} p(\varphi(m)) f(m) dm + \frac{\phi'(l(n))}{n} \\ \times \int_{n}^{n_1} p(\varphi(m)) &\left[1 - g_p(m) - \frac{p'(\varphi(m))}{p(\varphi(m))}\big(T(z(m)) + b + (1-(1-\tau)\Sigma)\pi(m)\big)\right] f(m) dm\Bigg] dn.\end{aligned} \tag{134}$$

Applying integration by parts (as in Appendix VII) gives:

$$\begin{aligned}\frac{\partial \mathcal{W}(\alpha)}{\partial \alpha} =& -(1-\tau)\Sigma \int_{n_0}^{n_1} \pi(n) f_p(n) dn \\ &+ \frac{\mu}{1-\mu}\int_{n_0}^{n_1} \upsilon(n)\Big[1 - g_p(n) - \hat{p}(n)\big(T(z(n)) + b + (1-(1-\tau)\Sigma)\pi(n)\big)\Big] f_p(n) dn \\ =& -\Sigma^k - \frac{\mu}{1-\mu}\int_{n_0}^{n_1} \upsilon(n)\hat{p}(n)\pi(n)(1-(1-\tau)\Sigma) f_p(n) dn \\ &+ \frac{\mu}{1-\mu}\int_{n_0}^{n_1} \upsilon(n)\Big[1 - g_p(n) - \hat{p}(n)\big(T(z(n)) + b\big)\Big] f_p(n) dn,\end{aligned} \tag{135}$$

where I used the definitions $f_p(n) = p(\varphi(n))f(n)$, $\hat{p}(n) = p'(\varphi(n))/p(\varphi(n))$ and $\Sigma^k = \Sigma(1-\tau)\bar{\pi}$. Next, multiply the final equation by $1-\mu$ and set the resulting expression larger than zero. Rearranging gives the result from Proposition 5.

Proof Proposition 6

If the government does not value redistribution, $\gamma(n, \sigma, \varphi) = 1$ for all (n, σ, φ) and hence, $g_p(n) = 1$ for all n and $\Sigma = 0$. Substituting this in equation (130) and rearranging gives

$$\frac{T'(z(n))}{1 - T'(z(n))}\frac{\varepsilon(n)}{1+\varepsilon(n)} n f_p(n) = -(1-\mu)\int_{n}^{n_1} \hat{p}(m)\big(T(z(m)) + b + \pi(m)\big) f_p(m) dm. \tag{136}$$

To demonstrate that optimal marginal tax rates and the welfare impact of monopsony power are non-positive if monopsony power does not vary with ability, all that is required is to show $\chi(n) \geq 0$ for all n. To see why, note that if $\Sigma = 0$ and $\mu'(m) = 0$ for all m, equation (127) implies $\lambda(n) = 0$ for all n. Equations (124) and (132) then simplify to

$$T'(z(n)) n f_p(n) = -\frac{\chi(n)}{n}(1-\mu)(\phi'(l(n)) + \phi''(l(n))l(n)), \tag{137}$$

$$\frac{\partial \mathcal{W}}{\partial \mu} = -\int_{n_0}^{n_1} \chi(n)\frac{\phi'(l(n))l(n)}{n} dn, \tag{138}$$

where $\partial\mathcal{W}/\partial\mu = (\partial\mathcal{W}/\partial\alpha)/\mu$ measures the change in welfare if monopsony power increases. From the above relationships it follows that marginal tax rates and the welfare impact of monopsony power are non-positive if $\chi(n) \geq 0$ for all n. The rest of this Appendix is devoted to showing this is indeed the case.

The transversality conditions from the optimal tax problem with Lagrangian (117) imply $\chi(n_0) = \chi(n_1) = 0$. In words, the function $\chi(n)$ starts at a value of zero at n_0 and ends at a value of zero at n_1. At intermediate values, the function may be positive, zero or negative. I now demonstrate that if $\mu > 0$, there does not exist an interval $[n', n''] \subseteq [n_0, n_1]$ with $n'' > n'$

where $\chi(n') = \chi(n'') = 0$ and $\chi(n) \leq 0$ for all $n \in [n', n'']$. If such an interval does not exist, it must be that $\chi(n) \geq 0$ for all $n \in [n_0, n_1]$ and $\chi(n) > 0$ for all $n \in (n_0, n_1)$, as required.

To construct a contradiction, suppose such an interval does exist. Hence, suppose there exists an interval $[n', n'']$ with $n'' > n'$, such that $\chi(n') = \chi(n'') = 0$ and $\chi(n) \leq 0$ for all $n \in [n', n'']$. Then, according to equation (137), $T'(z(n)) \geq 0$ and hence, $n \geq \phi'(l(n))$ for all $n \in [n', n'']$. Furthermore, by equation (122), the function $\chi(n)$ evolves according to

$$\chi'(n) = -\hat{p}(n) \underbrace{\left[nl(n) - v(n) - \phi(l(n)) + b\right]}_{=\,\Omega(n)} f_p(n), \tag{139}$$

where I substituted $g_p(n) = \eta = 1$ and $\Sigma = 0$ and used the definitions $f_p(n) = p(\varphi(n))f(n)$ and $\hat{p}(n) = p'(\varphi(n))/p(\varphi(n))$. Because $\hat{p}(n)f_p(n) > 0$ and $\chi(n') = \chi(n'') = 0$ by construction, it must be that $\Omega(n) \propto \chi'(n)$ switches sign at least once on the interval $[n', n'']$ or $\Omega(n) = 0$ for all $n \in [n', n'']$. To determine how $\Omega(n) = nl(n) - v(n) - \phi(l(n)) + b$ varies with ability, differentiate with respect to n and use the incentive constraint (53):

$$\begin{aligned} \Omega'(n) &= (n - \phi'(l(n)))l'(n) + l(n) - v'(n) \\ &= (n - \phi'(l(n)))l'(n) + l(n) - (1-\mu)\frac{\phi'(l(n))l(n)}{n}. \end{aligned} \tag{140}$$

Monotonicity of labor earnings $z(n) = nl(n) - \pi(n)$, in turn, implies

$$z'(n) = l(n) + nl'(n) - \pi'(n) = (1-\mu)l(n) + nl'(n) > 0 \quad \Leftrightarrow \quad l'(n) > \left(\frac{\mu - 1}{n}\right) l(n), \tag{141}$$

where I used the incentive constraint (52) to substitute out for $\pi'(n) = \mu l(n)$. Substituting the final result in equation (140) and using that $n - \phi'(l(n)) \geq 0$ if $\chi(n) \leq 0$,

$$\begin{aligned} \Omega'(n) &= (n - \phi'(l(n)))l'(n) + l(n) - (1-\mu)\frac{\phi'(l(n))l(n)}{n} \\ &\geq (n - \phi'(l(n)))\left(\frac{\mu - 1}{n}\right) l(n) + l(n) - (1-\mu)\frac{\phi'(l(n))l(n)}{n} = \mu l(n) > 0, \end{aligned} \tag{142}$$

provided $\mu > 0$. Hence, the function $\Omega(n)$ is increasing on the interval $[n', n'']$ where $\chi(n') = \chi(n'') = 0$ and $\chi(n) \leq 0$. If, as required, $\Omega(n)$ switches sign, it must be that $\Omega(n') < 0$. Equation (139) then implies $\chi'(n') > 0$. But because $\chi(n') = 0$, it must be that $\chi(n' + \delta) > 0$ for a small, positive number $\delta \in (0, n'' - n')$. This contradicts the requirement that $\chi(n) \leq 0$ for all values of $n \in [n', n'']$. Hence, there cannot exist an interval $[n', n''] \subseteq [n_0, n_1]$ with $n'' > n'$ where $\chi(n') = \chi(n'') = 0$ and $\chi(n) \leq 0$ for all $n \in [n', n'']$.

To summarize, it has been established that there does not exist a range of values where $\chi(n)$ is zero at the end-points and $\chi(n) \leq 0$ in between. But the transversality conditions imply that $\chi(n_0) = \chi(n_1) = 0$. Because the function $\chi(n)$ cannot stay on or below the horizontal axis in the (n, χ) plane, it must be that $\chi(n) \geq 0$ for all n. Furthermore, if $\mu > 0$, $\chi(n) > 0$ for all $n \in (n_0, n_1)$. From equations (137) and (138) it follows that optimal marginal tax rates and the welfare impact of monopsony power are non-positive.

X Investment distortions from profit taxes

Characterizing equilibrium and setting up the optimal tax problem

If a firm is matched to a worker with ability n, it chooses labor effort and earnings to maximize after-tax profits, subject to the requirement that the labor market payoff $\upsilon(n) = z(n) - T(z(n)) - \phi(l(n))$ exceeds some ability-specific threshold $\underline{\upsilon}(n)$. In addition, firms choose what fraction $I \in [0,1]$ of output to invest in order to generate productivity growth of $A(I)$ percent. As stated in the main text, the investment costs are not tax deductible. Ignoring the constraints that labor effort and earnings are non-negative, the Lagrangian associated with the profit maximization problem is

$$\mathcal{L}(n) = ((1 + A(I))nl - z)(1 - \tau) - Inl + \kappa\Big[z - T(z) - \phi(l) - \underline{\upsilon}(n)\Big], \tag{143}$$

where κ denotes the Lagrange multiplier. The first-order condition with respect to the investment rate can directly be rearranged to find equation (27) from the main text. In addition, combining the first-order conditions with respect to l and z gives

$$n(1 - T'(z(n)))\left[1 + A(I(\tau)) - \frac{I(\tau)}{1-\tau}\right] = \phi'(l(n)), \tag{144}$$

where $l(n)$ and $z(n)$ denote the optimal choice of labor effort and earnings and $I(\tau)$ is the optimal investment rate that solves equation (27). Equation (144) is the counterpart of equation (4) from the baseline version of the model.

Let $\hat{\pi}(n)$ denote after-tax profits, which coincides with the optimized Lagrangian (143) or, equivalently, with equation (26). I now relate monopsony power to after-tax profits in a very similar way as before (see Definition 1). In particular, monopsony power and after-tax profits are related through

$$\hat{\pi}(n) = \mu(n)r(\tau)\int_{n_0}^{n} l(m)dm, \tag{145}$$

where $r(\tau) = (1 + A(I(\tau)))(1-\tau) - I(\tau)$ is the investment-retention rate.[49] Equation (145) replaces equation (7) in the description of the equilibrium. Clearly, profits are zero if labor markets are competitive. Conversely, if firms have full monopsony power, equation (145) coincides with the expression for profits if the outside option equals $\underline{\upsilon}(n) = -T(0)$ and firms do not make profits from hiring the least productive worker: $\hat{\pi}(n_0) = 0$.

The government optimally chooses labor market payoffs $\upsilon(n)$, after-tax profits $\hat{\pi}(n)$ and labor effort $l(n)$ for all n to maximize social welfare subject to resource and incentive constraints. In addition, it chooses the profit tax τ taking into account the impact on investment, as captured by equation (27). The welfare function is almost the same as before:

$$\mathcal{W} = \int_{n_0}^{n_1}\Big[g(n)\upsilon(n) + (1 - \Sigma)\hat{\pi}(n)\Big]f(n)dn, \tag{146}$$

---

[49] Without investment, $A(I) = I = 0$ and $r(\tau) = 1 - \tau$. Equation (145) then coincides with equation (7) after imposing the relationship $\hat{\pi}(n) = \pi(n)(1-\tau)$.

which differs from equation (9) only because equation (146) is written in terms of after-tax profits. The incentive constraints can be derived in the same way as before (see Appendix II):

$$\upsilon'(n)=\frac{\phi'(l(n))}{n}\left[(1-\mu(n))l(n)-\frac{\mu'(n)}{\mu(n)}\frac{\hat{\pi}(n)}{r(\tau)}\right], \tag{147}$$

$$\hat{\pi}'(n)=r(\tau)\mu(n)l(n)+\frac{\mu'(n)}{\mu(n)}\hat{\pi}(n). \tag{148}$$

The government's budget constraint, in turn, is given by

$$\int_{n_0}^{n_1}\left[T(z(n))+\tau\Big((1+A(I(\tau)))nl(n)-z(n)\Big)\right]f(n)dn=G. \tag{149}$$

Using the property that $z(n)(1-\tau)=r(\tau)nl(n)-\hat{\pi}(n)$, the aggregate resource constraint can be written as

$$\int_{n_0}^{n_1}\left[(1+A(I(\tau))-I(\tau))nl(n)-\upsilon(n)-\phi(l(n))-\hat{\pi}(n)\right]f(n)dn=G. \tag{150}$$

For analytical convenience, I focus on the case where monopsony power does not vary with ability: $\mu(n)=\mu$ and hence, $\mu'(n)=0$ for all n. The Lagrangian associated with the government's maximization problem is then (see equation (60)):

$$\begin{aligned}\mathscr{L}=\int_{n_0}^{n_1}\Bigg[\bigg(&g(n)\upsilon(n)+(1-\Sigma)\hat{\pi}(n) \\ &+\eta\Big((1+A(I(\tau))-I(\tau))nl(n)-\upsilon(n)-\phi(l(n))-\hat{\pi}(n)-G\Big)\bigg)f(n) \\ &+\chi(n)(1-\mu)\frac{\phi'(l(n))l(n)}{n}+\chi'(n)\upsilon(n)+\lambda(n)\mu r(\tau)l(n)+\lambda'(n)\hat{\pi}(n)+\psi(n)l(n)\Bigg]dn \\ &+\chi(n_0)\upsilon(n_0)-\chi(n_1)\upsilon(n_1)+\lambda(n_0)\hat{\pi}(n_0)-\lambda(n_1)\hat{\pi}(n_1)+\xi\hat{\pi}(n_0).\end{aligned} \tag{151}$$

Derivation of equation (28)

The first-order condition with respect to the state and control variables are almost the same as before: see Appendix IV. The first-order condition with respect to the profit tax is

$$\int_{n_0}^{n_1}\eta I'(\tau)(A'(I(\tau))-1)nl(n)f(n)dn+\int_{n_0}^{n_1}r'(\tau)\mu\lambda(n)l(n)dn=0. \tag{152}$$

To simplify this expression, combine the first-order conditions with respect to $\upsilon(n)$ and $\hat{\pi}(n)$, the transversality conditions and the property that the average welfare weight equals one to find $\eta=1$, $\chi(n)=-(1-\bar{g}(n))(1-F(n))$ and $\lambda(n)=-\Sigma(1-F(n))$. Furthermore, the first-order condition with respect to the investment rate (27) implies

$$A'(I(\tau))-1=\frac{1}{1-\tau}-1=\frac{\tau}{1-\tau}. \tag{153}$$

Substituting this in equation (152) and rearranging gives

$$\frac{\tau}{1-\tau}\frac{I'(\tau)}{I(\tau)r'(\tau)}\underbrace{\int_{n_0}^{n_1} I(\tau)nl(n)f(n)dn}_{=\mathcal{I}} = \frac{1}{r(\tau)}\int_{n_0}^{n_1}\underbrace{\mu r(\tau)l(n)}_{=\hat{\pi}'(n)}\Sigma(1-F(n))dn. \tag{154}$$

Multiplying both sides by $r(\tau)$ and applying integration by parts with boundary condition $\hat{\pi}(n_0)=0$:

$$\frac{\tau}{1-\tau}\times\frac{I'(\tau)r(\tau)}{I(\tau)r'(\tau)}\times\mathcal{I}=\Sigma^k, \tag{155}$$

where $\Sigma^k = \Sigma\int_{n_0}^{n_1}\hat{\pi}(n)f(n)dn$ is the negative covariance between capital income and welfare weights, as before. Next, define $\hat{I}(r(\tau)) = I(\tau)$ for all values of τ. Differentiating both sides with respect to τ gives $\hat{I}'(r(\tau))r'(\tau) = I'(\tau)$ and hence, $\hat{I}'(r(\tau)) = I'(\tau)/r'(\tau)$. Therefore,

$$\frac{\tau}{1-\tau}\times\underbrace{\frac{\hat{I}'(r(\tau))r(\tau)}{\hat{I}(r(\tau))}}_{=\epsilon_{I,r}}\times\mathcal{I}=\Sigma^k. \tag{156}$$

Rearranging gives equation (28) from Proposition 7.

Derivation of equation (29)

The expression for the optimal marginal tax rate can be obtained from the first-order condition of the Lagrangian (151) with respect to labor effort $l(n)$. Assuming the non-negativity constraint on labor effort is not binding (i.e., $\psi(n)=0$), the first-order condition can be rearranged to find

$$\begin{aligned}&\Big[(1+A(I(\tau))-I(\tau))n-\phi'(l(n))\Big]f(n) = \\ &(1-\mu)(1-\bar{g}(n))(1-F(n))\left(\frac{\phi'(l(n))}{n}\right)\left(\frac{\phi'(l(n))+\phi''(l(n))l(n)}{\phi'(l(n))}\right)+\mu r(\tau)\Sigma(1-F(n)).\end{aligned} \tag{157}$$

where I substituted out for $\eta=1$, $\chi(n) = -(1-\bar{g}(n))(1-F(n))$ and $\lambda(n) = -\Sigma(1-F(n))$. Next, multiply both sides of the equation with $(1-\tau)/r(\tau)$ and use the definition of $\varepsilon(n)$ and the first-order condition $n(1-T'(z(n)))r(\tau) = \phi'(l(n))(1-\tau)$ (see equation (144)). Equation (157) can then be written as

$$\begin{aligned}&\frac{1-\tau}{r(\tau)}\Big[(1+A(I(\tau))-I(\tau))n-\phi'(l(n))\Big]f(n) = \\ &\Big[\mu(1-\tau)\Sigma+(1-\mu)(1-T'(z(n)))(1+1/\varepsilon(n))(1-\bar{g}(n))\Big](1-F(n)).\end{aligned} \tag{158}$$

To proceed, rearrange equation (144) to find

$$n(1-T'(z(n)))\left(1+A(I(\tau))-I(\tau))-I(\tau)\frac{\tau}{1-\tau}\right)-\phi'(l(n))=0. \tag{159}$$

Therefore, the left-hand side of equation (158) simplifies to

$$\begin{aligned}
&\frac{1-\tau}{r(\tau)}\Big[(1+A(I(\tau))-I(\tau))n-\phi'(l(n))\Big]f(n)=\frac{1-\tau}{r(\tau)}\left[T'(z(n))n\frac{r(\tau)}{1-\tau}+nI(\tau)\frac{\tau}{1-\tau}\right]f(n)\\
&=nf(n)\left(T'(z(n))+\frac{\tau I(\tau)}{r(\tau)}\right). \qquad (160)
\end{aligned}$$

Combining equations (158) and (160) gives equation (29) from Proposition 7.

Derivation of equation (30)

The last step is to demonstrate that the expression for the welfare impact of monopsony power is the same as without investment distortions from profit taxes. From the Lagrangian (151),

$$\begin{aligned}
&\frac{\partial\mathcal{W}(\mu)}{\partial\mu}=\frac{\partial\mathscr{L}(\mu)}{\partial\mu}=\int_{n_0}^{n_1}\left(-\chi(n)\frac{\phi'(l(n))l(n)}{n}+\lambda(n)r(\tau)l(n)\right)dn \qquad (161)\\
&=\int_{n_0}^{n_1}\left(-\frac{1}{1-\mu}\chi(n)(1-\mu)\frac{\phi'(l(n))l(n)}{n}+\frac{1}{\mu}\lambda(n)\mu r(\tau)l(n)\right)dn\\
&=\int_{n_0}^{n_1}\left(\frac{1}{1-\mu}\int_{n}^{n_1}(1-g(m))f(m)dm\,\underbrace{(1-\mu)\frac{\phi'(l(n))l(n)}{n}}_{=\,v'(n)}-\frac{1}{\mu}\Sigma\int_{n}^{n_1}f(m)dm\,\underbrace{\mu r(\tau)l(n)}_{=\,\hat{\pi}'(n)}\right)dn,
\end{aligned}$$

where I substituted out for $\chi(n)=-\int_n^{n_1}(1-g(m))f(m)dm$ and $\lambda(n)=-\Sigma\int_n^{n_1}f(m)dm$. Next, apply integration by parts with boundary conditions $\bar{g}(n_0)=1$ and $\hat{\pi}(n_0)=0$:

$$\frac{\partial\mathcal{W}(\mu)}{\partial\mu}=\int_{n_0}^{n_1}\left[\frac{1}{1-\mu}(1-g(n))v(n)-\frac{1}{\mu}\Sigma\hat{\pi}(n)\right]f(n)dn=\frac{\Sigma^v}{1-\mu}-\frac{\Sigma^k}{\mu}, \qquad (162)$$

where, as before, Σ^v denotes the negative covariance between labor market payoffs and welfare weights and Σ^k denotes the negative covariance between capital income and welfare weights. As a final step, multiply equation (162) by $\mu(1-\mu)$ and set the resulting expression larger than zero. Rearranging gives equation (30) from Proposition 7.

XI Tax havens and profit shifting opportunities

Setting up the optimal tax problem

If firms shift a fraction $s(\tau)\in[0,1]$ of pretax profits to tax havens according to equation (32), the social welfare function (9) is modified to

$$\mathcal{W}=\int_{n_0}^{n_1}\Big[g(n)v(n)+(1-\Sigma)(1-(1-s(\tau))\tau-\rho(s(\tau)))\pi(n)\Big]f(n)dn, \qquad (163)$$

where the term $(1-\tau)\pi(n)$ is replaced by $(1-(1-s(\tau))\tau-\rho(s(\tau)))\pi(n)$: see equation (31). The government budget constraint, in turn, is given by

$$\int_{n_0}^{n_1}\Big[T(z(n))+\tau(1-s(\tau))\pi(n)\Big]f(n)dn=G, \qquad (164)$$

which differs from equation (10) because only a fraction $1 - s(\tau)$ of profits are taxed. To derive the aggregate resource constraint, use the property $T(z(n)) = z(n) - \upsilon(n) - \phi(l(n)) = nl(n) - \pi(n) - \upsilon(n) - \phi(l(n))$:

$$\int_{n_0}^{n_1} \Big[nl(n) - \upsilon(n) - \phi(l(n)) - (1 - (1 - s(\tau))\tau)\pi(n)\Big] f(n)dn = G. \tag{165}$$

The incentive constraints are the same as before (see Appendix II) and for analytical convenience, I focus on the case where monopsony power does not vary with ability: $\mu'(n) = 0$. The Lagrangian of the optimal tax problem is then

$$\begin{aligned}\mathscr{L} = \int_{n_0}^{n_1} \Bigg[&\bigg(g(n)\upsilon(n) + (1 - \Sigma)(1 - (1 - s(\tau))\tau - \rho(s(\tau)))\pi(n) \\ &+ \eta\Big(nl(n) - \upsilon(n) - \phi(l(n)) - (1 - (1 - s(\tau))\tau)\pi(n) - G\Big)\bigg) f(n) \\ &+ \chi(n)(1 - \mu)\frac{\phi'(l(n))l(n)}{n} + \chi'(n)\upsilon(n) + \lambda(n)\mu l(n) + \lambda'(n)\pi(n) + \psi(n)l(n)\Bigg] dn \\ &+ \chi(n_0)\upsilon(n_0) - \chi(n_1)\upsilon(n_1) + \lambda(n_0)\pi(n_0) - \lambda(n_1)\pi(n_1) + \xi\pi(n_0).\end{aligned} \tag{166}$$

Derivation of equation (33)

The first-order condition with respect to the profit tax τ is

$$\int_{n_0}^{n_1} \Big[- (1 - \Sigma)(1 - s(\tau)) + \eta\left(1 - s(\tau) - s'(\tau)\tau\right) \Big]\pi(n)f(n)dn = 0, \tag{167}$$

where I used the envelope condition that s maximizes the term $1 - (1 - s)\tau - \rho(s)$. As before, the first-order condition for $\upsilon(n)$, the transversality conditions and the normalization of welfare weights imply $\eta = 1$. Next, divide equation (167) by aggregate profits $\int_{n_0}^{n_1} \pi(n)f(n)dn$ and rearrange to find

$$s'(\tau)\tau = \Sigma(1 - s(\tau)). \tag{168}$$

As a final step, use the property $s'(\tau) = \frac{\partial s}{\partial \tau} = \frac{\partial(1-s)}{\partial(1-\tau)}$ and define $\epsilon_{1-s,1-\tau} = \frac{\partial(1-s)}{\partial(1-\tau)}\frac{1-\tau}{1-s}$. Rearranging gives equation (33) from Proposition 8.

Derivation of equation (34)

The expression for the optimal marginal tax rate is obtained from the first-order condition of the Lagrangian (166) with respect to labor effort:

$$\eta(n - \phi'(l(n)))f(n) + (1 - \mu)\frac{\chi(n)}{n}(\phi'(l(n)) + \phi''(l(n))l(n)) + \mu\lambda(n) = 0. \tag{169}$$

To simplify this expression, substitute $\eta = 1$ and use the property $n(1 - T'(z(n))) = \phi'(l(n))$ and the definition of $\varepsilon(n)$. Rearranging gives

$$T'(z(n)))nf(n) = -(1 - \mu)(1 - T'(z(n)))(1 + 1/\varepsilon(n))\chi(n) - \mu\lambda(n). \tag{170}$$

As before, the first-order condition for $v(n)$ and the transversality condition $\chi(n_1) = 0$ can be combined to find $\chi(n) = -(1-\bar{g}(n))(1-F(n))$. To obtain an expression for $\lambda(n)$, use the first-order condition for $\pi(n)$:

$$\Big[(1-\Sigma)(1-(1-s(\tau))\tau-\rho(s(\tau))) - \eta(1-(1-s(\tau))\tau)\Big] f(n) + \lambda'(n) = 0. \tag{171}$$

Next, substitute $\eta = 1$ and use the transversality condition $\lambda(n_1) = 0$ to find

$$\lambda(n) = -\Big[(1-(1-s(\tau))\tau-\rho(s(\tau)))\Sigma + \rho(s)\Big](1-F(n)). \tag{172}$$

Substituting the solutions for $\chi(n)$ and $\lambda(n)$ in equation (170) and rearranging gives equation (34) from Proposition 8.

Derivation of equation (35)

The welfare impact of raising monopsony power is (see equation (161)):

$$\begin{aligned}\frac{\partial \mathcal{W}(\mu)}{\partial \mu} = \frac{\partial \mathscr{L}(\mu)}{\partial \mu} &= \int_{n_0}^{n_1}\left(-\chi(n)\frac{\phi'(l(n))l(n)}{n} + \lambda(n)l(n)\right)dn \qquad (173)\\ &= \int_{n_0}^{n_1}\Bigg(\frac{1}{1-\mu}\int_{n}^{n_1}(1-g(m))f(m)dm\,\underbrace{(1-\mu)\frac{\phi'(l(n))l(n)}{n}}_{=v'(n)}\\ &\quad -\frac{1}{\mu}\int_{n}^{n_1}\Big[(1-(1-s(\tau))\tau-\rho(s(\tau)))\Sigma+\rho(s)\Big]f(m)dm\,\underbrace{\mu l(n)}_{=\pi'(n)}\Bigg)dn,\end{aligned}$$

where I substituted out for $\chi(n)$ and $\lambda(n)$. To proceed, apply integration by parts with boundary conditions $\bar{g}(n_0) = 1$ and $\pi(n_0) = 0$:

$$\begin{aligned}\frac{\partial \mathcal{W}(\mu)}{\partial \mu} &= \int_{n_0}^{n_1}\left[\frac{1}{1-\mu}(1-g(n))v(n) - \frac{1}{\mu}\Big[(1-(1-s(\tau))\tau-\rho(s(\tau)))\Sigma+\rho(s)\Big]\pi(n)\right]f(n)dn\\ &= \frac{\Sigma^{v}}{1-\mu} - \frac{\Sigma^{k}}{\mu} - \frac{\kappa(s(\tau))\bar{\pi}}{\mu}, \qquad (174)\end{aligned}$$

where Σ^k is the negative covariance between welfare weights and capital income, taking into account profit shifting. Denote by $R = \kappa(s(\tau))\bar{\pi}$ the total costs of profit shifting. Multiply equation (174) by $\mu(1-\mu)$ and set the resulting expression larger than zero. Rearranging gives the final result from Proposition 8.

XII Optimal linear taxation

This Appendix analyzes the optimal linear tax problem. The reason for doing so is that it clearly illustrates how a change in the tax function affects welfare through its impact on labor market outcomes, much in the spirit of the sufficient statistics, or tax perturbation approach (see, e.g., Chetty (2009) and Golosov et al. (2014)).[50] As in the calibrated version of the model,

---

[50] Because of the specific way of modeling monopsony power, characterizing the optimal non-linear tax system via tax perturbation methods turns out to be particularly challenging. The reason is that finding the equilibrium labor market outcomes (see Definition 2) requires solving an integral equation if the tax function $T(\cdot)$ is non-

for simplicity I assume the utility function is iso-elastic and monopsony power does not vary with ability: $\phi(l) = l^{1+1/\varepsilon}/(1+1/\varepsilon)$ and $\mu(n) = \mu \in [0,1]$ for all n. For a given tax function $T(z) = -q + tz$ for $z > 0$, equilibrium labor effort follows from equation (4):

$$l(n) = (1-t)^{\varepsilon} n^{\varepsilon}. \tag{175}$$

The profits firms generate from hiring a worker with ability n are then given by

$$\pi(n) = \mu \int_{n_0}^{n} l(m)dm = \mu \int_{n_0}^{n} m^{\varepsilon}(1-t)^{\varepsilon} dm = \left(\frac{\mu}{1+\varepsilon}\right)(1-t)^{\varepsilon}\left[n^{1+\varepsilon} - n_0^{1+\varepsilon}\right]. \tag{176}$$

Labor earnings, in turn, are equal to

$$z(n) = nl(n) - \pi(n) = \left(1 - \frac{\mu}{1+\varepsilon}\right)(1-t)^{\varepsilon} n^{1+\varepsilon} + \left(\frac{\mu}{1+\varepsilon}\right)(1-t)^{\varepsilon} n_0^{1+\varepsilon}. \tag{177}$$

From equation (176) it follows that firms do not generate profits from hiring the least productive workers: $\pi(n_0) = 0$ and $z(n_0) = n_0 l(n_0)$. As explained in the main text, the government can always guarantee this is the case (and finds it optimal to do so) by separately optimizing a benefit $-T(0)$ that is paid to individuals if they reject the contract offered by firms. The value of this benefit, which is never paid in equilibrium as individuals always accept the contract offered to them, can be found by equating $-T(0)$ to the labor market payoff $v(n_0)$ of an individual with ability n_0 if she were paid an hourly wage equal to her productivity:

$$\begin{aligned} -T(0) &= z(n_0) - T(z(n_0)) - \phi(l(n_0)) = n_0 l(n_0) - T(n_0 l(n_0)) - \phi(l(n_0)) \\ &= q + (1-t)n_0 l(n_0) - \frac{l(n_0)^{1+1/\varepsilon}}{1+1/\varepsilon} = q + \left(\frac{1}{1+\varepsilon}\right)(1-t)^{1+\varepsilon} n_0^{1+\varepsilon}, \end{aligned} \tag{178}$$

where I substituted out for $l(n_0)$ using equation (175). According to equation (178), the value of $-T(0)$ depends on the lump-sum transfer q and the linear tax rate t. Hence, whenever the government changes either of these, it also adjusts $-T(0)$ to make sure firms do not earn profits from hiring the least productive workers: $\pi(n_0) = 0$.

The labor market payoff for an individual with ability n is given by

$$\begin{aligned} v(n) &= q + (1-t)z(n) - \frac{l(n)^{1+1/\varepsilon}}{1+1/\varepsilon} \\ &= q + (1-t)\left[\left(1 - \frac{\mu}{1+\varepsilon}\right)(1-t)^{\varepsilon} n^{1+\varepsilon} + \left(\frac{\mu}{1+\varepsilon}\right)(1-t)^{\varepsilon} n_0^{1+\varepsilon}\right] - \frac{\varepsilon}{1+\varepsilon}(1-t)^{1+\varepsilon} n^{1+\varepsilon} \\ &= q + \left(\frac{1-\mu}{1+\varepsilon}\right)(1-t)^{1+\varepsilon} n^{1+\varepsilon} + \left(\frac{\mu}{1+\varepsilon}\right)(1-t)^{1+\varepsilon} n_0^{1+\varepsilon}. \end{aligned} \tag{179}$$

In equilibrium, the latter coincides with the outside option $\underline{v}(n)$. As explained in the main text, the outside option is taken as given by firms but not by the government as it depends on the tax function. Furthermore, equation (179) illustrates how the labor market payoff $v(n)$ (and hence, the outside option $\underline{v}(n)$) is related to the degree of monopsony power μ. The

---

linear. I instead use the mechanism-design approach introduced by Mirrlees (1971) to solve the optimal non-linear tax problem: see Appendices II and IV.

higher is the degree of monopsony power, the lower is the labor market payoff.

We conclude the characterization of the equilibrium for a given set of tax instruments by requiring the government's budget constraint is satisfied:

$$\int_{n_0}^{n_1} \Big[tz(n) - q + \tau\pi(n) \Big] f(n)dn = G. \tag{180}$$

The government chooses the linear tax rate t and the lump-sum transfer q to maximize social welfare

$$\mathcal{W} = \int_{n_0}^{n_1} \Big[g(n)v(n) + (1-\tau)(1-\Sigma)\pi(n) \Big] f(n)dn \tag{181}$$

subject to the requirement that the budget constraint (180) holds and taking into account how the tax instruments affect the labor market outcomes cf. equations (175), (176), (177) and (179). The value of the non-employment benefit $-T(0)$, which, as stated, is never paid in equilibrium, adjusts according to equation (178) to make sure $\pi(n_0) = 0$.

Denote by $\tilde{v}(n,t,q)$, $\tilde{z}(n,t)$ and $\tilde{\pi}(n,t)$ the labor market payoff of an individual with ability n, her labor earnings and the profits firms generate from hiring a worker with ability n. These are obtained from equations (179), (177) and (176), respectively. The Lagrangian of the government's problem can then be written as:

$$\begin{aligned} \mathcal{L} = &\int_{n_0}^{n_1} \Big[g(n)\tilde{v}(n,t,q) + (1-\tau)(1-\Sigma)\tilde{\pi}(n,t) \Big] f(n)dn \\ &+ \lambda \Big[\int_{n_0}^{n_1} \Big(t\tilde{z}(n,t) - q + \tau\tilde{\pi}(n,t) \Big) f(n)dn - G \Big]. \end{aligned} \tag{182}$$

The first-order condition with respect to the lump-sum transfer q immediately implies the marginal costs of public funds equals one:

$$\frac{\partial \mathcal{L}}{\partial q} = \int_{n_0}^{n_1} \Big[g(n) \underbrace{\frac{\partial \tilde{v}}{\partial q}}_{=1} - \lambda \Big] f(n)dn = \int_{n_0}^{n_1} \Big[g(n) - \lambda \Big] f(n)dn = 0 \quad \leftrightarrow \quad \lambda = 1. \tag{183}$$

The first-order condition with respect to the linear tax rate t is:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial t} = &\int_{n_0}^{n_1} \Big[g(n)\frac{\partial \tilde{v}}{\partial t} + (1-\tau)(1-\Sigma)\frac{\partial \tilde{\pi}}{\partial t} \Big] f(n)dn \\ &+ \underbrace{\lambda}_{=1} \int_{n_0}^{n_1} \Big[z(n) + t\frac{\partial \tilde{z}}{\partial t} + \tau\frac{\partial \tilde{\pi}}{\partial t} \Big] f(n)dn = 0. \end{aligned} \tag{184}$$

This condition clearly illustrates how a change in the tax rate t affects welfare through its impact on labor market outcomes and payoffs, much in the spirit of the sufficient statistics or tax perturbation approach (see Chetty (2009) and Golosov et al. (2014)).[51] It can be simplified further in a number of steps. First, use the property that both earnings $z(n)$ and profits $\pi(n)$

---

[51] Equation (184) can be written as a more standard sufficient statistics optimal tax formula (expressed in terms of welfare weights, behavioral responses and the income distribution) in two steps. First, integrate over the labor income as opposed to the ability distribution. Second, denote by $\tilde{l}(n,t)$ the labor effort of an individual with

are proportional to $(1-t)^{\varepsilon}$ and hence, iso-elastic with respect to a change in the net-of-tax rate:

$$\frac{\partial \tilde{z}}{\partial t} = -\varepsilon \frac{z(n)}{1-t}, \quad \frac{\partial \tilde{\pi}}{\partial t} = -\varepsilon \frac{\pi(n)}{1-t}. \tag{185}$$

Furthermore, using equations (175)–(177) and (179), it follows that

$$\begin{aligned} \frac{\partial \tilde{v}}{\partial t} &= -(1+\varepsilon)\left[\left(1 - \frac{\varepsilon}{1+\varepsilon} - \frac{\mu}{1+\varepsilon}\right)(1-t)^{\varepsilon} n^{1+\varepsilon} + \left(\frac{\mu}{1+\varepsilon}\right)(1-t)^{\varepsilon} n_0^{1+\varepsilon} \right] \\ &= -(1+\varepsilon) z(n) + \varepsilon (1-t)^{\varepsilon} n^{1+\varepsilon} = -(1+\varepsilon) z(n) + \varepsilon n l(n) = -z(n) + \varepsilon \pi(n). \end{aligned} \tag{186}$$

Substituting these relationships in equation (184):

$$\begin{aligned} 0 = &\int_{n_0}^{n_1} (1-g(n)) z(n) f(n) dn + \varepsilon \int_{n_0}^{n_1} g(n) \pi(n) f(n) dn \\ &- \varepsilon \int_{n_0}^{n_1} (1-(1-\tau)\Sigma) \frac{\pi(n)}{1-t} f(n) dn - \varepsilon t \int_{n_0}^{n_1} \frac{z(n)}{1-t} f(n) dn. \end{aligned} \tag{187}$$

Multiplying both sides by $1-t$ and rearranging terms gives

$$\begin{aligned} &t \times \left[\int_{n_0}^{n_1} (1-g(n)) z(n) f(n) dn + \varepsilon \int_{n_0}^{n_1} (z(n) + \pi(n) g(n)) f(n) dn \right] \\ &= \int_{n_0}^{n_1} (1-g(n)) z(n) f(n) dn + \varepsilon \int_{n_0}^{n_1} \pi(n) \Big[g(n) - (1-(1-\tau)\Sigma) \Big] f(n) dn. \end{aligned} \tag{188}$$

The optimal linear tax rate is therefore given by

$$t = \frac{\int_{n_0}^{n_1} (1-g(n)) z(n) f(n) dn + \varepsilon \int_{n_0}^{n_1} \pi(n) \Big[g(n) - (1-(1-\tau)\Sigma) \Big] f(n) dn}{\int_{n_0}^{n_1} (1-g(n)) z(n) f(n) dn + \varepsilon \int_{n_0}^{n_1} (z(n) + \pi(n) g(n)) f(n) dn} = \frac{N_t}{D_t}. \tag{189}$$

To proceed, use equations (176) and (177) to substitute out for $\pi(n)$ and $z(n)$ and the property that the average welfare weight equals one: $\int_{n_0}^{n_1} g(n) f(n) dn = 1$. The numerator then simplifies to:

$$\begin{aligned} N_t = (1-t)^{\varepsilon} \Bigg[&\left(1 - \frac{\mu}{1+\varepsilon}\right) \int_{n_0}^{n_1} (1-g(n)) n^{1+\varepsilon} f(n) dn + \left(\frac{\mu\varepsilon}{1+\varepsilon}\right) \int_{n_0}^{n_1} (g(n)-1) n^{1+\varepsilon} f(n) dn \\ &+ \left(\frac{\mu\varepsilon}{1+\varepsilon}\right)(1-\tau)\Sigma \int_{n_0}^{n_1} \left(n^{1+\varepsilon} - n_0^{1+\varepsilon}\right) f(n) dn \Bigg] \\ = (1-t)^{\varepsilon} \Bigg[&(1-\mu) \int_{n_0}^{n_1} (1-g(n)) n^{1+\varepsilon} f(n) dn \\ &+ \left(\frac{\mu\varepsilon}{1+\varepsilon}\right)(1-\tau)\Sigma \int_{n_0}^{n_1} \left(n^{1+\varepsilon} - n_0^{1+\varepsilon}\right) f(n) dn \Bigg]. \end{aligned} \tag{190}$$

---

ability n and use the relationship $\tilde{v}(n,t,q) = q + (1-t)\tilde{z}(n,t) - \phi(\tilde{l}(n,t))$ to substitute out for

$$\frac{\partial \tilde{v}}{\partial t} = -z(n) + (1-t)\frac{\partial \tilde{z}}{\partial t} - \frac{\phi'}{n} n \frac{\partial \tilde{l}}{\partial t} = -z(n) + (1-t)\frac{\partial \tilde{z}}{\partial t} - (1-t)\left[\frac{\partial \tilde{z}}{\partial t} + \frac{\partial \tilde{\pi}}{\partial t}\right] = -z(n) - (1-t)\frac{\partial \tilde{\pi}}{\partial t}.$$

Using similar steps, the denominator can be written as

$$
\begin{aligned}
D_t &= \int_{n_0}^{n_1} (1-g(n))z(n)f(n)dn + \varepsilon \int_{n_0}^{n_1} (z(n)+\pi(n)+\pi(n)(g(n)-1))f(n)dn \\
&= (1-t)^{\varepsilon}\Bigg[\left(1-\frac{\mu}{1+\varepsilon}\right)\int_{n_0}^{n_1}(1-g(n))n^{1+\varepsilon}f(n)dn + \left(\frac{\mu\varepsilon}{1+\varepsilon}\right)\int_{n_0}^{n_1}(g(n)-1)n^{1+\varepsilon}f(n)dn \\
&\quad + \varepsilon\int_{n_0}^{n_1} n^{1+\varepsilon}f(n)dn\Bigg]. \\
&= (1-t)^{\varepsilon}\left[(1-\mu)\int_{n_0}^{n_1}(1-g(n))n^{1+\varepsilon}f(n)dn + \varepsilon\int_{n_0}^{n_1} n^{1+\varepsilon}f(n)dn\right]. \qquad (191)
\end{aligned}
$$

Substituting these results in the optimal tax formula (189):

$$
t = \frac{(1-\mu)\int_{n_0}^{n_1}(1-g(n))n^{1+\varepsilon}f(n)dn + \left(\frac{\mu\varepsilon}{1+\varepsilon}\right)(1-\tau)\Sigma\int_{n_0}^{n_1}\left(n^{1+\varepsilon}-n_0^{1+\varepsilon}\right)f(n)dn}{(1-\mu)\int_{n_0}^{n_1}(1-g(n))n^{1+\varepsilon}f(n)dn + \varepsilon\int_{n_0}^{n_1} n^{1+\varepsilon}f(n)dn}. \qquad (192)
$$

To proceed, divide the numerator and denominator by $\int_{n_0}^{n_1} n^{1+\varepsilon}f(n)dn \times \varepsilon/(1+\varepsilon)$. This gives:

$$
t = \frac{\mu(1-\tau)\Sigma\int_{n_0}^{n_1}\left(1-\left(\frac{n_0}{n}\right)^{1+\varepsilon}\right)f(n)dn + (1-\mu)(1+1/\varepsilon)(1-\hat{g})}{(1+\varepsilon)+(1-\mu)(1+1/\varepsilon)(1-\hat{g})}, \qquad (193)
$$

where

$$
\hat{g} = \int_{n_0}^{n_1} g(n)\left(\frac{n^{1+\varepsilon}f(n)}{\int_{n_0}^{n_1} m^{1+\varepsilon}f(m)dm}\right)dn. \qquad (194)
$$

Because the average welfare weight equals one (i.e., $\int_{n_0}^{n_1} g(n)f(n)dn = 1$) and $g(n)$ is decreasing in ability, it follows that $\hat{g} \leq 1$. Intuitively, when computing the weighted average $\hat{g}$, more weight is given to small values of $g(n)$. The term $\hat{g}$ captures how much the government values redistributing from high-ability to low-ability individuals. The stronger is the desire to redistribute from high-ability to low-ability individuals, the lower is the value of $\hat{g}$.

As a final step, suppose that the lowest ability n_0 is small compared to the average ability level. In that case, $n_0^{1+\varepsilon}/\int_{n_0}^{n_1} n^{1+\varepsilon}f(n)dn$ is close to zero and the optimal tax formula (193) simplifies to

$$
t = \frac{\mu(1-\tau)\Sigma + (1-\mu)(1+1/\varepsilon)(1-\hat{g})}{(1+\varepsilon)+(1-\mu)(1+1/\varepsilon)(1-\hat{g})}. \qquad (195)
$$

This expression for the optimal linear tax rate is very similar to the expression for the optimal non-linear tax rate (14) and hence, the explanation is not repeated here. The main differences are that, naturally, (i) the local Pareto parameter of the ability distribution no longer plays a role in the optimal linear tax formula and (ii) the redistributive preferences of the government are captured by $\hat{g}$ instead of $\bar{g}(n)$.

XIII Additional figures

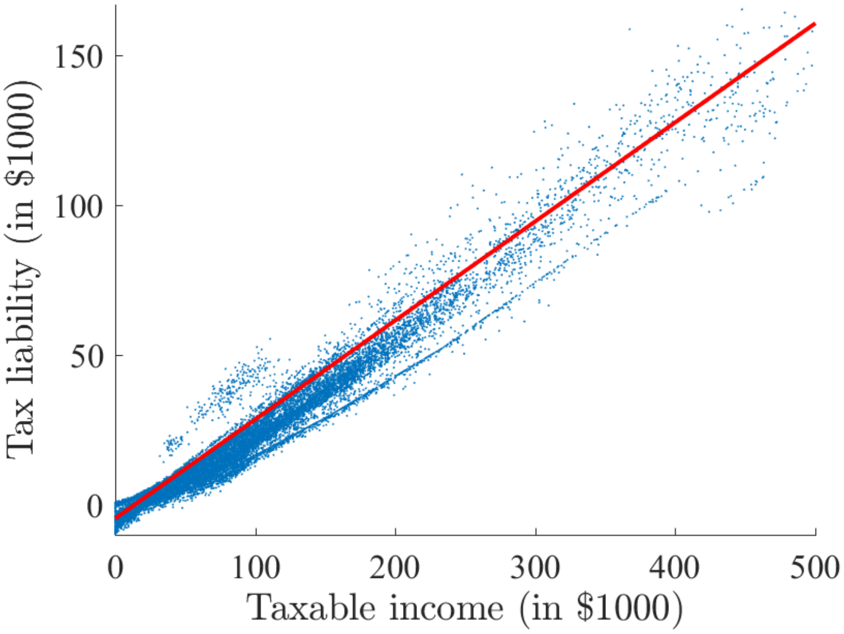


Figure 4: Current tax schedule

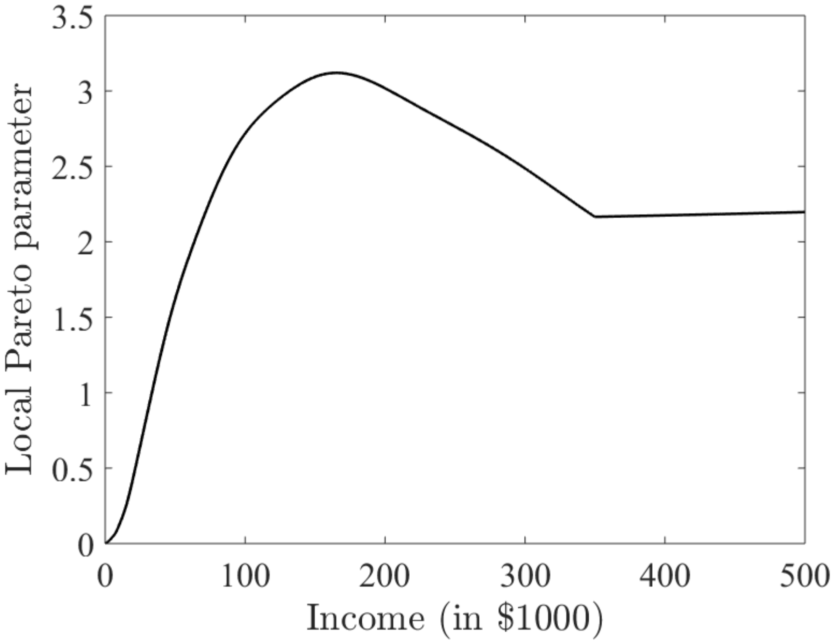


Figure 6: Local Pareto parameter

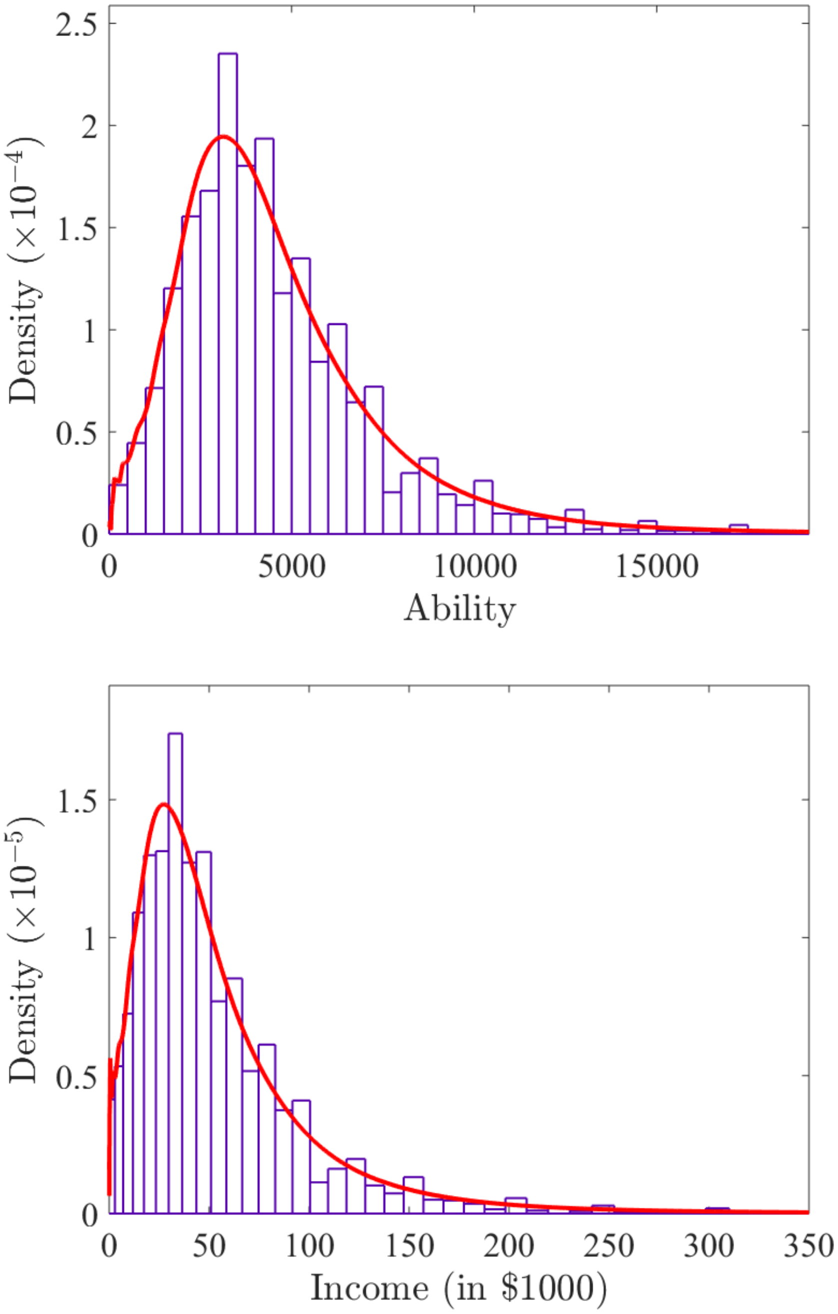


Figure 5: Distribution of ability and income