

Monopsony power, income taxation and welfare*

Albert Jan Hummel[†]

September 30, 2022

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, Florian Scheuer, Dirk Schindler, Kevin Spiritus, Christian Stoltenberg and seminar participants at Nederlandse Economendag 2019, IIPF Annual Congress 2020, CESifo Area Conference on Public Economics 2020 and Ludwig Maximilian University for many helpful comments and suggestions.

[†]University of Amsterdam, Tinbergen Institute and CESifo. 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., Arnold (2021), Azar et al. (2019, 2020, 2022), Benmelech et al. (2022), Hershbein et al. (2019), Lipsius (2018), Qiu and Sojourner (2019), Rinz (2022), Schubert et al. (2022), Thoresson (2021)). 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. In the baseline, monopsony power does not generate efficiency losses, but determines what share of the labor market surplus is translated into pure economic profits. Put differently, monopsony power does not reduce the size of the pie, but only the slice that goes to workers. 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 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 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 prof-

¹The 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>.

²For 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))

its, i.e., capital income. The reason is that part of the tax incidence falls on firms if they have monopsony power. Monopsony power thus makes taxes on labor earnings less effective in redistributing labor income, but more effective in redistributing capital income. As a result, optimal marginal tax rates with monopsony power are higher (lower) if the government has a strong preference for redistributing capital (labor) income than would be the case if labor markets are competitive. 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.

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 benchmark 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 to stimulate labor participation. Moreover, the welfare impact of raising monopsony power is lower if it distorts employment. I relax the second of these assumptions by analyzing an extension where profit taxes lead firms to engage in costly profit shifting. 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 \(2018\)](#). The results suggests 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 with monopsony power is less progressive than would be the case if labor markets are competitive. Moreover, taking monopsony power into account when designing tax policy leads to welfare gains that range between 0.13% and 1.12% of GDP in the calibrated economy, depending on the covariance between welfare weights and shareholdings. Furthermore, changing the degree of monopsony power from its value in the calibrated economy to zero has a welfare impact that ranges between -2.49% and +8.90% of GDP, again depending on the covariance between welfare weights and shareholdings. These figures, however, are clear upper bounds and only serve an illustrative purpose. The reason is that in the calibrated economy, *all* pure economic profits are attributed to monopsony power. If profits come from other sources as well, the implications of monopsony power for tax policy and welfare are smaller.

Related literature. A few papers study optimal income taxation in an environment where firms have monopsony power. [Hariton and Piaser \(2007\)](#), [Sleet and Yazici \(2017\)](#), [da Costa and Maestri \(2019\)](#) and [Hummel \(2021\)](#) analyze a model with productivity differences and labor supply responses on the intensive (hours, effort) margin, whereas [Cahuc and Laroque \(2014\)](#) study the desirability of minimum wages in a model with differences in participation costs and labor supply responses on the extensive (participation) margin, which I add in an extension. In these studies, monopsony power leads to a downward distortion in employment, either in hours worked or the number of individuals employed.³ Off-setting this distortion requires negative marginal (participation) tax rates if labor supply responds on the intensive (extensive) margin. By contrast, to focus on the distributional implications of monopsony power, employment is not distorted in the baseline version of the model.⁴ Optimal marginal tax rates only serve to redistribute income and are generally positive. Moreover, in my model monopsony power might raise welfare because firms have an informational advantage compared to the government about their workers' abilities. This is not possible in [Hariton and Piaser \(2007\)](#), [Cahuc and Laroque \(2014\)](#) and [da Costa and Maestri \(2019\)](#), since firms do not observe more information about their workers than the government.⁵

³An exception is [Sleet and Yazici \(2017\)](#), who study optimal taxation in a model with labor market frictions as in [Burdett and Mortensen \(1998\)](#). There, market power does not lead to a downward distortion in hours worked. Their focus, however, is on the impact of wage and profit dispersion and job ladders for optimal taxation, not so much the implications of monopsony power for optimal taxation and welfare.

⁴Monopsony power does generate distortions in labor participation in the extension where individuals also supply labor on the extensive margin.

⁵[Hummel \(2021\)](#) also assumes firms observe their workers' abilities while the government does not, but does not study the welfare impact of monopsony power.

Some recent papers study optimal taxation with mark-ups, see [Eeckhout et al. \(2021\)](#), [Kaplow \(2021\)](#), [Kushnir and Zubrickas \(2020\)](#) and [Atesagaoglu and Yazici \(2022\)](#). [Eeckhout et al. \(2021\)](#), [Kaplow \(2021\)](#) and [Kushnir and Zubrickas \(2020\)](#) do so in a static environment with unobserved heterogeneity and non-linear taxes. 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, which is not present in my framework. [Atesagaoglu and Yazici \(2022\)](#) study a dynamic representative-agent Ramsey setting with proportional taxes. They also find that taxes on labor income can be used to indirectly tax pure economic profits, as is the case in my model. The most important difference compared to these studies is that I assume firms offer workers a combination of earnings and labor effort instead of charging a (constant) mark-up that depresses the real wage. As a result, the outcome in the absence of taxation is efficient in the baseline version of my model. Tax policy is then exclusively aimed at redistribution and not to reduce inefficiencies from market power.

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 \(2003\)](#) (which [Manning \(2004\)](#) uses to study progressive income taxation, among other things). The first similarity is that firms can increase labor supply by paying higher wages. 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. The most important difference is that in [Robinson \(1933\)](#) and [Manning \(2003\)](#), monopsony power generates distortions. By contrast, the equilibrium in the absence of taxation is efficient in the baseline version of my model. 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 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 characterizes optimal tax policy and studies the welfare effects from monopsony power. Section 4 studies extensions where monopsony power generates efficiency losses and profit taxes are distortionary. Section 5 illustrates the findings by calibrating the baseline version of the model to the US economy. Section 6 concludes. An appendix contains all proofs and additional details of the analysis.

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 in the labor market.⁶ 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 proportions aggregate profits flow back to individuals. Both ability and shareholdings are taken to be exogenous. Let $H(n, \sigma)$ denote the joint distribution over ability and shareholdings and $h(n, \sigma)$ the corresponding density. Moreover, denote by $F(n)$ the marginal distribution of ability with density $f(n)$. The latter is assumed to be strictly positive on its entire support.

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

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

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

⁶Because all firms produce the same consumption good, there is no market power in the goods market.

⁷This 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 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

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

Here, $\sigma(1 - \tau)\bar{\pi}$ is after-tax capital income and $v(n) = z(n) - T(z(n)) - \phi(l(n))$ 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. As in [Mirrlees \(1971\)](#), I assume firms, unlike the government, perfectly observe the ability of their workers.⁸ While admittedly a strong assumption, what is 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, high-ability 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 $l \geq 0$.⁹ 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 $\underline{v}(n)$ that is weakly increasing in ability. Importantly, the outside option is taken as given by firms. Moreover, as will be made clear below, the outside option is related to 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

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

Despite that firms and workers are matched exogenously, each firm faces an upward-sloping labor supply curve in the sense that it can induce higher effort by promising higher earnings. In what follows, 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

⁸See [Stantcheva \(2014\)](#), [Bastani et al. \(2015\)](#) and [Craig \(2021\)](#) 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.

⁹This set-up differs from [Hummel \(2021\)](#), where a monopsonist can set an hourly wage $w = z/l$ and workers can choose how many hours l to work. This leads to a markdown and a downward distortion in hours worked, which is not the case if firms can offer a bundle (z, l) , as in the current paper.

$z(n)$.¹⁰ At an interior solution, labor effort and earnings are related through

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

At the optimum, firms offer bundles which equate an individual's productivity (on the left-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 the case 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: 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). As will be made clear below, this notion of monopsony power resembles bargaining power of firms. A high degree of firms' bargaining power could reflect that they face few competitors or that employees face significant switching costs. 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.¹¹ 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 to her. The Lagrangian associated with the firm's problem (3) is then

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

¹⁰See Appendix IV for additional details on the second-order conditions.

¹¹Formally, the outside option, which depends on the tax function $T(\cdot)$, is then given by

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

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: $\pi(n_0) = 0$.¹² 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 expression for profits if firms have full monopsony power:

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

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

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

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.¹³

The above definition of monopsony power does not have a clear interpretation or game-theoretic foundation. Despite this, it satisfies a number of useful properties. First, 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. Second, if taxes on labor income are linear, $\mu(n) \in [0, 1]$ corresponds to the generalized Nash bargaining weight: see Appendix I. Such a foundation is lacking, however, if the tax function is non-linear. The reason for nevertheless choosing this definition of monopsony power is that it keeps the optimal tax problem tractable and makes it possible to study the welfare impact of raising monopsony power.

The degree of monopsony power $\mu(n)$ is allowed to 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 perform more specialized tasks (see Caldwell and Danieli (2021) and Bachmann et al. (2022) 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.¹⁴ As shown in Appendix

¹²As is shown in Appendix VI, from an optimal tax perspective the assumption that firms do not earn profits from hiring the least productive workers is without loss of generality.

¹³To 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 the firm's maximization problem (3).

¹⁴Recall that the outside option (which in equilibrium coincides with the labor market payoff) weakly increases

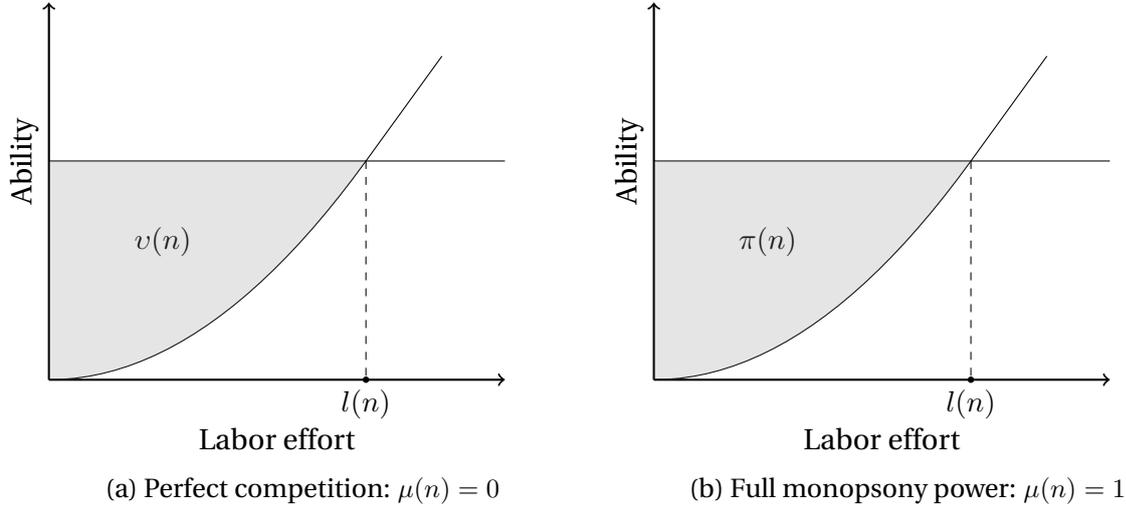


Figure 1: Labor market equilibrium

III, 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 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 there are no income taxes. 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). This assumption is relaxed in Section 4.1. In the baseline, 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.¹⁵

2.4 Government

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

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

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.

ability: $v'(n) \geq 0$ and hence $v'(n) \geq 0$. Thus, individuals with higher ability are not worse off. Equation (58) from Appendix III demonstrates that this assumption implies $\mu'(n)$ is bounded from above.

¹⁵The 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.

To make sure the government wishes to redistribute from individuals with high to individuals 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 .¹⁶ 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*

$$\mathcal{W} = \int_{n_0}^{n_1} \left[g(n)v(n) + (1 - \Sigma)(1 - \tau)\pi(n) \right] f(n)dn, \quad (9)$$

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

Proof. See Appendix II. □

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 then given by

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

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

¹⁶An 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 obtain a closed-form solution for the optimal marginal tax rate, as will be made clear below.

pure economic profits that can be taxed. Without a restriction on profit taxation, $\tau = 1$. 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 engage in costly profit shifting. Naturally, in that case, 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 describes 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 must coincide 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.¹⁷ 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

¹⁷The 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.

shareholdings. This is because monopsony power increases the share of the labor market 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 III 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.¹⁸ Hence, monopsony power mitigates inequality driven by differences in ability, but exacerbates inequality driven by differences in shareholdings.

The evidence that speaks to these hypotheses is scarce. A key challenge is that one needs variation in monopsony power, which should then be linked to measures of tax incidence and inequality. Some studies attempt to do this. Consistent with the model presented in the current paper, Berger et al. (2022) find that wage dispersion is lower in labor markets that are more concentrated. By contrast, Webber (2015) and Rinz (2022) 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.¹⁹ Regarding the impact of monopsony power on tax incidence, Fuest et al. (2018) find that workers bear a smaller share of the *corporate* tax burden if firms have more monopsony power. 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. (2022) 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 $\mu(n)$ 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. To solve this problem, I follow the approach pioneered by Mirrlees (1971) and

¹⁸From equations (57) and (58) 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).

¹⁹It should be noted, however, 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 Webber (2015) and Rinz (2022), i.e., the variance in log earnings and the P90/P10 earnings ratio: see Section 5. Moreover, the model can accommodate these findings if individuals with higher ability suffer less from monopsony (i.e., if $\mu'(n) < 0$).

characterize the allocation that maximizes welfare subject to resource and incentive constraints. The details can be found in Appendix III. 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*

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

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, which measures the percentage increase in labor effort $l(n)$ following a one percent increase in the net-of-tax rate $1 - T'(z(n))$.

Proof. See Appendix VI. □

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.²⁰ 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

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

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 labor 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 $z(n)$.²¹ 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.²² This is valuable provided profit taxation is restricted and the negative covariance between welfare weights and shareholdings is positive: $\tau < 1$ and $\Sigma > 0$. 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).

²⁰Hence, 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.

²¹Note 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.

²²A similar mechanism is operative in Ales and Sleet (2016) in their study on the taxation of CEO incomes. They also show that taxes on labor income can be used to indirectly tax profits. The lower the social value of profits, i.e., the higher $(1 - \tau)\Sigma$, the higher should be the tax rate on labor income *ceteris paribus*.

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 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

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

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 tax incidence that falls on firms and decreases the tax incidence that falls on workers. Monopsony power therefore makes labor income taxes less (more) effective in redistributing labor (capital) income. This result bears resemblance to Sleet and Yazici (2017), who study optimal labor income taxation in a framework à la Burdett and Mortensen (1998) with frictions and on and off the job search. In their framework, search frictions give rise to labor market power. The incidence of labor income taxes then too falls partly on firms' profits. Because profits are taxed at a confiscatory rate, this provides a force for lower tax rates on labor incomes. In my model profit taxation is restricted and as a result, the impact of monopsony power on optimal marginal tax rates is generally ambiguous. As it turns out, the impact 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*

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

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 μ raises the marginal tax rate $T'(z(n))$ if and only if

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

Proof. See Appendix VII. □

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 optimal 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).²³

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. 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

²³Monopsony 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.

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 monopsony power 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 VI for details. Hence, some individuals may not work at the second-best 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 optimal marginal tax rates satisfy 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*

$$\mu\Sigma^v > (1 - \mu)\Sigma^k, \quad (16)$$

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

Proof. See Appendix VIII. □

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, this 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 alleviates the equity-efficiency trade-off. 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.

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, 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. Consequently, 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 providing a universal basic income $-T(0)$ that is financed by a confiscatory tax on profits.²⁴ 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. One reason is that it is very difficult for policymakers to distinguish between normal returns and above-normal returns. In addition, 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

²⁴It is worth pointing out that the universal basic income should not be taxed away if individuals earn labor income. Put differently, optimal marginal tax rates are zero. To see this, substitute $\tau = \mu = 1$ in equation (11). Taxing labor income only distorts labor effort without generating any distributional benefits.

reduces inequality in labor market payoffs but exacerbates inequality in capital income.

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 VIII that the welfare effect of raising monopsony power can be written solely as a function of exogenous variables if the elasticity of labor supply 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.²⁵ 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*

$$\mu\Sigma^\ell > (1 - \mu)\Sigma^k, \quad (17)$$

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

Proof. See Appendix VIII. □

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 VIII 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 potentially easier to assess for policymakers, as it depends on after-tax labor income and not on the disutility of working.

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,

²⁵The welfare weights that make the current tax system optimal can be calculated using the inverse optimal tax method: see Bourguignon and Spadaro (2012).

individuals with higher ability do not suffer more from monopsony to an extent they are worse off than individuals whose ability is lower.

Proposition 3. *Suppose monopsony power $\mu(n)$ varies with ability. At an interior solution, the optimal marginal tax rate satisfies*

$$T'(z(n)) = \frac{1 - F(n)}{nf(n)} \left[\bar{\mu}(n)(1 - \tau)\Sigma + (1 - \mu(n))(1 - T'(z(n)))(1 + 1/\varepsilon(n))(1 - \bar{g}(n)) \right. \\ \left. - \frac{\mu'(n)\pi(n)(1 - T'(z(n)))(1 - \bar{g}(n))}{\mu(n)\varepsilon(n)l(n)} - \frac{\int_n^{n_1} \mu'(m)(1 - T'(z(m))) \left(\int_m^{n_1} (1 - g(s))f(s)ds \right) dm}{1 - F(n)} \right], \quad (18)$$

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

$$\frac{\partial \mathcal{W}(\alpha)}{\partial \alpha} = \int_{n_0}^{n_1} \left[(1 - T'(z(n))) \int_n^{n_1} (1 - g(m))f(m)dm - (1 - \tau)\Sigma \int_n^{n_1} \frac{\mu(m)}{\mu(n)} f(m)dm \right. \\ \left. + \int_n^{n_1} \frac{\mu'(m)}{\mu(n)} (1 - T'(z(m))) \left(\int_m^{n_1} (1 - g(s))f(s)ds \right) dm \right] \mu(n)l(n)dn. \quad (19)$$

Proof. See Appendix VI and VIII. □

Compared to the result from Proposition 1, two additional effects show up in the optimal tax formula (18). 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 (19) 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 (19) 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 than would be the case if 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 induce firms to engage in costly profit shifting (Section 4.2). 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.²⁶

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 ,

²⁶It 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: $\pi(n_0) = 0$.

shareholdings σ and participation costs φ is then

$$\mathcal{U}(n, \sigma, \varphi) = \max\{v(n) - \varphi, b\} + \sigma(1 - \tau)\bar{\pi}. \quad (20)$$

Participation costs φ are subtracted from the labor market payoff $v(n)$ as they lower the utility from working. Equation (20) determines a participation threshold $\varphi(n) = v(n) - b$ at every 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

$$p(\varphi(n)) = \frac{\int_{\sigma_0}^{\sigma_1} \int_{\varphi_0}^{\varphi(n)} h(n, \sigma, \varphi) d\varphi d\sigma}{\int_{\sigma_0}^{\sigma_1} \int_{\varphi_0}^{\varphi_1} h(n, \sigma, \varphi) d\varphi d\sigma}, \quad (21)$$

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 $v(n)$ 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: $v(n) + \pi(n) \geq \varphi$. However, individuals become employed only if their *individual* labor market payoff exceeds the participation costs: $v(n) \geq \varphi$. 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 4. *Suppose monopsony power does not vary with ability and individuals supply labor on the extensive margin according to equation (20). At an interior solution, the optimal marginal tax rate satisfies*

$$T'(z(n)) = \frac{1 - F_p(n)}{n f_p(n)} \left[\mu(1 - \tau)\Sigma + (1 - \mu)(1 - T'(z(n)))(1 + 1/\varepsilon(n)) \right. \\ \left. \times \mathbb{E} [1 - g_p(m) - \hat{p}(m)\pi(m)(1 - (1 - \tau)\Sigma) - \hat{p}(m)(T(z(m)) + b) | m \geq n] \right], \quad (22)$$

where the conditional expectation $\mathbb{E}[\cdot]$ is taken using the distribution of employed individuals $F_p(n) = 1 - \int_n^{\sigma_1} p(\varphi(m)) f(m) dm$ 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 $\varphi(n) = v(n) - b$. Furthermore,

an increase in monopsony power μ raises welfare if and only if

$$\begin{aligned} & \mu \int_{n_0}^{n_1} v(n)(1 - g_p(n))f_p(n)dn > (1 - \mu)\Sigma^k \\ & + \mu \int_{n_0}^{n_1} v(n)\hat{p}(n)\pi(n)(1 - (1 - \tau)\Sigma)f_p(n)dn + \mu \int_{n_0}^{n_1} v(n)\hat{p}(n)(T(z(n)) + b)f_p(n)dn. \end{aligned} \quad (23)$$

Proof. See Appendix IX. □

The optimal marginal tax rate (22) 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 (22). 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 (22). 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 (2021), 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 (22) gives an optimal tax formula that is very similar to the ones derived in these papers.²⁷

Equation (23) 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 $\mu\Sigma^v = -\mu\text{Cov}[v, \gamma]$ 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. Pro-

²⁷There 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 (2021) consider a model with a discrete set of ability levels.

vided 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 (23), 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.

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 from monopsony power does not come at zero efficiency costs if it distorts labor participation. This explains why monopsony power has an ambiguous effect on welfare even if profit taxation is unrestricted: it alleviates the classic 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.

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.

To study how the results are affected by distortions from profit taxes, 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:

$$\Pi(\bar{\pi}, \tau) = \max_{s \in [0,1]} \{s\bar{\pi}(1 - \tilde{\rho}(s)) + (1 - s)\bar{\pi}(1 - \tau)\} = \max_{s \in [0,1]} \{(1 - (1 - s)\tau - \rho(s))\bar{\pi}\}. \quad (24)$$

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

$$\rho'(s) = \tau. \quad (25)$$

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 study the welfare impact of raising monopsony power in Appendix X. The main findings are summarized below.

Proposition 5. *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 (25). At an interior solution, the optimal profit tax τ satisfies*

$$\frac{\tau}{1 - \tau} = \frac{\Sigma}{\epsilon_{1-s,1-\tau}}, \quad (26)$$

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

$$T'(z(n)) = \frac{1 - F(n)}{nf(n)} \left[\mu \left([1 - (1 - s(\tau))\tau - \rho(s(\tau))] \Sigma + \rho(s(\tau)) \right) + (1 - \mu)(1 - T'(z(n)))(1 + 1/\varepsilon(n))(1 - \bar{g}(n)) \right]. \quad (27)$$

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

$$\mu \Sigma^v > (1 - \mu) \Sigma^k + (1 - \mu) R, \quad (28)$$

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 X. □

Equation (26) gives an inverse elasticity rule for the optimal profit tax. 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 (26), 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 (27) 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: $(1 - (1 - s)\tau - \rho(s))\Sigma \geq (1 - \tau)\Sigma$.²⁸ 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 5 states that the existence of profit shifting opportunities makes it less likely that an increase in monopsony power raises welfare: compare equations (28) 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 welfare impact of raising monopsony power is lower if firms engage in costly profit shifting.

5 Numerical illustration

This Section illustrates the main findings in the baseline version of the model presented in Section 2 where monopsony power does not vary with ability, there are no efficiency losses from monopsony power and profit taxes do not affect any economic decisions. After presenting the calibration (Section 5.1), I analyze how monopsony power affects optimal income taxation (Section 5.2) and welfare (Section 5.3).

5.1 Calibration

The model is calibrated 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. The CPS data is supplemented with estimates of the share of pure economic profits in GDP and a number of parameters based on existing literature. Below I describe how these inputs are used to construct a baseline economy with the same set of primitives as the model presented in Section 2. The calibration strategy is summarized in Table 1.

²⁸To 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 $1 - (1 - s)\tau - \rho(s) \geq 1 - \tau$.

5.1.1 Functional forms

To calibrate the model, I first characterize the equilibrium for a given tax schedule $T(z)$ and the disutility of labor $\phi(l)$. Regarding the former, as in [Saez \(2001\)](#) I approximate the current US tax schedule on labor earnings by a linear function

$$T(z(n)) = -q + tz(n), \quad (29)$$

provided labor earnings are positive: $z(n) > 0$. Values for the lump-sum transfer q and the constant marginal tax rate t are obtained by regressing the tax liability (computed as the sum of state and federal taxes) on taxable income, both of which are obtained from CPS.²⁹ This gives $q = \$4,590$ and $t = 33.1\%$ with an R^2 of approximately 0.94. [Figure 4](#) in [Appendix XI](#) plots the actual and fitted values for incomes up to \$500,000.

Recall that the individual utility function is quasi-linear in consumption: $u(c, l) = c - \phi(l)$. For the disutility of labor, I use the following iso-elastic form:

$$\phi(l) = \frac{l^{1+1/\varepsilon}}{1 + 1/\varepsilon}, \quad (30)$$

where ε governs the convexity. It measures the percentage increase in labor effort (and given the choice of the tax schedule [\(29\)](#), the percentage increase in labor earnings as well) following a one percent increase in the net-of-tax rate: see [equation \(4\)](#). Furthermore, as will be demonstrated below, ε also captures the percentage increase in labor effort if firms pay a one percent higher wage. A natural target thus appears to be the *firm-level* elasticity of labor supply, which is a crucial statistic in the new monopsony models ([Manning \(2003\)](#)). This statistic, however, is typically estimated using variation in the number of individuals available to each firm (see, e.g., [Falch \(2010\)](#), [Hirsch et al. \(2010\)](#), [Staiger et al. \(2010\)](#), [Ransom and Sims \(2010\)](#), [Webber \(2015\)](#), [Dube et al. \(2020\)](#)). By contrast, in my model the number of workers available to each firm is fixed and firms can only incentivize them to work longer hours by offering a contract that implies a higher hourly wage. Therefore, a more suitable target for ε is the elasticity of labor supply on the intensive margin. I select a value of $\varepsilon = 0.33$ based on [Chetty \(2012\)](#). This value is also well within the range of common estimates for the elasticity of taxable income (ETI), which measures the responsiveness of earnings to a change in the net-of-tax rate. See [Saez et al. \(2012\)](#) for an overview.

5.1.2 Equilibrium outcomes

With the above specification of the utility function and the tax schedule, it is straightforward to solve for the equilibrium outcomes (cf. [Definition 2](#)). Labor effort follows directly from [equation \(4\)](#):

$$l(n) = (1 - t)^\varepsilon n^\varepsilon. \quad (31)$$

²⁹In the analysis I include all individuals between 25 years and 65 years who earn strictly positive labor income and whose hourly wage exceeds half the federal minimum wage of \$7.25

From this relationship it can be seen that a one percent increase in the net-of-tax rate $1 - t$ leads to an ε percent increase in labor effort. The same is true for labor earnings. To see this, substitute out for labor effort in equation (7) and use the relationship $z(n) = nl(n) - \pi(n)$ to find

$$\begin{aligned} z(n) &= \left(1 - \frac{\mu}{1 + \varepsilon}\right) (1 - t)^\varepsilon n^{1+\varepsilon} + \left(\frac{\mu}{1 + \varepsilon}\right) (1 - t)^\varepsilon n_0^{1+\varepsilon} \\ &= \left(1 - \frac{\mu}{1 + \varepsilon}\right) (1 - t)^\varepsilon n^{1+\varepsilon} + \left(\frac{\mu}{1 + \varepsilon}\right) z(n_0). \end{aligned} \quad (32)$$

According to equation (32), 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).³⁰ The profits $\pi(n) = nl(n) - z(n)$ firms generate from hiring a worker with ability n are then given by

$$\pi(n) = \left(\frac{\mu}{1 + \varepsilon}\right) (1 - t)^\varepsilon [n^{1+\varepsilon} - n_0^{1+\varepsilon}] = \left(\frac{\mu}{1 + \varepsilon - \mu}\right) (z(n) - z(n_0)). \quad (33)$$

For a given tax rate t , elasticity of labor supply ε and degree of monopsony power μ , equations (32) and (33) 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 labor income) 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 . If n_0 is small (and hence, $z(n_0)$ is small as well), the hourly wage is proportional to an individual's productivity. Using equation (33), the mark-up, i.e., the measure of "exploitation" introduced by [Pigou \(1920\)](#), is

$$\frac{n - w(n)}{w(n)} = \frac{\mu}{1 + \varepsilon - \mu}. \quad (34)$$

From the above relationship it follows that the mark-up is increasing in the degree of monopsony power μ and decreasing in the elasticity of labor supply ε . Second, equation (32) implies that if firms have monopsony 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 (32) it is clear that monopsony power mitigates inequality in labor earnings driven by differences in ability. Despite this, monopsony 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 scales down labor earnings proportionally for this particular choice of

³⁰The reason why the lowest income shows up in equation (32) is that the government provides a non-employment benefit which guarantees firms make no profits from hiring the least productive workers: $\pi(n_0) = 0$ and $z(n_0) = n_0 l(n_0)$. This can only be the case for *any* degree of monopsony 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.

the utility and tax function. In the more general case 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.

5.1.3 Degree of 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, or above-normal profits. Because my model abstracts from productive capital, I calibrate monopsony power μ to target the ratio of aggregate, above-normal 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 (33) over the ability distribution and dividing by aggregate labor income $\bar{z} = \int_{n_0}^{n_1} z(n)f(n)dn$. This gives

$$\left(\frac{\bar{\pi}}{\bar{z}}\right) = \left(\frac{\mu}{1 + \varepsilon - \mu}\right) \left(1 - \left(\frac{z(n_0)}{\bar{z}}\right)\right) \Leftrightarrow \mu = (1 + \varepsilon) \left[\frac{(\bar{\pi}/\bar{z})}{1 + (\bar{\pi}/\bar{z}) - (z(n_0)/\bar{z})}\right]. \quad (35)$$

To obtain values for $z(n_0)$ and \bar{z} , I define an individual's labor earnings in the CPS data as her income from wage and salary payments. Individuals with labor income below \$5,710 are excluded.³¹ Consequently, $z(n_0)$ is slightly above this value. For individuals whose labor earnings are 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 \(2018\)](#).³² Average annual labor earnings are then approximately \$65,488. Substituting out for the elasticity of labor supply, the estimate of the ratio of profits to wages from [Barkai and Benzell \(2018\)](#) and the ratio of the lowest earnings to average earnings gives a degree of monopsony power of approximately $\mu = 0.28$. It should be emphasized that this value is an upper bound. The reason is that, through the lens of my model, *all* above-normal profits originate from monopsony power. However, in reality these profits come from a variety of sources, monopsony power being one of them. Clearly, the fact that the degree of monopsony power in the calibrated economy is an upper bound has implications for the quantitative results. I get back to this point below.

³¹This is approximately the annual income for someone who works full-time (on average 35 hours per week) during the full year (at least 45 weeks) at an hourly wage half the federal minimum wage of \$7.25.

³²If 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 $\mathbb{E}[z|z \geq z'] = \tilde{a}z'/(\tilde{a} - 1)$.

5.1.4 Ability distribution

As in [Saez \(2001\)](#), I calibrate the ability distribution to match the empirical income distribution. To do so, I use equation (32) and calculate the ability n for each individual whose annual labor income exceeds \$5,710. 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 XI. 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.³³ This 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.

5.1.5 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 $G = \$22,782$, which in the calibrated economy corresponds to approximately 28.7% of aggregate output. The calibration strategy is summarized in Table 1.

Variable	Target	Source	Value
μ	Aggregate profits over wages	Barkai and Benzell (2018)	0.28
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	\$22,782
$T(z)$	Tax liability	CPS 2018	Figure 4
$F(n)$	Income distribution	CPS 2018	Figure 5

Table 1: Calibration

³³Let $\tilde{F}(z(n))$ 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 $\tilde{a}(z(n)) = z(n)\tilde{f}(z(n))/(1 - \tilde{F}(z(n)))$ 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 (32).

5.2 Optimal tax policy

5.2.1 Welfare function

The government's preferences for redistribution are described by the average welfare weight $g(n) \geq 0$ it attaches to individuals with the same ability $n \in [n_0, n_1]$ and the negative covariance between shareholdings and welfare weights, as captured by $\Sigma \in [0, 1]$. The first (second) of these determines how much the government wishes to reduce inequality generated by differences in ability (shareholdings). In what follows, 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 has a strong desire to redistribute capital income because all profits flow back to individuals whose welfare weight equals zero. Regarding the average welfare weight of individuals with the same ability, I use the following specification:

$$g(n) = \zeta n^{-\beta}. \quad (36)$$

Here, $\zeta > 0$ is a scaling parameter which ensures the average welfare weight for all individuals equals one 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 attaches the same average weight to individuals of all ability levels. Conversely, if $\beta \rightarrow \infty$, the government only cares about individuals with the lowest ability. I select the value of β such that the average marginal tax rate at the optimal tax system with perfectly competitive labor markets (obtained by setting $\mu = 0$ in equation (14)) equals the current rate $t = 33.1\%$.

5.2.2 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, all results are plotted against *current* labor earnings. The red, solid line shows 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%. Because all individuals work, the marginal tax rate equals zero at the bottom of the income distribution, cf. Seade (1977). Intuitively, a non-zero marginal tax rate for the least-skilled workers only distorts their labor supply and does not generate distributional benefits. After that, marginal tax rates increase rapidly to values slightly above 40% before displaying the conventional U-shape pattern, see Diamond (1998) and Saez (2001). This reflects mostly the behavior of the local Pareto parameter $a(n)$, as shown in Figure 6. At earnings levels above \$350,000, the local Pareto parameter is assumed to be constant and as a result, marginal tax rates are approximately constant as well.

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: $\mu = 0.28$ 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. Moreover, there are no benefits from indirectly taxing profits if $\Sigma = 0$. Hence, for similar distortions in labor effort, the distributional benefits of a higher marginal tax rate are smaller. It follows that if $\Sigma = 0$, optimal marginal tax rates with monopsony power are lower than would be the case if labor markets are perfectly competitive. The average reduction in optimal marginal tax rates brought about by monopsony power is 6.8 percentage points. It should be emphasized that this difference is likely to be an upper bound, because in the calibrated economy *all* economic profits are attributed to monopsony power. If the actual degree of monopsony power is lower, then so will be the difference between the optimal tax schedules with and without monopsony power.

The black, dotted line plots the optimal marginal tax rates if firms have monopsony power and the government has a very strong preference for redistributing capital income: $\Sigma = 1$. Naturally, marginal tax rates are higher compared to the case with $\Sigma = 0$ because there are distributional benefits from indirectly taxing profits. The average increase brought about by a change in the covariance between welfare weights and shareholdings is approximately 17.0 percentage points. The difference is particularly striking at low earnings levels. With a preference for redistributing capital income, the marginal tax rate for the least-skilled workers is not zero: see equation (14). In fact, the low value of the local Pareto parameter $a(n)$ at the bottom of the skill distribution implies that marginal tax rates are so high that individuals whose current labor earnings are below \$13,082 do not work at optimal allocation. For these workers, the constraint $l(n) \geq 0$ in the optimal tax problem is binding. The government prefers to have some individuals not work in order to reduce aggregate profits. At the lowest level of positive earnings, marginal tax rates are very high and close to 100%. Intuitively, the low value of the local Pareto parameter makes marginal tax rates on labor earnings a particularly effective tool to indirectly tax profits. After that, marginal tax rates follow a U-shape pattern and remain approximately constant from the point where the Pareto tail starts.

According to Corollary 1, the impact of monopsony power on optimal marginal tax rates is generally ambiguous. This can be seen by comparing the red, solid and black, dotted lines in Figure 2. At low levels of earnings, marginal tax rates with monopsony power and a preference for redistributing capital income are higher than would be the case if labor markets are perfectly competitive. If firms have monopsony power, marginal tax rates on labor earnings can be used to indirectly tax profits. This provides a force for higher marginal tax rates that is especially relevant at the bottom of the income distribution (where the local Pareto parameter is low). At higher levels of earnings, marginal tax rates with monopsony power are lower than would be the case if labor markets are perfectly competitive. In Figure 2, this happens for individuals whose current labor earnings are above \$70,000. Because part of the incidence falls on firms, monopsony power makes taxes on labor earnings less effective in redistributing labor income. This provides a force for lower marginal tax rates that is most relevant at higher levels of labor income. The analysis here suggests that if the government has a preference for redistributing *both* labor and capital income, monopsony power tends to increase optimal marginal tax rates at lower earnings levels and to decrease optimal marginal tax rates at higher earnings levels. In that sense, and in line with the theoretical prediction

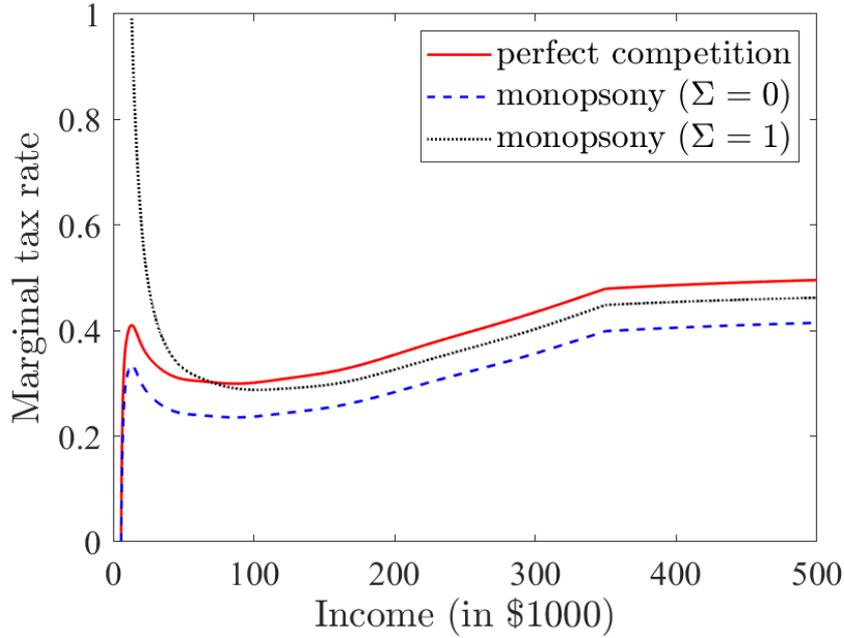


Figure 2: Optimal marginal tax rates

from Corollary 1, monopsony power makes the optimal tax schedule less progressive.

5.3 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 from its value in the calibrated economy (i.e., $\mu = 0.28$) 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 to zero.

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 \$103 and \$889 in consumption equivalents, or between 0.13% and 1.12% 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

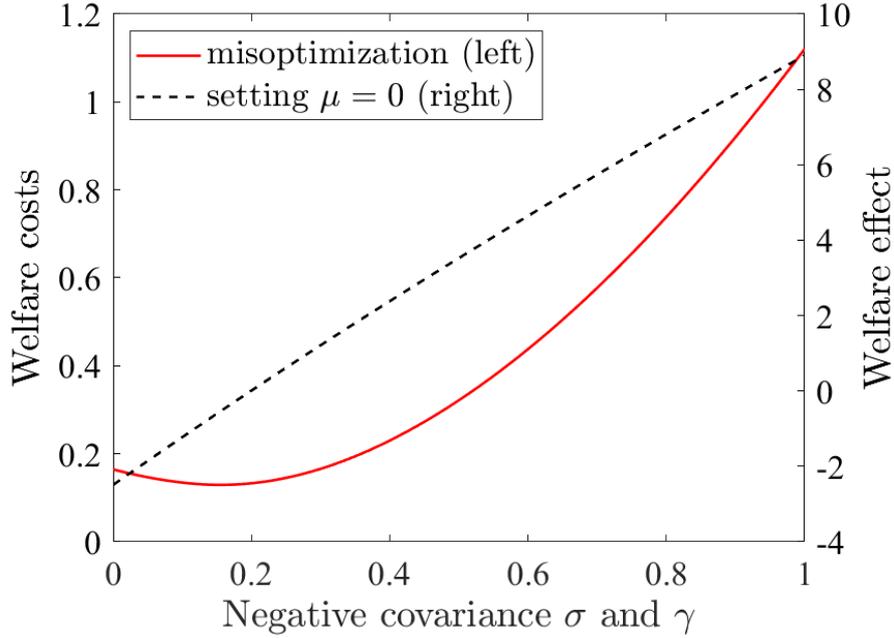


Figure 3: Welfare impact in consumption equivalents (% of GDP)

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 \$130, or 0.16% 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 \$889, or 1.12% of GDP. It should be emphasized that these figures are upper bounds for the possible welfare gains that can be reaped from taking monopsony power into account when designing tax policy. The reason is that the degree of monopsony power is calibrated to match the share of pure economic profits in GDP. If these profits come from other sources as well, monopsony power is lower in the actual than in the calibrated economy.

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 \$1,977 in consumption equivalents, or 2.49% of GDP. This loss occurs because the reduction in monopsony power exacerbates inequality in labor market payoffs and the government does not value the associated reduction in capital income inequality. By contrast, the welfare impact of getting rid of monopsony power is positive if the government has a preference for redistributing capital income. In the calibrated economy, this happens whenever $\Sigma \geq 0.20$. If $\Sigma = 1$, the welfare gain of firms losing monopsony power equals \$7,061 in consumption equivalents, or 8.90% of GDP. Again, it should be emphasized that these numbers are upper bounds for the potential welfare effects associated with changing the degree of monopsony power. As explained before, the reason is that in the calibrated economy, *all* economic profits are attributed to monopsony power. If the actual degree of monopsony power is lower than in the calibrated economy, then so will be the welfare effects from getting rid of monopsony power.

6 Conclusion

This paper extends [Mirrlees' \(1971\)](#) non-linear tax framework with monopsony power and studies the implications for optimal income taxation and welfare. Unlike the government, firms observe workers' abilities and monopsony power determines what share of the labor market surplus is translated into profits. In the baseline, monopsony power does not generate efficiency losses and profit taxes are non-distortionary. These assumptions are subsequently relaxed. Lastly, I calibrate the baseline version of the model to the US economy.

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 preferences for redistribution. In the typical case where the government wishes to redistribute both labor and capital income, optimal marginal tax rates with monopsony power are higher (lower) at low (high) levels of labor earnings than would be the case if labor markets are perfectly competitive. In that sense, monopsony power makes the optimal tax schedule less progressive. In the calibrated economy where all economic profits are attributed to monopsony power, the welfare benefits of taking monopsony power into account when designing tax policy range between 0.13% and 1.12% of GDP depending on the covariance between welfare weights and shareholdings.

Monopsony power has an ambiguous impact on welfare. On the one hand, it 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. This is because monopsony power reduces inequality in workers' labor market payoffs generated by differences in ability. On the other hand, monopsony power exacerbates inequality in capital income generated by differences in shareholdings. Moreover, there is an additional negative welfare effect of monopsony power if it distorts labor participation or induces firms to engage in costly profit shifting. The calibration exercise suggests that getting rid of monopsony power has a welfare effect that ranges between -2.49% and $+8.90\%$ of GDP in the baseline economy, again depending on the covariance between welfare weights and shareholdings. It should be emphasized that these figures are clear upper bounds, because the calibration attributes all pure economic profits to monopsony power. If these profits come from other sources as well, the welfare effects are smaller.

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). I conjecture that adding these features 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 above-normal profits if firms have monopsony power and (ii) monopsony power has an ambiguous impact on welfare. Second, there is no meaningful role for firm heterogeneity: firms operate an identical

technology and are matched exogenously to workers who differ in their ability and may suffer more or less from monopsony. However, in reality 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 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.
- Ales, L. and C. Sleet (2016). Taxing top CEO incomes. *American Economic Review* 106(11), 3331–3366.
- Arnold, D. (2021). Mergers and acquisitions, local labor market concentration, and worker outcomes. mimeo, Princeton University.
- Atesagaoglu, O. and H. Yazici (2022). Optimal fiscal policy in the presence of declining labor share. mimeo, Istanbul Bilgi University.
- 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 (2022). Labor market concentration. *Journal of Human Resources* 57(S), S167–S199.
- Azar, J., I. Marinescu, M. Steinbaum, and B. Taska (2020). Concentration in US labor markets: Evidence from online vacancy data. *Labour Economics* 66, 101886.
- Bachmann, R., G. Demir, and H. Frings (2022). Labor market polarization, job tasks, and monopsony power. *Journal of Human Resources* 57(S), S11–S49.
- Barkai, S. (2020). Declining labor and capital shares. *Journal of Finance* 75(5), 2421–2463.
- 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 (2022). Strong employers and weak employees: How does employer concentration affect wages? *Journal of Human Resources* 57(S), S200–S250.
- Berger, D., K. Herkenhoff, and S. Mongey (2022). Labor market power. *American Economic Review* 112(4), 1147–93.

- Bourguignon, F. and A. Spadaro (2012). Tax-benefit revealed social preferences. *Journal of Economic Inequality* 10(1), 75–108.
- Burdett, K. and D. Mortensen (1998). Wage differentials, employer size, and unemployment. *International Economic Review* 39(2), 257–273.
- 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 (2021). Outside options in the labor market. mimeo, UC Berkeley.
- 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. (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. (2021). Optimal income taxation with spillovers from employer learning. *American Economic Journal: Economic Policy*. Forthcoming.
- 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.
- Dube, A., J. Jacobs, S. Naidu, and S. Suri (2020). Monopsony in online labor markets. *American Economic Review: Insights* 2(1), 33–46.
- Eeckhout, J., C. Fu, W. Li, and X. Weng (2021). Optimal taxation and market power. mimeo, UPF Barcelona.
- Falch, T. (2010). The elasticity of labor supply at the establishment level. *Journal of Labor Economics* 28(2), 237–266.
- 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.
- Fuest, C., A. Peichl, and S. Siegloch (2018). Do higher corporate taxes reduce wages? Micro evidence from Germany. *American Economic Review* 108(2), 393–418.
- Hansen, E. (2021). Optimal income taxation with labor supply responses at two margins: When is an Earned Income Tax Credit optimal? *Journal of Public Economics* 195, 104365.
- 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.
- Hirsch, B., T. Schank, and C. Schnabel (2010). Differences in labor supply to monopsonistic firms and the gender pay gap: An empirical analysis using linked employer-employee data from Germany. *Journal of Labor Economics* 28(2), 291–330.
- Hummel, A. J. (2021). Tax curvature. CESifo Working Paper No. 9220, Munich: CESifo.
- 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. (2021). Market power and income taxation. *American Economic Journal: Economic Policy* 13(4), 329–54.
- 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.

- Kushnir, A. and R. Zubrickas (2020). Optimal income taxation with endogenous prices. mimeo, Carnegie Mellon University.
- 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.
- Manning, A. (2004). Monopsony and the efficiency of labour market interventions. *Labour Economics* 11(2), 145–163.
- 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.
- Ransom, M. and D. Sims (2010). Estimating the firm’s labor supply curve in a “new monopsony” framework: Schoolteachers in Missouri. *Journal of Labor Economics* 28(2), 331–355.
- Rinz, K. (2022). Labor market concentration, earnings, and inequality. *Journal of Human Resources* 57(S), S251–S283.
- 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., J. Slemrod, and S. Giertz (2012). The elasticity of taxable income with respect to marginal tax rates: a critical review. *Journal of Economic Literature* 50(1), 3–50.
- 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 (2022). Employer concentration and outside options. mimeo, UCLA Anderson School of Management.

- Seade, J. (1977). On the shape of optimal tax schedules. *Journal of Public Economics* 7(2), 203–235.
- Sleet, C. and H. Yazici (2017). Taxation, redistribution and frictional labor supply. mimeo, Carnegie Mellon University.
- Song, J., D. Price, F. Guvenen, N. Bloom, and T. Von Wachter (2019). Firming up inequality. *Quarterly Journal of Economics* 134(1), 1–50.
- Staiger, D., J. Spetz, and C. Phibbs (2010). Is there monopsony in the labor market? Evidence from a natural experiment. *Journal of Labor Economics* 28(2), 211–236.
- Stantcheva, S. (2014). Optimal income taxation with adverse selection in the labour market. *Review of Economic Studies* 81(3), 1296–1329.
- Thoresson, A. (2021). Employer concentration and wages for specialised workers. mimeo, Uppsala University.
- 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 Relation $\mu(n)$ and the Nash bargaining weight

This appendix demonstrates that if the tax function $T(\cdot)$ is linear, $\mu(n)$ correspond to the bargaining power of firms from the generalized Nash bargaining solution. The bargaining power of workers with ability n is then given by $1 - \mu(n)$.

Suppose the tax function is linear: $T(z) = -q + tz$ for $z > 0$. Furthermore, denote by $-T(0)$ the payoff a worker obtains when she rejects the contract offered by the firm. Under Nash bargaining, the equilibrium outcomes can be found by solving

$$\max_{l,z} \Gamma(n) = \mu(n) \log(nl - z) + (1 - \mu(n)) \log(z(1 - t) + q - \phi(l) - (-T(0))), \quad (37)$$

where $\mu(n)$ is the bargaining power of firms and $1 - \mu(n)$ the bargaining power of workers with ability n . Note that the worker's payoff is taken in deviation from the payoff of not working.

Denote the solution by $l(n)$ and $z(n)$. These are pinned down by the first-order conditions

$$l : \frac{\mu(n)}{\pi(n)} n - \frac{1 - \mu(n)}{v(n) + T(0)} \phi'(l(n)) = 0, \quad (38)$$

$$z : -\frac{\mu(n)}{\pi(n)} + \frac{1 - \mu(n)}{v(n) + T(0)} (1 - t) = 0. \quad (39)$$

Combining gives

$$\phi'(l(n)) = n(1 - t), \quad (40)$$

which corresponds to equation (4) if the tax function $T(\cdot)$ is linear.

Differentiating the labor market payoff $v(n) = z(n)(1 - t) + q - \phi(l(n))$ with respect to ability gives

$$v'(n) = (1 - t)z'(n) - \phi'(l(n))l'(n) = \frac{\phi'(l(n))}{n} [z'(n) - nl'(n)] = \frac{\phi'(l(n))}{n} [l(n) - \pi'(n)], \quad (41)$$

where I used equation (40) and the definition for profits $\pi(n) = nl(n) - z(n)$. Next, rewrite equation (39):

$$\mu(n)(v(n) + T(0)) = (1 - \mu(n))\pi(n)(1 - t). \quad (42)$$

Differentiate both sides with respect to ability:

$$\mu'(n)(v(n) + T(0)) + \mu(n)v'(n) = (1 - \mu(n))\pi'(n)(1 - t) - \mu'(n)\pi(n)(1 - t). \quad (43)$$

Inserting $(1 - t) = \phi'(l(n))/n$ and rewriting

$$\mu(n)v'(n) = (1 - \mu(n))\frac{\phi'(l(n))}{n}\pi'(n) - \mu'(n) \left[(v(n) + T(0)) + \frac{\phi'(l(n))}{n}\pi(n) \right]. \quad (44)$$

Equation (38), in turn, implies

$$(v(n) + T(0)) + \frac{\phi'(l(n))}{n}\pi(n) = \frac{1}{\mu(n)} \frac{\phi'(l(n))}{n}\pi(n). \quad (45)$$

Substituting back into equation (44):

$$\mu(n)v'(n) = (1 - \mu(n))\frac{\phi'(l(n))}{n}\pi'(n) - \frac{\mu'(n)}{\mu(n)} \frac{\phi'(l(n))}{n}\pi(n). \quad (46)$$

Combining with equation (41) gives:

$$v'(n) = \frac{\phi'(l(n))}{n} \left[(1 - \mu(n))l(n) - \frac{\mu'(n)}{\mu(n)}\pi(n) \right], \quad (47)$$

$$\pi'(n) = \mu(n)l(n) + \frac{\mu'(n)}{\mu(n)}\pi(n). \quad (48)$$

Integrating equation (48) and imposing the boundary condition $\pi(n_0) = 0$ gives the following expression for profits:

$$\pi(n) = \mu(n) \int_{n_0}^n l(m)dm, \quad (49)$$

which coincides with equation (7) from the paper. Hence, if the tax function is linear, $\mu(n)$ corresponds to the firm's bargaining weight from the generalized Nash bargaining solution.

If the tax function is non-linear, $\mu(n)$ no longer coincides with the Nash bargaining weight. This can be seen from inspecting equation (39). With a non-linear tax function, t is replaced by $T'(z(n))$. Differentiating with respect to ability then yields an additional term proportional to $T''(z(n))z'(n)$. This additional term significantly complicates the optimal tax problem if the degree of monopsony power is defined as the Nash bargaining parameter. To keep the optimal tax problem tractable, I therefore operationalize monopsony power in a different way, cf. Definition 1.

II 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:

$$\begin{aligned}
\mathcal{W} &= \int_{\sigma_0}^{\sigma_1} \int_{n_0}^{n_1} \gamma(n, \sigma) \mathcal{U}(n, \sigma) h(n, \sigma) dn d\sigma \\
&= \int_{\sigma_0}^{\sigma_1} \int_{n_0}^{n_1} \gamma(n, \sigma) \left[v(n) + \sigma(1 - \tau)\bar{\pi} \right] h(n, \sigma) dn d\sigma \\
&= \int_{n_0}^{n_1} v(n) \underbrace{\left(\int_{\sigma_0}^{\sigma_1} \gamma(n, \sigma) h(n, \sigma) d\sigma \right)}_{= g(n)f(n)} dn + (1 - \tau)\bar{\pi} \int_{\sigma_0}^{\sigma_1} \int_{n_0}^{n_1} \sigma \gamma(n, \sigma) h(n, \sigma) dn d\sigma \\
&= \int_{n_0}^{n_1} g(n)v(n)f(n)dn + (1 - \tau)\bar{\pi} \underbrace{\left(1 + \int_{\sigma_0}^{\sigma_1} \int_{n_0}^{n_1} (\sigma - 1)(\gamma(n, \sigma) - 1)h(n, \sigma) dn d\sigma \right)}_{= \text{Cov}[\sigma, \gamma] = -\Sigma} \\
&= \int_{n_0}^{n_1} g(n)v(n)f(n)dn + (1 - \tau)(1 - \Sigma) \int_{n_0}^{n_1} \pi(n)f(n)dn, \tag{50}
\end{aligned}$$

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

$$\Sigma = - \int_{\sigma_0}^{\sigma_1} \int_{n_0}^{n_1} (\sigma - 1)(\gamma(n, \sigma) - 1)h(n, \sigma) dn d\sigma = 1 - \int_{\sigma_0}^{\sigma_1} \int_{n_0}^{n_1} \sigma \gamma(n, \sigma) h(n, \sigma) dn d\sigma. \tag{51}$$

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

$$\begin{aligned}
\Sigma &= \int_{\sigma_0}^{\sigma_1} \int_{n_0}^{n_1} (1 - \sigma)\gamma(n, \sigma)h(n, \sigma) dn d\sigma = \int_{\sigma_0}^{\sigma_1} (1 - \sigma) \int_{n_0}^{n_1} \gamma(n, \sigma)h(n, \sigma) dn d\sigma \\
&= \int_{\sigma_0}^{\sigma_1} (1 - \sigma) \underbrace{\left(\frac{\int_{n_0}^{n_1} \gamma(n, \sigma)h(n, \sigma) dn}{\int_{n_0}^{n_1} h(n, \sigma) dn} \right)}_{= \mathbb{E}[\gamma(n, \sigma)|\sigma]} \underbrace{\left(\int_{n_0}^{n_1} h(n, \sigma) dn \right)}_{= \hat{h}(\sigma)} d\sigma. \tag{52}
\end{aligned}$$

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

$$\Sigma = \int_{\sigma_0}^{\sigma_1} (1 - \sigma)\mathbb{E}[\gamma(n, \sigma)|\sigma]\hat{h}(\sigma)d\sigma \geq \int_{\sigma_0}^{\sigma_1} (1 - \sigma)\hat{h}(\sigma)d\sigma = 0. \tag{53}$$

III 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

$$\int_{n_0}^{n_1} nl(n)f(n)dn = \int_{n_0}^{n_1} \left[v(n) + \phi(l(n)) + (1 - \tau)\pi(n) \right] f(n)dn + G. \quad (54)$$

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

$$v'(n) = (1 - T'(z(n)))z'(n) - \phi'(l(n))l'(n). \quad (55)$$

Next, use the first-order condition from the profit maximization problem (4) and the relationship $\pi(n) = nl(n) - z(n)$. Condition (55) can then be written as

$$v'(n) = \frac{\phi'(l(n))}{n} \left[l(n) - \pi'(n) \right]. \quad (56)$$

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

$$\pi'(n) = \mu(n)l(n) + \frac{\mu'(n)}{\mu(n)}\pi(n). \quad (57)$$

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 (56) and (57) gives

$$v'(n) = \frac{\phi'(l(n))}{n} \left[(1 - \mu(n))l(n) - \frac{\mu'(n)}{\mu(n)}\pi(n) \right]. \quad (58)$$

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: $v'(n) \geq 0$.³⁴ The labor market payoff does not

³⁴From equation (58), 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.

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, $v'(n) = 0$.

The government's problem consists of choosing the allocation variables $v(n)$, $\pi(n)$ and $l(n)$ at each ability level n to maximize welfare (9), subject to the resource constraint (54) and the incentive constraints (57) – (58). As it turns out, it is important to take the non-negativity constraint $l(n) \geq 0$ explicitly into account.³⁵ 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.³⁶ It is shown in Appendix VI 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 $v(n)$ 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 V.

To make sure that the optimal allocation (as implicitly characterized in Appendix V) 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 IV for details.

IV 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 $l = \hat{l}(z, v(n))$, where $v(n) = \underline{v}(n)$ for all n . The profit maximization problem is

$$\max_{z \geq 0} n\hat{l}(z, v(n)) - z. \quad (59)$$

³⁵To ensure consumption is non-negative, one could also include the constraint $v(n) + \phi(l(n)) \geq 0$ for all n . I assume the revenue requirement G and preferences for redistribution are such that this constraint never binds.

³⁶To see why, note that the general solution to the differential equation (57) 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 .

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

$$\frac{n(1 - T'(z))}{\phi'(\hat{l}(z, v(n)))} - 1 = 0. \quad (60)$$

The second-order condition is strictly satisfied if the left-hand side of equation (60) is strictly decreasing in earnings z . The latter is true if and only if

$$-\phi''(l) - n^2 T''(z) < 0, \quad (61)$$

where I used the first-order condition (60) and substituted out for $\hat{l}(z, v(n)) = l$. Because $\phi(\cdot)$ is strictly convex, condition (61) is satisfied as long as the tax function is not too concave.

To determine how earnings z vary with ability, rewrite equation (60) and define

$$L(z, n) \equiv n(1 - T'(z)) - \phi'(\hat{l}(z, v(n))) = 0. \quad (62)$$

Next, apply the implicit function theorem and use the first-order condition (60) and the property $\hat{l}_v = -1/\phi'$ to find

$$z'(n) = -\frac{L_n(z, n)}{L_z(z, n)} = \frac{\phi'(l) + \frac{\phi''(l)}{\phi'(l)} n v'(n)}{\phi''(l) + n^2 T''(z)}. \quad (63)$$

From the incentive constraint (58), $v'(n) \geq 0$ as long as monopsony power is not too quickly increasing in ability (which is assumed throughout). The numerator in (63) 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 (61) is satisfied. Therefore, if the allocation satisfies the monotonicity condition $z'(n) > 0$, it follows that the first-order condition for profit maximization (60) is both necessary and sufficient.

V Lagrangian and first-order conditions

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

$$\begin{aligned} & \max_{[v(n), \pi(n), l(n)]_{n_0}^{n_1}} \mathcal{W} = \int_{n_0}^{n_1} \left[g(n)v(n) + (1 - \Sigma)(1 - \tau)\pi(n) \right] f(n) dn, & (64) \\ \text{s.t.} & \int_{n_0}^{n_1} \left[nl(n) - v(n) - \phi(l(n)) - (1 - \tau)\pi(n) \right] f(n) dn = G, \\ \forall n : & v'(n) = \frac{\phi'(l(n))}{n} \left[(1 - \mu(n))l(n) - \frac{\mu'(n)}{\mu(n)}\pi(n) \right], \\ \forall n : & \pi'(n) = \mu(n)l(n) + \frac{\mu'(n)}{\mu(n)}\pi(n), \\ \forall n : & l(n) \geq 0, \\ & \pi(n_0) \geq 0. \end{aligned}$$

The corresponding Lagrangian is given by

$$\begin{aligned} \mathcal{L} = & \int_{n_0}^{n_1} \left[\left(g(n)v(n) + (1 - \Sigma)(1 - \tau)\pi(n) + \eta \left(nl(n) - v(n) - \phi(l(n)) - (1 - \tau)\pi(n) - G \right) \right) f(n) \right. \\ & + \chi(n) \frac{\phi'(l(n))}{n} \left((1 - \mu(n))l(n) - \frac{\mu'(n)}{\mu(n)}\pi(n) \right) + \chi'(n)v(n) + \lambda(n) \left(\mu(n)l(n) + \frac{\mu'(n)}{\mu(n)}\pi(n) \right) \\ & \left. + \lambda'(n)\pi(n) + \psi(n)l(n) \right] dn + \chi(n_0)v(n_0) - \chi(n_1)v(n_1) + \lambda(n_0)\pi(n_0) - \lambda(n_1)\pi(n_1) + \xi\pi(n_0). \end{aligned} \quad (65)$$

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

$$v(n) : (g(n) - \eta) f(n) + \chi'(n) = 0, \quad (66)$$

$$\pi(n) : (1 - \tau)(1 - \Sigma - \eta) f(n) - \frac{\mu'(n)}{\mu(n)} \left(\chi(n) \frac{\phi'}{n} - \lambda(n) \right) + \lambda'(n) = 0, \quad (67)$$

$$\begin{aligned} l(n) : \eta (n - \phi') f(n) + \frac{\chi(n)}{n} \left((1 - \mu(n))(\phi' + \phi''l(n)) - \phi'' \frac{\mu'(n)}{\mu(n)} \pi(n) \right) \\ + \lambda(n)\mu(n) + \psi(n) = 0, \end{aligned} \quad (68)$$

$$\chi(n) : \frac{\phi'}{n} \left((1 - \mu(n))l(n) - \frac{\mu'(n)}{\mu(n)}\pi(n) \right) - v'(n) = 0, \quad (69)$$

$$\lambda(n) : \mu(n)l(n) + \frac{\mu'(n)}{\mu(n)}\pi(n) - \pi'(n) = 0, \quad (70)$$

$$\eta : \int_{n_0}^{n_1} (nl(n) - v(n) - \phi(l(n)) - (1 - \tau)\pi(n) - G) f(n) dn = 0, \quad (71)$$

$$v(n_0) : \chi(n_0) = 0, \quad (72)$$

$$v(n_1) : -\chi(n_1) = 0, \quad (73)$$

$$\pi(n_0) : \lambda(n_0) + \xi = 0, \quad (74)$$

$$\pi(n_1) : -\lambda(n_1) = 0, \quad (75)$$

$$\psi(n) : \psi(n)l(n) = 0, \quad \psi(n) \geq 0 \text{ and } l(n) \geq 0, \quad (76)$$

$$\xi : \xi\pi(n_0) = 0, \quad \xi \geq 0 \text{ and } \pi(n_0) \geq 0. \quad (77)$$

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 $l(n) > 0$.

VI 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 (66) and (73) gives

$$\chi(n) = \chi(n_1) - \int_n^{n_1} \chi'(m) dm = - \int_n^{n_1} (\eta - g(m)) f(m) dm. \quad (78)$$

Evaluate equation (78) at $n = n_0$ and use the transversality condition (72) and the normalization $\int_{n_0}^{n_1} g(n)f(n)dn = 1$ to find

$$\int_{n_0}^{n_1} (\eta - g(n)) f(n)dn = \eta - 1 = 0. \quad (79)$$

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

$$\bar{g}(n) = \frac{\int_n^{n_1} g(m)f(m)dm}{1 - F(n)} \quad (80)$$

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 (67):

$$\lambda'(n) + \frac{\mu'(n)}{\mu(n)}\lambda(n) = (1 - \tau)\Sigma f(n) - \frac{\mu'(n)}{\mu(n)} \frac{\phi'(l(n))}{n} \int_n^{n_1} (1 - g(m))f(m)dm, \quad (81)$$

where I used equations (78) and (79) to substitute out for $\chi(n)$ and η . Equation (81) is a linear differential equation in $\lambda(n)$. Using the transversality condition (75), the solution is

$$\lambda(n) = -\frac{\bar{\mu}(n)}{\mu(n)}(1 - \tau)\Sigma(1 - F(n)) + \int_n^{n_1} \frac{\mu'(m)}{\mu(n)} \frac{\phi'(l(m))}{m} \int_m^{n_1} (1 - g(s))f(s)dsdm, \quad (82)$$

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 (74) and (77) 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 (68). 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

$$nf(n) - \bar{\mu}(n)(1 - \tau)\Sigma(1 - F(n)) + \int_n^{n_1} \mu'(m) \frac{\phi'(l(m))}{m} \int_m^{n_1} (1 - g(s))f(s)dsdm < 0, \quad (83)$$

where I imposed $\eta = 1$ and substituted out for $\lambda(n)$ using equation (82). The latter is true if the local Pareto parameter

$$a(n) = \frac{nf(n)}{1 - F(n)} < \bar{\mu}(n)(1 - \tau)\Sigma - \frac{\int_n^{n_1} \mu'(m) \frac{\phi'(l(m))}{m} \int_m^{n_1} (1 - g(s))f(s)dsdm}{1 - F(n)}. \quad (84)$$

Hence, at ability levels where condition (84) 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 (84) 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 (68) gives

$$T'(z(n))nf(n) = -\frac{\chi(n)}{n} \left((1 - \mu(n))(\phi' + \phi''l(n)) - \phi'' \frac{\mu'(n)}{\mu(n)} \pi(n) \right) - \mu(n)\lambda(n). \quad (85)$$

Substituting $\chi(n)$ and $\lambda(n)$ from equations (78) and (82), equation (85) can be written as

$$\begin{aligned} T'(z(n))nf(n) &= (1 - \bar{g}(n))(1 - F(n)) \frac{\phi'}{n} \left[(1 - \mu(n)) \left(1 + \frac{\phi''l(n)}{\phi'} \right) - \pi(n) \frac{\phi''}{\phi'} \frac{\mu'(n)}{\mu(n)} \right] \\ &+ \bar{\mu}(n)(1 - \tau)\Sigma(1 - F(n)) - \int_n^{n_1} \mu'(m) \frac{\phi'(l(m))}{m} \left(\int_m^{n_1} (1 - g(s))f(s)ds \right) dm. \end{aligned} \quad (86)$$

Next, use the condition $n(1 - T') = \phi'$ and denote by $\varepsilon(n) = \frac{\phi'}{\phi''l(n)}$ the elasticity of labor supply. The latter measures the percentage increase in labor effort $l(n)$ following a one percent increase in the net-of-tax rate $1 - T'(z(n))$: see equation (4). Upon dividing equation (86) by $nf(n)$ and rearranging, we obtain equation (18) from Proposition 3:

$$\begin{aligned} T'(z(n)) &= \frac{1 - F(n)}{nf(n)} \left[\bar{\mu}(n)(1 - \tau)\Sigma + (1 - \mu(n))(1 - T'(z(n)))(1 + 1/\varepsilon(n))(1 - \bar{g}(n)) \right. \\ &\left. - \frac{\mu'(n)\pi(n)(1 - T'(z(n)))(1 - \bar{g}(n))}{\mu(n)\varepsilon(n)l(n)} - \frac{\int_n^{n_1} \mu'(m)(1 - T'(z(m))) \left(\int_m^{n_1} (1 - g(s))f(s)ds \right) dm}{1 - F(n)} \right]. \end{aligned} \quad (87)$$

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 (87) 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.

VII 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:

$$T'(z(n)) = \frac{\mu(1 - \tau)\Sigma + (1 - \mu)(1 + 1/\varepsilon)(1 - \bar{g}(n))}{a(n) + (1 - \mu)(1 + 1/\varepsilon)(1 - \bar{g}(n))}. \quad (88)$$

This is a closed-form solution for the optimal marginal tax rate. To determine how the latter varies with monopsony power, differentiate equation (88) with respect to μ to find

$$\frac{\partial T'(z(n))}{\partial \mu} = \frac{a(n)((1 - \tau)\Sigma - (1 + 1/\varepsilon)(1 - \bar{g}(n))) + (1 - \tau)\Sigma(1 + 1/\varepsilon)(1 - \bar{g}(n))}{(a(n) + (1 - \mu)(1 + 1/\varepsilon)(1 - \bar{g}(n)))^2}. \quad (89)$$

Equation (89) 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 monopsony 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 (89) by $a(n)(1 - \tau)\Sigma(1 - \bar{g}(n)) > 0$. The resulting expression is positive if and only if

$$((1 - \tau)\Sigma)^{-1} < ((1 + 1/\varepsilon)(1 - \bar{g}(n)))^{-1} + a(n)^{-1}. \quad (90)$$

VIII 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

$$\begin{aligned} \mathcal{L}(\alpha) = & \int_{n_0}^{n_1} \left[\left((g(n) - \eta)v(n) + (1 - \Sigma - \eta)(1 - \tau)\pi(n) + \eta(nl(n) - \phi(l(n))) - G \right) f(n) \right. \\ & + \chi'(n)v(n) + \chi(n)\frac{\phi'(l(n))}{n} \left((1 - \mu(n)(1 + \alpha))l(n) - \frac{\mu'(n)}{\mu(n)}\pi(n) \right) + \lambda'(n)\pi(n) \\ & + \lambda(n) \left(\mu(n)(1 + \alpha)l(n) + \frac{\mu'(n)}{\mu(n)}\pi(n) \right) + \psi(n)l(n) \left. \right] dn + \chi(n_0)v(n_0) - \chi(n_1)v(n_1) \\ & + \lambda(n_0)\pi(n_0) - \lambda(n_1)\pi(n_1) + \xi\pi(n_0), \end{aligned} \quad (91)$$

which is the optimized Lagrangian (65) evaluated at $\hat{\mu}(n) = \mu(n)(1 + \alpha)$. Here I used the fact that the increase in monopsony power is proportional, which implies

$$\frac{\hat{\mu}'(n)}{\hat{\mu}(n)} = \frac{\mu'(n)(1 + \alpha)}{\mu(n)(1 + \alpha)} = \frac{\mu'(n)}{\mu(n)}. \quad (92)$$

By the envelope theorem, the welfare effect is

$$\frac{\partial \mathcal{W}(\alpha)}{\partial \alpha} = \frac{\partial \mathcal{L}(\alpha)}{\partial \alpha} = \int_{n_0}^{n_1} \left(-\chi(n)\frac{\phi'}{n} + \lambda(n) \right) \mu(n)l(n)dn. \quad (93)$$

Next, use equations (78) and (82) to substitute out for $\chi(n)$ and $\lambda(n)$. This leads to

$$\begin{aligned} \frac{\partial \mathcal{W}(\alpha)}{\partial \alpha} = & \int_{n_0}^{n_1} \left[\frac{\phi'(l(n))}{n} \int_n^{n_1} (1 - g(m))f(m)dm - (1 - \tau)\Sigma \int_n^{n_1} \frac{\mu(m)}{\mu(n)}f(m)dm \right. \\ & \left. + \int_n^{n_1} \frac{\mu'(m)}{\mu(n)} \frac{\phi'(l(m))}{m} \left(\int_m^{n_1} (1 - g(s))f(s)ds \right) dm \right] \mu(n)l(n)dn, \end{aligned} \quad (94)$$

which coincides with equation (19) from Proposition 3 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 (94) cancels. Substituting $\mu(n) = \mu$ gives

$$\frac{\partial \mathcal{W}(\alpha)}{\partial \alpha} = \quad (95)$$

$$\int_{n_0}^{n_1} \left[\frac{\mu}{1-\mu} \underbrace{(1-\mu) \frac{\phi'(l(n))l(n)}{n}}_{=v'(n)} \int_n^{n_1} (1-g(m))f(m)dm - (1-\tau)\Sigma \underbrace{\mu l(n)}_{=\pi'(n)} \int_n^{n_1} f(m)dm \right] dn.$$

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

$$\frac{\partial \mathcal{W}(\alpha)}{\partial \alpha} = \int_{n_0}^{n_1} \left[\frac{\mu}{1-\mu} (1-g(n))v(n) - (1-\tau)\Sigma\pi(n) \right] f(n)dn. \quad (96)$$

The latter can be simplified further after defining

$$\Sigma^v = -\text{Cov}[v, \gamma] \geq 0, \quad (97)$$

$$\Sigma^k = -\text{Cov}[\sigma(1-\tau)\bar{\pi}, \gamma] = \Sigma(1-\tau)\bar{\pi} \geq 0. \quad (98)$$

The first measures the negative covariance between labor market payoffs $v(n)$ 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 (96) gives

$$\frac{\partial \mathcal{W}(\alpha)}{\partial \alpha} = \frac{\mu}{1-\mu} \Sigma^v - \Sigma^k. \quad (99)$$

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)

$$\mu \Sigma^v > (1-\mu) \Sigma^k. \quad (100)$$

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:

$$T'(z(n)) = \frac{\mu(1-\tau)\Sigma + (1-\mu)(1+1/\varepsilon)(1-\bar{g}(n))}{a(n) + (1-\mu)(1+1/\varepsilon)(1-\bar{g}(n))}, \quad (101)$$

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

$$l(n) = n^\varepsilon \left(\frac{a(n) - \mu(1-\tau)\Sigma}{a(n) + (1-\mu)(1+1/\varepsilon)(1-\bar{g}(n))} \right)^\varepsilon \quad (102)$$

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 (94) and setting $\mu(n) = \mu$ and hence, $\mu'(n) = 0$ gives

$$\frac{\partial \mathcal{W}(\alpha)}{\partial \alpha} = \int_{n'}^{n_1} \mu \left[\left(\frac{a(n) - \mu(1-\tau)\Sigma}{a(n) + (1-\mu)(1+1/\varepsilon)(1-\bar{g}(n))} \right) (1-\bar{g}(n)) - (1-\tau)\Sigma \right]$$

$$\times (1 - F(n))n^\varepsilon \left(\frac{a(n) - \mu(1 - \tau)\Sigma}{a(n) + (1 - \mu)(1 + 1/\varepsilon)(1 - \bar{g}(n))} \right)^\varepsilon dn, \quad (103)$$

which is expressed solely in terms of exogenous variables.

Proof Corollary 2

To derive the result from Corollary 2, note that equation (100) gives a necessary and sufficient condition to determine if an increase in monopsony power raises welfare. Next, write

$$\begin{aligned} \Sigma^v &= \int_{n_0}^{n_1} (1 - g(n))v(n)f(n)dn = \int_{n_0}^{n_1} (1 - g(n))(z(n) - T(z(n)) - \phi(l(n)))f(n)dn \\ &= \int_{n_0}^{n_1} (1 - g(n))(z(n) - T(z(n)))f(n)dn - \int_{n_0}^{n_1} (1 - g(n))\phi(l(n))f(n)dn \\ &= -\text{Cov}[z - T(z), \gamma] - \int_{n_0}^{n_1} (1 - g(n))\phi(l(n))f(n)dn \\ &= \Sigma^\ell - \int_{n_0}^{n_1} (1 - g(n))\phi(l(n))f(n)dn. \end{aligned} \quad (104)$$

Because $g(n)$ is weakly decreasing in ability and averages to one, the second term on the last line of equation (104) 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

$$\mu\Sigma^\ell > (1 - \mu)\Sigma^k. \quad (105)$$

Unlike equation (100), this condition is necessary but not sufficient.

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 III. 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 (20).

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 II:

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

$$\begin{aligned}
& + (1 - \tau)\bar{\pi} \int_{n_0}^{n_1} \int_{\sigma_0}^{\sigma_1} \int_{\varphi_0}^{\varphi_1} \gamma(n, \sigma, \varphi) \sigma h(n, \sigma, \varphi) d\varphi d\sigma dn \\
& = \int_{n_0}^{n_1} \left[\int_{\varphi_0}^{\varphi(n)} g(n, \varphi) (v(n) - \varphi) k(n, \varphi) d\varphi + \int_{\varphi(n)}^{\varphi_1} g(n, \varphi) b k(n, \varphi) d\varphi \right] dn \\
& + (1 - \tau) \int_{n_0}^{n_1} \int_{\varphi_0}^{\varphi(n)} \pi(n) k(n, \varphi) d\varphi dn \int_{n_0}^{n_1} \int_{\sigma_0}^{\sigma_1} \int_{\varphi_0}^{\varphi_1} \gamma(n, \sigma, \varphi) \sigma h(n, \sigma, \varphi) d\varphi d\sigma dn \\
& = \int_{n_0}^{n_1} \int_{\varphi_0}^{\varphi(n)} \left[g(n, \varphi) (v(n) - \varphi) + (1 - \tau)(1 - \Sigma)\pi(n) \right] k(n, \varphi) d\varphi dn \\
& + \int_{n_0}^{n_1} \int_{\varphi(n)}^{\varphi_1} g(n, \varphi) b k(n, \varphi) d\varphi dn. \tag{106}
\end{aligned}$$

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 :

$$g(n, \varphi) = \frac{\int_{\sigma_0}^{\sigma_1} \gamma(n, \sigma, \varphi) h(n, \sigma, \varphi) d\sigma}{\int_{\sigma_0}^{\sigma_1} h(n, \sigma, \varphi) d\sigma}, \quad k(n, \varphi) = \int_{\sigma_0}^{\sigma_1} h(n, \sigma, \varphi) d\sigma. \tag{107}$$

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

$$\int_{n_0}^{n_1} \int_{\varphi_0}^{\varphi(n)} (T(z(n)) + \tau\pi(n)) k(n, \varphi) d\varphi dn = \int_{n_0}^{n_1} \int_{\varphi(n)}^{\varphi_1} b k(n, \varphi) d\varphi dn + G. \tag{108}$$

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 $T(z(n)) = nl(n) - \pi(n) - v(n) - \phi(l(n))$:

$$\begin{aligned}
& \int_{n_0}^{n_1} \int_{\varphi_0}^{\varphi(n)} (nl(n) - v(n) - \phi(l(n)) - (1 - \tau)\pi(n)) k(n, \varphi) d\varphi dn \\
& = \int_{n_0}^{n_1} \int_{\varphi(n)}^{\varphi_1} b k(n, \varphi) d\varphi dn + G. \tag{109}
\end{aligned}$$

Using the definition for the participation rate (21), the final equation can be written as

$$\int_{n_0}^{n_1} \left[p(\varphi(n))(nl(n) - v(n) - \phi(l(n)) - (1 - \tau)\pi(n)) - (1 - p(\varphi(n)))b \right] f(n) dn = G. \tag{110}$$

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

$$f(n) = \int_{\varphi_0}^{\varphi_1} k(n, \varphi) d\varphi. \tag{111}$$

The incentive constraints are the same as before and given by equations (57)–(58). This is because firms cannot observe participation costs. Hence, as in the model without a par-

icipation 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 (106) subject to the resource constraint (109), incentive constraints (57)–(58) and the requirements $\pi(n_0) \geq 0$ and $l(n) \geq 0$ for all n . The corresponding Lagrangian is

$$\begin{aligned}
\mathcal{L} = & \int_{n_0}^{n_1} \left[\int_{\varphi_0}^{v(n)-b} \left(g(n, \varphi)(b + (v(n) - b) - \varphi) + (1 - \tau)(1 - \Sigma)\pi(n) \right. \right. \\
& + \eta \left[nl(n) - (v(n) - b) - b - \phi(l(n)) - (1 - \tau)\pi(n) - G \right] \left. \right) k(n, \varphi) d\varphi \\
& + \int_{v(n)-b}^{\varphi_1} \left(g(n, \varphi)b - \eta(b + G) \right) k(n, \varphi) d\varphi \\
& + \chi(n) \frac{\phi'(l(n))}{n} \left[(1 - \mu(n))l(n) - \frac{\mu'(n)}{\mu(n)}\pi(n) \right] + \chi'(n)(b + (v(n) - b)) \\
& + \lambda(n) \left[\mu(n)l(n) + \frac{\mu'(n)}{\mu(n)}\pi(n) \right] + \lambda'(n)\pi(n) + \psi(n)l(n) \Big] dn \\
& + \chi(n_0)(b + (v(n_0) - b)) - \chi(n_1)(b + (v(n_1) - b)) + \lambda(n_0)\pi(n_0) - \lambda(n_1)\pi(n_1) + \xi\pi(n_0),
\end{aligned} \tag{112}$$

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

Derivation of equation (22)

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

$$\int_{n_0}^{n_1} \int_{\varphi_0}^{\varphi_1} (g(n, \varphi) - \eta)k(n, \varphi) d\varphi dn + \int_{n_0}^{n_1} \chi'(n) dn + \chi(n_0) - \chi(n_1) = 0. \tag{113}$$

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 (110):

$$\int_{n_0}^{n_1} \left[p(nl(n) - v(n) - \phi(l(n)) - (1 - \tau)\pi(n)) - (1 - p)b - G \right] f(n) dn = 0. \tag{114}$$

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

$$\pi(n) : p(1 - \tau)(1 - \Sigma - \eta)f(n) - \frac{\mu'(n)}{\mu(n)} \left(\chi(n) \frac{\phi'}{n} - \lambda(n) \right) + \lambda'(n) = 0, \tag{115}$$

$$\begin{aligned}
l(n) : & p\eta(n - \phi') f(n) + \frac{\chi(n)}{n} \left((1 - \mu(n))(\phi' + \phi''l(n)) - \phi'' \frac{\mu'(n)}{\mu(n)} \pi(n) \right) \\
& + \lambda(n)\mu(n) + \psi(n) = 0.
\end{aligned} \tag{116}$$

These equations differ from the ones in Appendix V 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 $v(n)$ (or, equivalently, $v(n) - b$):

$$p \left[g_p(n) - \eta + \frac{p'}{p} \left((1 - \tau)(1 - \Sigma)\pi(n) + \eta(nl(n) - v(n) - \phi(l(n)) + b - (1 - \tau)\pi(n)) \right) \right] f(n) + \chi'(n) = 0. \quad (117)$$

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

$$g_p(n) = \frac{\int_{\varphi_0}^{v(n)-b} g(n, \varphi) k(n, \varphi) d\varphi}{\int_{\varphi_0}^{v(n)-b} k(n, \varphi) d\varphi}. \quad (118)$$

The first-order condition with respect to labor effort (116) 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

$$pT'(z(n))nf(n) = -\frac{\chi(n)}{n} \left((1 - \mu(n))(\phi' + \phi''l(n)) - \phi''\frac{\mu'(n)}{\mu(n)}\pi(n) \right) - \mu(n)\lambda(n). \quad (119)$$

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 $T(z(n)) = nl(n) - \pi(n) - v(n) - \phi(l(n))$ in equation (117) and use the transversality condition $\chi(n_1) = 0$:

$$\chi(n) = -\int_n^{n_1} p \left[1 - g_p(m) - \frac{p'}{p} (T(z(m)) + b + (1 - (1 - \tau)\Sigma)\pi(m)) \right] f(m) dm. \quad (120)$$

To derive an expression for $\lambda(n)$, substitute $\eta = 1$ in equation (115) and rearrange to find the following linear differential equation in $\lambda(n)$:

$$\lambda'(n) + \frac{\mu'(n)}{\mu(n)}\lambda(n) = p(1 - \tau)\Sigma f(n) + \frac{\mu'(n)}{\mu(n)}\frac{\phi'}{n}\chi(n). \quad (121)$$

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

$$\begin{aligned} \lambda(n) &= -\int_n^{n_1} \left[(1 - \tau)\Sigma p \frac{\mu(m)}{\mu(n)} f(m) + \frac{\mu'(m)}{\mu(n)} \frac{\phi'}{m} \chi(m) \right] dm \\ &= -(1 - \tau)\Sigma \int_n^{n_1} p \frac{\mu(m)}{\mu(n)} f(m) dm \\ &\quad + \int_n^{n_1} \frac{\mu'(m)}{\mu(n)} \frac{\phi'}{m} \left[\int_m^{n_1} p \left[1 - g_p(s) - \frac{p'}{p} (T(z(s)) + b + (1 - (1 - \tau)\Sigma)\pi(s)) \right] f(s) ds \right] dm. \end{aligned} \quad (122)$$

Substituting the solutions for $\chi(n)$ and $\lambda(n)$ in equation (119) gives, after rearranging,

$$pT'(z(n))nf(n) = (1 - \tau)\Sigma \int_n^{n_1} p\mu(m)f(m)dm + (1 - \mu(n))(1 - T'(z(n)))(1 + 1/\varepsilon(n)) \quad (123)$$

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

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

$$\begin{aligned}
p(\varphi(n))T'(z(n))nf(n) &= \tag{124} \\
\mu(1 - \tau)\Sigma \int_n^{n_1} p(\varphi(m))f(m)dm + (1 - \mu)(1 - T'(z(n)))(1 + 1/\varepsilon(n)) \\
& \times \int_n^{n_1} p(\varphi(m)) \left[1 - g_p(m) - \frac{p'(\varphi(m))}{p(\varphi(m))}(T(z(m)) + b + (1 - (1 - \tau)\Sigma)\pi(m)) \right] f(m) dm.
\end{aligned}$$

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 4:

$$\begin{aligned}
T'(z(n)) &= \frac{1 - F_p(n)}{nf_p(n)} \left[\mu(1 - \tau)\Sigma + (1 - \mu)(1 - T'(z(n)))(1 + 1/\varepsilon(n)) \right. \\
& \left. \times \mathbb{E} \left[1 - g_p(m) - \hat{p}(m)(T(z(m)) + b) - \hat{p}(m)\pi(m)(1 - (1 - \tau)\Sigma) \mid m \geq n \right] \right], \tag{125}
\end{aligned}$$

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

Derivation of equation (23)

To derive the welfare impact of a proportional increase in monopsony power from $\mu(n)$ to $\mu(n)(1 + \alpha)$, modify the Lagrangian (112) to

$$\begin{aligned}
\mathcal{L}(\alpha) &= \int_{n_0}^{n_1} \left[\int_{\varphi_0}^{v(n)-b} \left(g(n, \varphi)(b + (v(n) - b) - \varphi) + (1 - \tau)(1 - \Sigma)\pi(n) \right. \right. \tag{126} \\
& \left. \left. + \eta \left[nl(n) - (v(n) - b) - b - \phi(l(n)) - (1 - \tau)\pi(n) - G \right] \right) k(n, \varphi) d\varphi \right. \\
& \left. + \int_{v(n)-b}^{\varphi_1} \left(g(n, \varphi)b - \eta(b + G) \right) k(n, \varphi) d\varphi \right. \\
& \left. + \chi(n) \frac{\phi'(l(n))}{n} \left[(1 - \mu(n)(1 + \alpha))l(n) - \frac{\mu'(n)}{\mu(n)}\pi(n) \right] + \chi'(n)(b + (v(n) - b)) \right. \\
& \left. + \lambda(n) \left[\mu(n)(1 + \alpha)l(n) + \frac{\mu'(n)}{\mu(n)}\pi(n) \right] + \lambda'(n)\pi(n) + \psi(n)l(n) \right] dn \\
& + \chi(n_0)(b + (v(n_0) - b)) - \chi(n_1)(b + (v(n_1) - b)) + \lambda(n_0)\pi(n_0) - \lambda(n_1)\pi(n_1) + \xi\pi(n_0).
\end{aligned}$$

By the envelope theorem,

$$\frac{\partial \mathcal{W}(\alpha)}{\partial \alpha} = \frac{\partial \mathcal{L}(\alpha)}{\partial \alpha} = \int_{n_0}^{n_1} \left(-\chi(n) \frac{\phi'}{n} + \lambda(n) \right) \mu(n)l(n) dn. \tag{127}$$

Substituting in the solution for $\chi(n)$ and $\lambda(n)$ from equations (120) and (122):

$$\begin{aligned} \frac{\partial \mathcal{W}(\alpha)}{\partial \alpha} &= \int_{n_0}^{n_1} \mu(n) l(n) \left[- (1 - \tau) \Sigma \int_n^{n_1} p \frac{\mu(m)}{\mu(n)} f(m) dm + \int_n^{n_1} \frac{\mu'(m)}{\mu(n)} \frac{\phi'}{m} \right. \\ &\quad \times \left(\int_m^{n_1} p \left[1 - g_p(s) - \frac{p'}{p} (T(z(s)) + b + (1 - (1 - \tau) \Sigma) \pi(s)) \right] f(s) ds \right) dm \\ &\quad \left. + \frac{\phi'}{n} \int_n^{n_1} p \left[1 - g_p(m) - \frac{p'}{p} (T(z(m)) + b + (1 - (1 - \tau) \Sigma) \pi(m)) \right] f(m) dm \right] dn. \end{aligned} \quad (128)$$

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

$$\begin{aligned} \frac{\partial \mathcal{W}(\alpha)}{\partial \alpha} &= \int_{n_0}^{n_1} \mu l(n) \left[- (1 - \tau) \Sigma \int_n^{n_1} p(\varphi(m)) f(m) dm + \frac{\phi'(l(n))}{n} \right. \\ &\quad \left. \times \int_n^{n_1} p(\varphi(m)) \left[1 - g_p(m) - \frac{p'(\varphi(m))}{p(\varphi(m))} (T(z(m)) + b + (1 - (1 - \tau) \Sigma) \pi(m)) \right] f(m) dm \right] dn. \end{aligned} \quad (129)$$

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

$$\begin{aligned} \frac{\partial \mathcal{W}(\alpha)}{\partial \alpha} &= - (1 - \tau) \Sigma \int_{n_0}^{n_1} \pi(n) f_p(n) dn \\ &\quad + \frac{\mu}{1 - \mu} \int_{n_0}^{n_1} v(n) \left[1 - g_p(n) - \hat{p}(n) (T(z(n)) + b + (1 - (1 - \tau) \Sigma) \pi(n)) \right] f_p(n) dn \\ &= - \Sigma^k - \frac{\mu}{1 - \mu} \int_{n_0}^{n_1} v(n) \hat{p}(n) \pi(n) (1 - (1 - \tau) \Sigma) f_p(n) dn \\ &\quad + \frac{\mu}{1 - \mu} \int_{n_0}^{n_1} v(n) \left[1 - g_p(n) - \hat{p}(n) (T(z(n)) + b) \right] f_p(n) dn, \end{aligned} \quad (130)$$

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 4.

X 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 (25), the social welfare function (9) is modified to

$$\mathcal{W} = \int_{n_0}^{n_1} \left[g(n) v(n) + (1 - \Sigma) (1 - (1 - s(\tau)) \tau - \rho(s(\tau))) \pi(n) \right] f(n) dn, \quad (131)$$

where the term $(1 - \tau) \pi(n)$ is replaced by $(1 - (1 - s(\tau)) \tau - \rho(s(\tau))) \pi(n)$: see equation (24). The government budget constraint, in turn, is given by

$$\int_{n_0}^{n_1} \left[T(z(n)) + \tau (1 - s(\tau)) \pi(n) \right] f(n) dn = G, \quad (132)$$

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 $T(z(n)) = z(n) - v(n) - \phi(l(n)) =$

$nl(n) - \pi(n) - v(n) - \phi(l(n))$:

$$\int_{n_0}^{n_1} \left[nl(n) - v(n) - \phi(l(n)) - (1 - (1 - s(\tau))\tau)\pi(n) \right] f(n)dn = G. \quad (133)$$

The incentive constraints are the same as before (see Appendix III) 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

$$\begin{aligned} \mathcal{L} = \int_{n_0}^{n_1} & \left[\left(g(n)v(n) + (1 - \Sigma)(1 - (1 - s(\tau))\tau - \rho(s(\tau)))\pi(n) \right. \right. \\ & \left. \left. + \eta \left(nl(n) - v(n) - \phi(l(n)) - (1 - (1 - s(\tau))\tau)\pi(n) - G \right) \right) f(n) \right. \\ & \left. + \chi(n)(1 - \mu) \frac{\phi'(l(n))l(n)}{n} + \chi'(n)v(n) + \lambda(n)\mu l(n) + \lambda'(n)\pi(n) + \psi(n)l(n) \right] dn \\ & + \chi(n_0)v(n_0) - \chi(n_1)v(n_1) + \lambda(n_0)\pi(n_0) - \lambda(n_1)\pi(n_1) + \xi\pi(n_0). \end{aligned} \quad (134)$$

Derivation of equation (26)

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

$$\int_{n_0}^{n_1} \left[- (1 - \Sigma)(1 - s(\tau)) + \eta (1 - s(\tau) - s'(\tau)\tau) \right] \pi(n)f(n)dn = 0, \quad (135)$$

where I used the envelope condition that s maximizes the term $1 - (1 - s)\tau - \rho(s)$. As before, the first-order condition for $v(n)$, the transversality conditions and the normalization of welfare weights imply $\eta = 1$. Next, divide equation (135) by aggregate profits $\int_{n_0}^{n_1} \pi(n)f(n)dn$ and rearrange to find

$$s'(\tau)\tau = \Sigma(1 - s(\tau)). \quad (136)$$

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 (26) from Proposition 5.

Derivation of equation (27)

The expression for the optimal marginal tax rate is obtained from the first-order condition of the Lagrangian (134) with respect to labor effort:

$$\eta(n - \phi'(l(n)))f(n) + (1 - \mu) \frac{\chi(n)}{n} (\phi'(l(n)) + \phi''(l(n))l(n)) + \mu\lambda(n) = 0. \quad (137)$$

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

$$T'(z(n))nf(n) = -(1 - \mu)(1 - T'(z(n)))(1 + 1/\epsilon(n))\chi(n) - \mu\lambda(n). \quad (138)$$

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 $\pi(n)$:

$$\left[(1 - \Sigma)(1 - (1 - s(\tau))\tau - \rho(s(\tau))) - \eta(1 - (1 - s(\tau))\tau) \right] f(n) + \lambda'(n) = 0. \quad (139)$$

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

$$\lambda(n) = - \left[(1 - (1 - s(\tau))\tau - \rho(s(\tau)))\Sigma + \rho(s) \right] (1 - F(n)). \quad (140)$$

Substituting the solutions for $\chi(n)$ and $\lambda(n)$ in equation (138) and rearranging gives equation (27) from Proposition 5.

Derivation of equation (28)

The welfare impact of raising monopsony power is:

$$\begin{aligned} \frac{\partial \mathcal{W}(\mu)}{\partial \mu} &= \frac{\partial \mathcal{L}(\mu)}{\partial \mu} = \int_{n_0}^{n_1} \left(-\chi(n) \frac{\phi'(l(n))l(n)}{n} + \lambda(n)l(n) \right) dn \\ &= \int_{n_0}^{n_1} \left(\frac{1}{1 - \mu} \int_n^{n_1} (1 - g(m))f(m)dm \underbrace{(1 - \mu) \frac{\phi'(l(n))l(n)}{n}}_{= v'(n)} \right. \\ &\quad \left. - \frac{1}{\mu} \int_n^{n_1} \left[(1 - (1 - s(\tau))\tau - \rho(s(\tau)))\Sigma + \rho(s) \right] f(m)dm \underbrace{\mu l(n)}_{= \pi'(n)} \right) dn, \end{aligned} \quad (141)$$

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 $\pi(n_0) = 0$:

$$\begin{aligned} \frac{\partial \mathcal{W}(\mu)}{\partial \mu} &= \int_{n_0}^{n_1} \left[\frac{1}{1 - \mu} (1 - g(n))v(n) - \frac{1}{\mu} \left[(1 - (1 - s(\tau))\tau - \rho(s(\tau)))\Sigma + \rho(s) \right] \pi(n) \right] f(n)dn \\ &= \frac{\Sigma^v}{1 - \mu} - \frac{\Sigma^k}{\mu} - \frac{\kappa(s(\tau))\bar{\pi}}{\mu}, \end{aligned} \quad (142)$$

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 (142) by $\mu(1 - \mu)$ and set the resulting expression larger than zero. Rearranging gives the final result from Proposition 5.

XI Additional figures

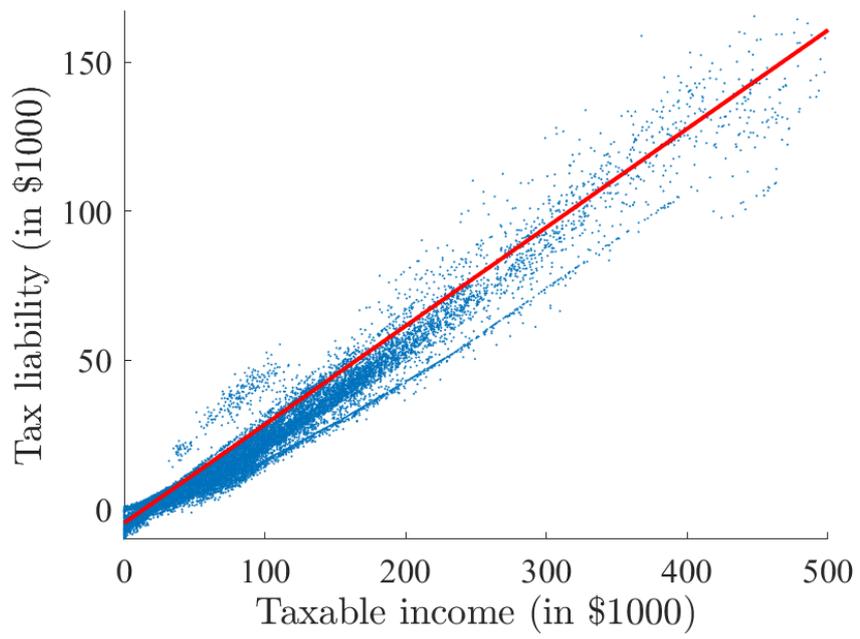


Figure 4: Current tax schedule

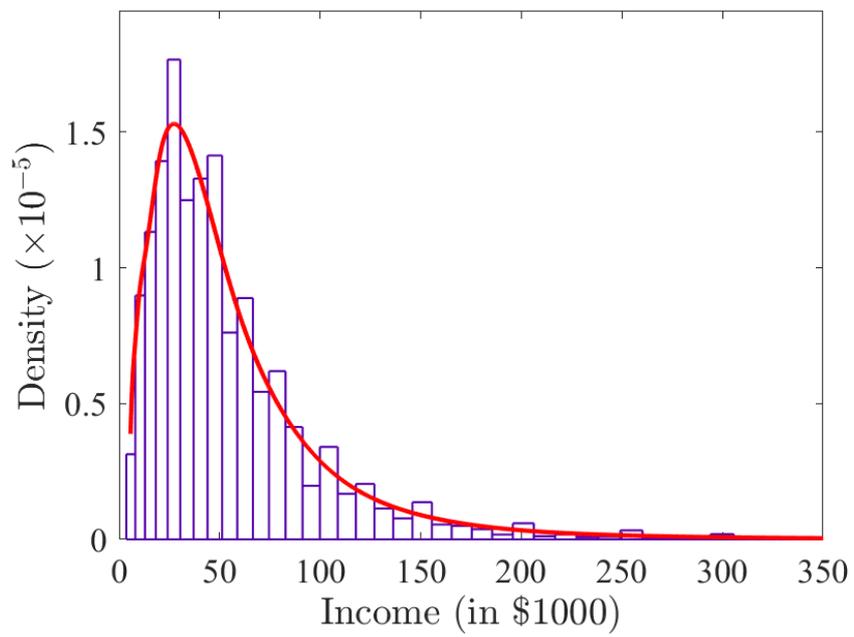
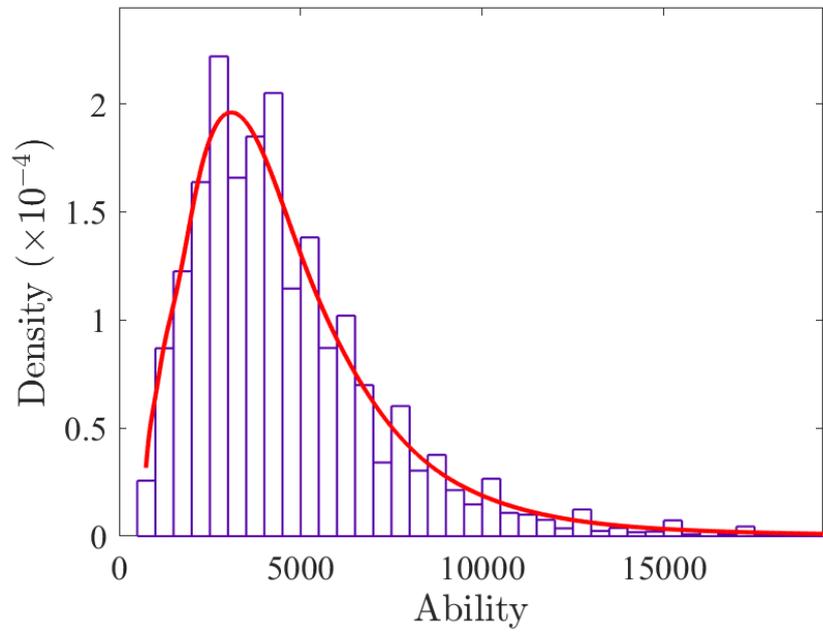


Figure 5: Distribution of ability and income

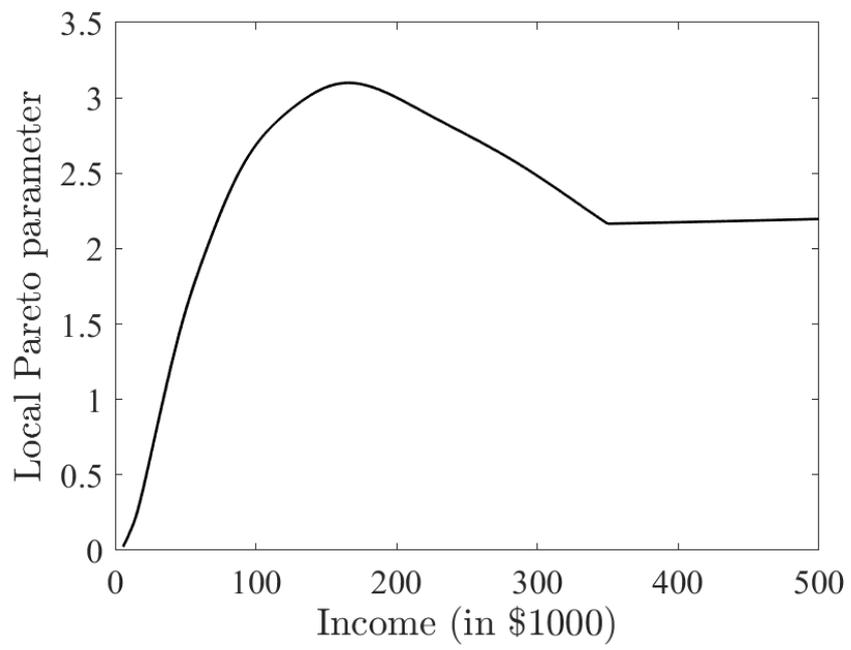


Figure 6: Local Pareto parameter