

# Monopsony power, income taxation, and welfare\*

Albert Jan Hummel<sup>†</sup>

April 30, 2026

## 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 at 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, IIPF Annual Congress, CESifo Area Conference on Public Economics and Ludwig Maximilian University for many helpful comments and suggestions.

<sup>†</sup>University of Groningen, CESifo, and Instituut voor Publieke Economie. E-mail: a.j.hummel@rug.nl.

# 1 Introduction

There is growing concern among economists and policymakers that firms exercise monopsony power in labor markets. In the US, 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.<sup>1</sup> 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), Lipsius (2018), Qiu and Sojourner (2023), Rinz (2022), Schubert et al. (2022), Thoesson (2023), Yeh et al. (2022)). 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.<sup>2</sup>

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 the model, firms observe workers' abilities while the government does not. To focus on distributional issues, monopsony power does not generate efficiency losses, but determines what share of the labor market surplus is translated into pure economic profits. After-tax profits flow back as capital income to individuals who differ in their ability and shareholdings. The government levies an exogenous tax on profits and optimizes a non-linear tax on labor income. 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 surplus that can be shared between workers and firms, 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 income. 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 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 indirectly tax profits, i.e., capital income. The reason is that part of the tax incidence falls on firms if they

---

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

<sup>2</sup>For example, Alan Krueger noted in his address at the 2018 Fed conference in Jackson Hole:

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

have monopsony power. Monopsony power thus makes taxes on labor earnings less effective at redistributing labor income but more effective at redistributing capital income. As a result, monopsony power raises (lowers) optimal marginal tax rates if the government has a strong preference for redistributing capital (labor) income. I derive a condition which can be used to determine if monopsony power raises the optimal marginal tax rate and show that this condition is more likely to be satisfied at lower income 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 effect 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 income driven by differences in ability. The associated effect on welfare is positive and proportional to the covariance between welfare weights and labor market payoffs, i.e., after-tax labor earnings minus the disutility of working. 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. This reduction in inequality comes at zero efficiency costs, which can never be achieved with distortionary taxes on labor income. By enabling the government to exploit the informational advantage of firms, monopsony power thus alleviates the equity-efficiency trade-off that occurs because the government does not observe ability. This, however, comes at the costs of higher inequality in capital income.

Two critical assumptions in the model 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 distortion in employment along the extensive margin, 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. Naturally, monopsony power has a smaller positive (or stronger negative) effect on welfare if it generates distortions in employment. I relax the second of these assumptions by analyzing an extension where profit taxes leads to costly profit shifting. Optimal tax rates on labor income are higher when firms shift profits, as they can be used to reduce wasteful profit shifting. In addition, an increase in monopsony power is more likely to reduce 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 model to the US economy. The joint distribution of ability and shareholdings is obtained from the joint distribution of labor and capital income, and the degree of monopsony power is used to target an estimate of the pure profit share from [Barkai and Benzell \(2018\)](#). I find that monopsony power significantly raises marginal tax rates at low income levels, but lowers them at middle and higher incomes. Monopsony power thus makes the optimal tax schedule less progressive. In the calibrated economy, taking monopsony power into account when designing tax policy leads to a welfare gain of 0.38% of GDP. Moreover, getting rid of monopsony power generates a welfare gain of about

3.41% of GDP. For a given choice of the welfare function, however, these figures are clear upper bounds. The reason is that, in the calibrated economy, *all* 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 labor market power. [Hariton and Piaser \(2007\)](#), [da Costa and Maestri \(2019\)](#), and [Hummel \(2023\)](#) analyze a model with labor supply responses on the intensive (hours, effort) margin, whereas [Cahuc and Laroque \(2014\)](#) study the optimal joint design of income taxes and minimum wages in a model with with labor supply responses on the extensive (participation) margin. In these studies, monopsony power leads to a downward distortion in employment, either in hours worked or the number of individuals employed. The optimal tax schedule features negative marginal (participation) tax rates if labor supply responds on the intensive (extensive) margin. To focus on distributional issues, in the baseline model I abstract from distortions in employment generated by monopsony power. In this case, marginal tax rates are generally positive and serve to distribute not only labor but also capital income, because part of the tax incidence falls on firms. In addition, I study the welfare effects of monopsony power.

Another closely related paper is [Sleet and Yazici \(2024\)](#), who study optimal income taxation in a rich frictional model that features unemployment, on and off the job search, and labor income inequality both across and within ability levels. They also find that part of the tax incidence falls on firms (the “profit squeeze” effect), which, given that profits are taxed at a confiscatory rate, calls for lower tax rates on labor income. By contrast, I study a much simpler, frictionless model where the main departure from the [Mirrlees \(1971\)](#) benchmark is that firms capture part of the labor market surplus. This allows me to characterize theoretically how a change in monopsony power affects optimal marginal tax rates and, ultimately, welfare. Another important difference is that I allow for heterogeneity in capital income. Provided profits are not fully taxed, monopsony power then raises marginal tax rates at low income levels, which makes the optimal tax schedule less progressive.

Some recent papers study optimal taxation when firms have market power in output markets. [Eeckhout et al. \(2021\)](#), [Kaplow \(2021\)](#), [Gürer \(2022\)](#), [Tarasov and Zubrickas \(2023\)](#), and [Kushnir and Zubrickas \(2023\)](#) do so in a static Mirrleesian environment with private information on types and non-linear taxes. As in the classic model of monopoly, output and employment are inefficiently low, which calls for a downward adjustment in optimal taxes. [Atesagaoglu and Yazici \(2023\)](#) study a dynamic representative-agent Ramsey setting with linear taxes. Like [Kushnir and Zubrickas \(2023\)](#), they find that taxes on labor income can be used to indirectly tax profits, a force that is also present in my model. Compared to these papers, an important difference is that I focus on market power in the labor market. In addition, to focus on the distributional consequences from monopsony power, I mostly abstract from inefficiencies that originate from market power. Tax policy is then exclusively aimed at redistribution and not to reduce distortions.

The model of labor market monopsony I analyze features important similarities and dif-

ferences 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. 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 a combination of earnings and labor effort that implies a higher hourly wage. Second, 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 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 model. Section 3 characterizes optimal tax policy and analyzes 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 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 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 (there is no market power in the market for goods). Whenever this is the case, firms earn pure economic profits. These profits are taxed linearly at an exogenous rate 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 pro-

duces per unit of effort, and shareholdings determine how much profits flow back to each individual. For simplicity, both ability and shareholdings are taken to be exogenous but they can be correlated in an arbitrary fashion. Let  $H(n, \sigma)$  denote the joint distribution of 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 exerting 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 but not with shareholdings.<sup>3</sup> I denote by  $l(n) \geq 0$  the labor effort exerted by an individual with ability  $n$ . In exchange for their services, they receive 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 therefore depends on 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 an exogenous 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 to one, the utility of an individual with ability  $n$  and shareholdings  $\sigma$  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. I assume that firms, unlike the government, perfectly observe the ability of their workers.<sup>4</sup> 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 their labor income, which firms also observe. Second, high-ability individuals typically have an incentive to truthfully reveal their ability

<sup>3</sup>This would also be the case with Greenwood-Hercowitz-Huffman (GHH) preferences, so that the utility function is of the form  $u(c, l) = V(c - \phi(l))$ , where  $V(\cdot)$  is increasing.

<sup>4</sup>See [Stantcheva \(2014\)](#), [Bastani et al. \(2015\)](#), and [Craig \(2023\)](#) for an analysis of optimal income taxation if firms do not (perfectly) observe ability. These studies assume that firms have no monopsony power and competition drives (expected) profits to zero.

to firms. On the contrary, if the government attempts to tax ability, high-ability individuals would try to mimic low-ability individuals.

To a potential employee with ability  $n$ , a firm offers a bundle  $(z, l)$  which specifies labor earnings  $z \geq 0$  and effort  $l \geq 0$ .<sup>5</sup> 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 but not by the government, as it depends on the tax schedule  $T(\cdot)$ . Moreover, as will be made clear below, the outside option is related to how much monopsony power firms have. If a firm is matched to a worker with ability  $n$ , it solves

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

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 I denote the solution to the maximization problem (3) by  $l(n)$  and  $z(n)$ .<sup>6</sup> At an interior solution with strictly positive labor effort and earnings, the following condition must hold:

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

At the optimum, firms offer bundles which equate an individual's productivity (on the left-hand side) to their 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, 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. To focus on distributional issues, the model thus abstracts from efficiency losses due to monopsony power. Section 4.1 analyzes an extension where this is not the case.

### 2.3 Monopsony power

Monopsony power determines what share of the labor market surplus is translated into pure economic profits or, relatedly, 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

<sup>5</sup>This set-up differs from [Hummel \(2023\)](#), where a monopsonist sets an hourly wage  $w = z/l$  and individuals choose how many hours  $l$  to work. This leads to a markdown of productivity over wages and a downward distortion in hours worked, which is not the case in the current paper.

<sup>6</sup>For details on the second-order conditions, see Appendix IV.

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 their preferred number of hours at an hourly wage equal to their productivity.<sup>7</sup> 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 they reject the contract offered to them, which does not vary with ability. The Lagrangian associated with the firm's problem (3) is then

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

where the  $\kappa$ 's are Lagrange multipliers. I assume the benefit  $-T(0)$  is such that firms do not make profits from hiring the least productive workers:  $\pi(n_0) = 0$ .<sup>8</sup> To derive an expression for the profits  $\pi(n)$  firms generate from hiring a worker with ability  $n$  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  $n$ .<sup>9</sup> 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.** *The degree of **monopsony power**  $\mu \in [0, 1]$  and the profits  $\pi(n) = nl(n) - z(n)$  firms generate from hiring a worker with ability  $n$  are related through*

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

Equation (7) implicitly defines in a reduced-form way the degree of monopsony power, which is assumed to be invariant to tax policy.<sup>10</sup> Clearly, if the degree of monopsony power is  $\mu = 0$ , labor markets are perfectly competitive and firms make zero profits. Conversely, if firms have full monopsony power, i.e., if  $\mu = 1$ , 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. It is demonstrated formally in Appendix I that  $\mu \in [0, 1]$  corresponds to the generalized Nash bargaining weight if taxes are linear. Such a foundation is lacking, however, if the tax function

<sup>7</sup>Formally, the outside option is then given by  $\underline{v}(n) = \max_{\ell} \{nl - T(n\ell) - \phi(\ell)\}$ .

<sup>8</sup>As 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.

<sup>9</sup>This relationship uses the assumption that, if firms have *full* monopsony power, the outside option is  $-T(0)$ , which does not vary with ability. If firms do not have full monopsony power, the outside option  $\underline{v}(n)$  does vary with ability.

<sup>10</sup>Equation (7) relates the degree of monopsony power to firms' profits. Equation (3), in turn, relates profits to the outside option of workers. Monopsony power can thus equivalently be interpreted as determining the outside option of workers (which, in equilibrium, equals their payoff from working).

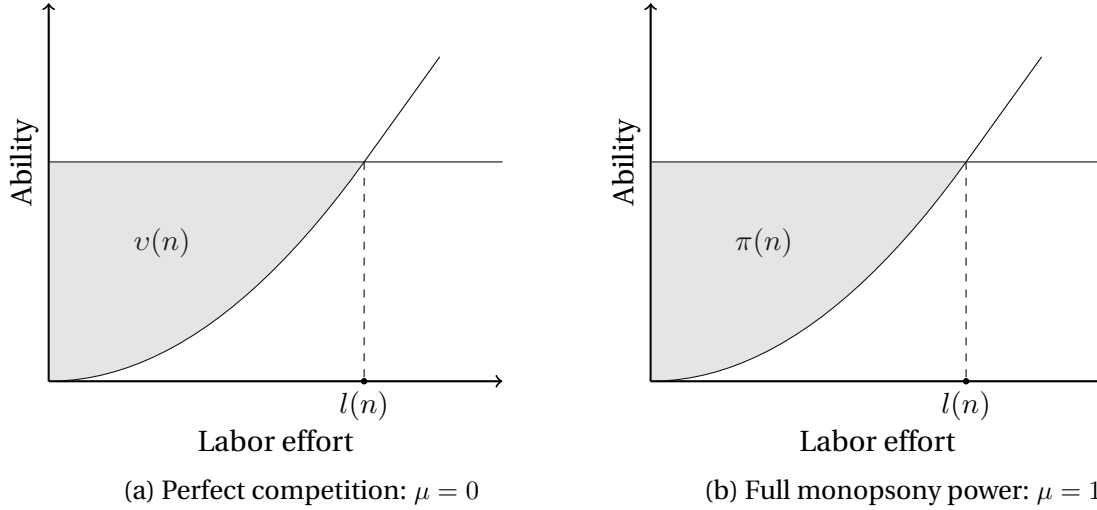


Figure 1: Labor market equilibrium

is non-linear. In this case, the above definition of monopsony power does not have a clear economic interpretation or game-theoretic foundation. 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 effect of raising monopsony power.

Figure 1 graphically illustrates how monopsony power affects the payoffs of workers and firms. Here, I assume income taxes are absent.<sup>11</sup> 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. By construction, monopsony power does not reduce the size of the pie, i.e., does not generate efficiency losses, but only how it 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.<sup>12</sup>

## 2.4 Government

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

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

<sup>11</sup>The reason for doing so is that, through its effect on labor income and thereby the total tax burden, monopsony power has a mechanical effect on the part of the labor market surplus that can be shared between workers and firms. To ease the exposition, Figure 1 abstracts from this mechanical effect that goes via income taxes.

<sup>12</sup>The equilibrium with full monopsony power also occurs if firms engage in first-degree price discrimination. In that case, firms pay workers their reservation wage *for every hour worked*. Hence, the hourly wage depends on the number of hours worked. Firms then continue to demand labor effort up to the point where the worker's productivity is high enough to compensate for their marginal disutility of working.

Here,  $\gamma(n, \sigma) \geq 0$  is the welfare weight, or Pareto weight, the government attaches to an individual with ability  $n$  and shareholdings  $\sigma$ . The average welfare weight is normalized to one. 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  $\sigma$ . 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  $\mathbb{E}[\gamma(n, \sigma)|n]$  is weakly decreasing in  $n$ . 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  $g(n) = \mathbb{E}[\gamma(n, \sigma)|n]$  is the average welfare weight of individuals with the same ability  $n$  and  $\Sigma = -\text{Cov}[\sigma, \gamma] \in [0, 1]$  is the negative covariance between shareholdings and welfare weights, which is bounded between zero and one.

*Proof.* See Appendix II. □

Individuals derive utility from 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  $\Sigma$ , therefore lowers the contribution of after-tax 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, I assume the government observes the total profits made by firms, but cannot infer the abilities of the workers that are employed by each firm. These profits are taxed linearly at an exogenous rate  $\tau \in [0, 1]$ , either at the firm or the individual level. 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 exogenous spending. Because the profit tax is non-distortionary and the government wishes to redistribute from individuals with high to individuals with low shareholdings, it is optimal to levy a confiscatory tax on profits. One can therefore interpret the exogenous rate  $\tau$  as the *maximum* share of 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 an extension where profit taxes are distortionary by inducing 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 a given degree of monopsony power  $\mu$  and given labor income taxes  $T(\cdot)$ , a profit tax rate  $\tau$ , 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 given tax instruments and a given degree of monopsony power. 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.<sup>13</sup> 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 comparative static results regarding the effect of monopsony power or tax reforms on equilibrium 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  $\mu = 0$  (perfect competition) and  $\mu = 1$  (full monopsony power). If labor markets are perfectly competitive, firms earn zero profits – irrespective of the level of taxation. The full incidence of labor income taxes then falls on workers. Conversely, if firms have full monopsony power, all workers are put on their participation constraint:  $v(n) = -T(0)$  for all  $n$ . An increase in the tax burden  $T(z(n))$  at  $z(n) > 0$  must then be compensated one-for-one by higher labor earnings as otherwise workers prefer non-employment. In this case, the full incidence of labor income taxes falls on firms. Second, monopsony power decreases inequality in labor market payoffs generated by differences in ability, but increases inequality in capital income generated by differences in shareholdings. This is because monopsony power increases the share of the labor market surplus that accrues to firms. An increase in monopsony power thus raises aggregate profits and lowers the aggregate wage bill.

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, are associated with higher inequality in labor earnings. Regarding the effect of monopsony power on tax incidence, [Fuest et al. \(2018\)](#) find

<sup>13</sup>The integral equation is  $\pi(n) = \mu \int_{n_0}^n l(m) dm$ , where  $l(m)$  solves the first-order condition for profit maximization  $m(1 - T'(ml(m) - \pi(m))) = \phi'(l(m))$ .

that workers bear a smaller share of the *corporate* tax burden if firms have more monopsony power. Saez et al. (2019) and Lobel (2024) find that a payroll tax cuts in Sweden and Brazil significantly raised profits, which suggests that firms have substantial monopsony power. Furthermore, Lobel (2024) presents evidence that wage and employment pass-through effects from payroll tax cuts are larger in small firms, which suggests that 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. Section 3.1 derives results for optimal income taxation, and Section 3.2 analyzes how monopsony power affects welfare.

#### 3.1 Optimal income taxation

The government chooses the non-linear tax function  $T(\cdot)$  that maximizes welfare. To solve this problem, I follow the approach pioneered by Mirrlees (1971) and 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.

**Proposition 1.** *At an interior solution with positive labor effort and earnings, 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. 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.

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  $z(n)$ .<sup>14</sup> 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.<sup>15</sup> 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 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).<sup>16</sup>

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. The optimal marginal tax rate trades off distributional benefits from taxing labor income against the costs of distorting labor effort. 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 at redistributing labor (capital) income. This result bears close resemblance to Sleet and Yazici (2024), who study optimal income taxation in a framework with frictions and on and off the job search. In their model, the incidence of labor income taxes too falls partly on firms' profits (the "profit squeeze" channel), which are taxed at a confiscatory rate ( $\tau = 1$ ). This makes labor income taxes less effective at redistributing labor income, which provides a force for lower tax rates. By con-

<sup>14</sup>Note that individuals with different abilities do not earn the same labor income if firms have full monopsony power. This is because firms demand more labor effort from individuals with higher ability. To compensate them, i.e., to ensure the participation constraint holds, firms must pay higher labor earnings to these individuals.

<sup>15</sup>In recent work, Kushnir and Zubrickas (2023) show that a similar term is added to an otherwise standard optimal tax formula if firms make pure profits due to decreasing returns to scale and unequal firm-ownership. A 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*.

<sup>16</sup>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.

trast, in my model, profit taxation is restricted ( $\tau \leq 1$ ) and, as a result, the effect of monopsony power on optimal marginal tax rates is ambiguous. The effect depends crucially on the government's preferences for redistribution, as shown 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 with positive labor effort and earnings, 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 income levels, an increase in monopsony power  $\mu$  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 faced by individuals with ability  $n$ . Because monopsony power makes income taxes more (less) effective at redistributing capital (labor) income, the effect 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  $\tau$  is low, and if the government has a strong preference for redistributing capital income, i.e., if  $\Sigma$  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.<sup>17</sup>

The effect 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. *Ceteris paribus*, monopsony power therefore has a smaller positive or a larger negative effect on optimal tax rates at

<sup>17</sup>Monopsony power also lowers the optimal marginal tax rate if the local Pareto parameter  $a(n)$  is high. The reason is quite mechanical. In the second component of equation (11), monopsony power affects optimal marginal tax rates through the term  $T'(z(n))/(1-T'(z(n)))$ . The latter changes faster and, hence, implies a smaller change in the marginal tax rate, the higher is  $T'(z(n))$ . This is the case if the local Pareto parameter is low. Therefore, a lower Pareto parameter makes it easier for condition (15) to be satisfied.

higher income levels. In that sense, monopsony power makes the optimal tax schedule less progressive. Intuitively, by making income taxes less effective at redistributing labor income, monopsony power dampens the force for increasing marginal tax rates.<sup>18</sup>

### 3.2 Welfare effect 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 the tax function  $T(\cdot)$  is optimized. An increase in monopsony power  $\mu$  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 after-tax capital income and welfare weights.

*Proof.* See Appendix VIII. □

Monopsony power raises aggregate profits and lowers the aggregate wage bill. The associated effect 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 can be good for 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, which 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. The combination of monopsony power and firms' informational advantage 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.<sup>19</sup> *Ceteris paribus*, the associated effect on welfare is positive.

<sup>18</sup>Another reason why monopsony power typically makes the optimal tax schedule less progressive is that, 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.

<sup>19</sup>Monopsony power does not have a positive welfare effect if firms, like the government, do not observe ability either. [da Costa and Maestri \(2019\)](#) show that in this case, the best a planner can achieve is to implement the allocation that maximizes welfare subject to the same resource and incentive constraints it would face if labor markets are competitive.

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 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 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, since all workers are put on their identical participation constraint. A confiscatory tax on profits, in turn, guarantees there is no inequality in capital income either. Hence, all individuals are equally well off. The government can implement the first-best allocation by levying a confiscatory tax on profits to provide a universal basic income  $-T(0)$  that should not be taxed away if individuals earn higher incomes. However, in the typical case where taxing profits at a confiscatory rate would be either unfeasible (e.g., due to profit shifting opportunities, introduced in Section 4.2) or undesirable (e.g., due to the inability to distinguish normal and above-normal returns), monopsony power has an ambiguous effect on welfare: it 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. 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. 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 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  $\mu$  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 after-tax labor income and welfare weights.

*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) does not depend on the disutility of working, but instead on after-tax labor income.

## 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, generate distortions by inducing firms to engage in costly profit shifting (Section 4.2).

### 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 correctly take into account how a change in earnings and effort affect their workers' utility. As a result, monopsony power does not distort labor supply. 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.<sup>20</sup>

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 differ not only in their ability but also 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,

<sup>20</sup>For an analysis of optimal income taxation when firms have monopsony power and distort labor supply on the intensive margin, see Hummel (2023).

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 their labor income if employed. Consequently, in addition to a linear tax  $\tau$  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.<sup>21</sup>

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 threshold pins down the level of profits, which is again related to monopsony power cf. Definition 1. Equilibrium labor effort, earnings, and profits can again be found by solving equations (4) and (7) together with the relationship  $\pi(n) = nl(n) - z(n)$ . The utility of an individual with ability  $n$ , shareholdings  $\sigma$  and participation costs  $\varphi$  is then

$$\mathcal{U}(n, \sigma, \varphi) = \max\{v(n) - \varphi, b\} + \sigma(1 - \tau)\bar{\pi}. \quad (18)$$

Participation costs  $\varphi$  are subtracted from the labor market payoff  $v(n)$  as they lower the utility from working. Equation (18) determines a participation threshold  $\varphi(n) = v(n) - b$  at every ability level. Hence, an individual with ability  $n$  and participation costs  $\varphi$  becomes employed if and only if  $\varphi \leq \varphi(n)$ . I denote the participation rate of individuals with ability  $n$  by

$$p(\varphi(n)) = \frac{\int_{\sigma_0}^{\sigma_1} \int_{\varphi_0}^{\varphi(n)} h(n, \sigma, \varphi) d\varphi d\sigma}{\int_{\sigma_0}^{\sigma_1} \int_{\varphi_0}^{\varphi_1} h(n, \sigma, \varphi) d\varphi d\sigma}, \quad (19)$$

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

As before, monopsony power does not distort labor supply on the intensive (effort) margin. However, by lowering the labor market payoff  $v(n)$  that accrues to workers, it does lead to a distortion in labor supply on the extensive (participation) margin. To see this, suppose there are no taxes and benefits:  $T(z(n)) = b = 0$ . In this case, an allocation is Pareto efficient if a worker becomes employed whenever the *joint* surplus that can be shared between workers and firms 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:  $v(n) \geq \varphi$ . When making their participation decision, individuals do not internalize the profits made by firms. Consequently, labor participation is distorted downwards whenever firms make positive profits, i.e., whenever firms have monopsony power.

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 effect of monopsony power.

**Proposition 3.** *Suppose individuals supply labor on the extensive margin according to equa-*

<sup>21</sup>It is useful to distinguish between a benefit  $b$  paid to non-participants and the transfer  $-T(0)$  an individual receives if they reject 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:  $\pi(n_0) = 0$ .

tion (18). At an interior solution with positive labor effort and earnings, the optimal marginal tax rate satisfies

$$T'(z(n)) = \frac{1 - F_p(n)}{n f_p(n)} \left[ \mu(1 - \tau)\Sigma + (1 - \mu)(1 - T'(z(n)))(1 + 1/\varepsilon(n)) \right. \\ \left. \times \mathbb{E}[1 - g_p(m) - \hat{p}(m)\pi(m)(1 - (1 - \tau)\Sigma) - \hat{p}(m)(T(z(m)) + b)|m \geq n] \right], \quad (20)$$

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

$$\mu \int_{n_0}^{n_1} v(n)(1 - g_p(n))f_p(n)dn > (1 - \mu)\Sigma^k \\ + \mu \int_{n_0}^{n_1} v(n)\hat{p}(n)\pi(n)(1 - (1 - \tau)\Sigma)f_p(n)dn + \mu \int_{n_0}^{n_1} v(n)\hat{p}(n)(T(z(n)) + b)f_p(n)dn. \quad (21)$$

*Proof.* See Appendix IX. □

Compared to Proposition 1, there are two substantive changes in the optimal tax formula (20). The first modification is standard and occurs because changes in the participation rate generate a fiscal externality proportional to  $T(z(m)) + b$ , as captured by the final term on the second line.<sup>22</sup> Second, and more importantly, 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 (20) as well. 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. 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.

Equation (21) 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 gen-

<sup>22</sup>To understand this term, note that a higher marginal tax rate  $T'(z(n))$  raises the tax burden for individuals with earnings at least equal to  $z(n)$ . 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  $T(z(m)) + b$ . This modification of the optimal tax formula is also present if firms do not have monopsony power ( $\mu = 0$ ). 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 (20) gives an optimal tax formula that is very similar to the ones derived in these papers.

erates efficiency losses if  $\hat{p}(n)\pi(n) > 0$ , as captured by the first term on the second line. Provided the government values profits to some extent, i.e., provided  $(1 - \tau)\Sigma < 1$ , the fact that monopsony power distorts labor participation makes 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 (21), the welfare effect 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 effect on government finances. Again, this makes it less likely that an increase in monopsony power raises welfare.

## 4.2 Distortionary profit taxes

The results up to this point have been derived assuming the profit tax is exogenous and does not affect any economic decisions. Given there are distributional benefits associated with taxing profits if  $\Sigma > 0$ , 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:

$$\Pi(\bar{\pi}, \tau) = \max_{s \in [0, 1]} \{s\bar{\pi}(1 - \tilde{\rho}(s)) + (1 - s)\bar{\pi}(1 - \tau)\} = \max_{s \in [0, 1]} \{(1 - (1 - s)\tau - \rho(s))\bar{\pi}\}. \quad (22)$$

Here,  $\rho(s) = s\tilde{\rho}(s)$  denotes the total shifting costs per unit of pretax profits  $\bar{\pi}$ . At an interior solution with  $s \in (0, 1)$ , the first-order condition is

$$\rho'(s) = \tau. \quad (23)$$

Firms continue to shift profits until the marginal costs of doing so (on the left-hand side) equal 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 positive whenever  $\tau > 0$  and increasing in the profit tax. 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 effect of raising monopsony power in [Appendix X](#). The main findings are summarized below.

**Proposition 4.** *Suppose firms shift a fraction  $s(\tau) \in [0, 1]$  of pretax profits to a tax haven according to equation (23). At an interior solution with  $\tau \in (0, 1)$ , the optimal profit tax  $\tau$  satisfies*

$$\frac{\tau}{1 - \tau} = \frac{\Sigma}{\epsilon_{1-s, 1-\tau}}, \quad (24)$$

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

$$T'(z(n)) = \frac{1 - F(n)}{nf(n)} \left[ \mu \left( [1 - (1 - s(\tau))\tau - \rho(s(\tau))] \Sigma + \rho(s(\tau)) \right) + (1 - \mu)(1 - T'(z(n)))(1 + 1/\varepsilon(n))(1 - \bar{g}(n)) \right]. \quad (25)$$

Furthermore, an increase in monopsony power  $\mu$  raises welfare if and only if

$$\mu \Sigma^v > (1 - \mu) \Sigma^k + (1 - \mu) R, \quad (26)$$

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 (24) 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. 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 (25) 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:  $(1 - (1 - s)\tau - \rho(s))\Sigma \geq (1 - \tau)\Sigma$ .<sup>23</sup> 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 4 states that the welfare effect of raising monopsony power is lower if there are profit shifting opportunities: compare equations (26) and (16). Intuitively, an increase in monopsony power raises aggregate profits and thereby the aggregate

<sup>23</sup>To see this, note that  $s \in [0, 1]$  is chosen to maximize  $\Lambda(s) = 1 - (1 - s)\tau - \rho(s)$ , the term multiplied by  $\Sigma$ , and that  $\Lambda(0) = 1 - \tau$  because  $\rho(0) = 0$ . Together, these observations imply  $1 - (1 - s)\tau - \rho(s) \geq 1 - \tau$ .

gate costs of profit shifting. Doing so is costly from a social perspective, because it reduces the resources available for consumption by an amount equal to  $R$ . Hence, an increase in monopsony power is less likely to raise welfare 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 there are no efficiency losses from monopsony power and profit taxes do not affect any economic decisions. After discussing the calibration (Section 5.1) and welfare function (Section 5.2), I analyze how monopsony power affects optimal income taxation (Section 5.3) and welfare (Section 5.4).

### 5.1 Calibration

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 different sources of 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.

#### 5.1.1 Functional forms

To calibrate the model, I first characterize the equilibrium for a given tax schedule  $T(z)$  and 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

$$T(z(n)) = -q + tz(n), \quad (27)$$

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. 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 (cf. Corollary 1),

$$\phi(l) = \frac{l^{1+1/\varepsilon}}{1 + 1/\varepsilon}, \quad (28)$$

where  $\varepsilon$  governs the convexity. It measures the percentage increase in labor effort and, given the choice of the tax schedule (27), the percentage increase in labor earnings as well, following a one percent increase in the net-of-tax rate: see equation (4). I select a value of  $\varepsilon = 0.33$

based on [Chetty \(2012\)](#). This value for the intensive-margin elasticity is 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. \(2012\)](#).

### 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 \(4\)](#):

$$l(n) = (1 - t)^\varepsilon n^\varepsilon. \quad (29)$$

To derive an expression for labor income, substitute out for labor effort in [equation \(7\)](#) and use the relationship  $z(n) = nl(n) - \pi(n)$  to find

$$\begin{aligned} z(n) &= \left(1 - \frac{\mu}{1 + \varepsilon}\right) (1 - t)^\varepsilon n^{1+\varepsilon} + \left(\frac{\mu}{1 + \varepsilon}\right) (1 - t)^\varepsilon n_0^{1+\varepsilon} \\ &= \left(1 - \frac{\mu}{1 + \varepsilon}\right) (1 - t)^\varepsilon n^{1+\varepsilon} + \left(\frac{\mu}{1 + \varepsilon}\right) z(n_0). \end{aligned} \quad (30)$$

According to [equation \(30\)](#), an individual's labor income equals a weighted average of the output they produce (first term) and the labor income of the individuals with the lowest ability (second term).<sup>24</sup> The profits  $\pi(n) = nl(n) - z(n)$  firms generate from hiring a worker with ability  $n$  are then given by

$$\pi(n) = \left(\frac{\mu}{1 + \varepsilon}\right) (1 - t)^\varepsilon [n^{1+\varepsilon} - n_0^{1+\varepsilon}] = \left(\frac{\mu}{1 + \varepsilon - \mu}\right) (z(n) - z(n_0)). \quad (31)$$

From [equations \(30\)–\(31\)](#) it can readily be seen that a higher degree of monopsony power  $\mu$  amplifies (mitigates) the negative effect of the labor tax rate on profits (labor income) that operates through its effect on effort, cf. [equation \(29\)](#). These equations, in turn, give a mapping from (observable) labor income to (unobservable) ability and the pure economic profits from hiring a worker with ability  $n$ , for a given tax rate  $t$ , elasticity of labor supply  $\varepsilon$ , and degree of monopsony power  $\mu$ , which is discussed next.

### 5.1.3 Degree of monopsony power

Monopsony power  $\mu$  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. Because my model abstracts from productive capital, I calibrate monopsony power  $\mu$  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

<sup>24</sup>The reason why the lowest income shows up in [equation \(30\)](#) 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.

recent year 2015, [Barkai and Benzell \(2018\)](#) calculate that the ratio of aggregate pure profits to aggregate wages is approximately  $\bar{\pi}/\bar{z} = 24.2\%$ . Using their estimate, the value for monopsony power  $\mu$  can be calculated by integrating equation (31) over the ability distribution and dividing by average labor income  $\bar{z} = \int_{n_0}^{n_1} z(n)f(n)dn$ . This gives

$$\left(\frac{\bar{\pi}}{\bar{z}}\right) = \left(\frac{\mu}{1 + \varepsilon - \mu}\right) \left(1 - \left(\frac{z(n_0)}{\bar{z}}\right)\right) \Leftrightarrow \mu = (1 + \varepsilon) \left[\frac{(\bar{\pi}/\bar{z})}{1 + (\bar{\pi}/\bar{z}) - (z(n_0)/\bar{z})}\right]. \quad (32)$$

For a given elasticity of labor supply  $\varepsilon$  and ratio of lowest to average labor income  $z(n_0)/\bar{z}$ , equation (32) gives a relationship between the degree of monopsony power  $\mu$  and the ratio of the profit share to the labor share  $\bar{\pi}/\bar{z}$ . The implied degree of monopsony power is approximately  $\mu = 0.28$ . It should be emphasized, however, 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. In reality, of course, these profits come from other sources as well.

#### 5.1.4 Joint distribution of ability and shareholdings

I calibrate the joint distribution of ability and shareholdings  $H(n, \sigma)$  to match the joint distribution of labor and capital income obtained from the CPS. An individual's labor income is defined as their income from wage and salary payments. An individual's capital income, in turn, is defined as their income from interest, dividends, and rents. For individuals whose labor (capital) income is top-coded, I multiply the reported income with a factor of 2.67 (3.63), consistent with an estimate of the Pareto parameter of 1.6 (1.38) for the distribution of labor (capital) income at the top, obtained by [Saez and Stantcheva \(2018\)](#). In the final sample, I include individuals between 25 years and 65 years whose hourly wage exceeds half the federal minimum wage of \$7.25 and whose labor income exceeds \$5,710.<sup>25</sup> For each individual, I use equation (30) to calculate their ability  $n$ . As in [Saez \(2001\)](#), this approach ensures the ability distribution is consistent with the empirical distribution of labor income. To calculate the shareholdings  $\sigma$  of each individual, I divide their capital income by average capital income. This approach is valid under the assumption that above-normal profits flow back to individuals in proportion to their capital income from the CPS (which includes the normal returns to capital, which are absent in my model, but excludes some other forms of capital income, such as pensions and capital gains). The implied correlation between ability and shareholdings is small but significantly positive at approximately  $\rho = 0.07$  ( $p < 0.01$ ).

Based on the calculated ability  $n$  for each individual, I obtain the marginal distribution  $F(n)$  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 labor income. I choose the tail parameter of the *ability* distribution to be consistent with a tail parameter of 1.6 of the *labor income*

<sup>25</sup>This is the labor 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.

distribution at the top, cf. [Saez and Stantcheva \(2018\)](#).<sup>26</sup> 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. Figure 6 plots the local Pareto parameter.

### 5.1.5 Profit taxation and revenue requirement

In the model, there is no productive capital and  $\tau$  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 assume in the baseline that all capital income is taxed at a rate  $\tau = 36\%$ , taken from [Trabandt and Uhlig \(2011\)](#), but also study how sensitive the results are to this choice. For a given value of  $\tau$ , 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
$\mu$	Aggregate profits over wages	<a href="#">Barkai and Benzell (2018)</a>	0.28
$\varepsilon$	Elasticity of labor supply	<a href="#">Chetty (2012)</a>	0.33
$\tau$	Tax rate on capital income	<a href="#">Trabandt and Uhlig (2011)</a>	0.36
$G$	Government budget constraint	Equilibrium condition	\$22,782
$T(z)$	Tax liability	CPS 2018	Figure 4
$F(n)$	Labor income distribution	CPS 2018	Figure 5

Table 1: Calibration

## 5.2 Welfare function

From Lemma 1, the government's preferences for redistribution are summarized by the average welfare weight  $g(n) \geq 0$  it attaches to individuals with the same ability  $n$  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). To obtain values for these, I assume the Pareto weight  $\gamma(n, \sigma)$  the government attaches to an individual with ability  $n$  and shareholdings  $\sigma$  takes the following form:

$$\gamma(n, \sigma) = \zeta(n^{1+\varepsilon} + \psi\sigma)^{-\beta}. \quad (33)$$

Here,  $\zeta$  is a parameter which ensures the average welfare weight equals one, and  $\psi$  scales shareholdings in such a way that, in the calibrated economy, welfare weights are declining

<sup>26</sup>Let  $\tilde{F}(z(n))$  denote the labor income distribution with density  $\tilde{f}(z(n))$ . Monotonicity of labor earnings implies  $F(n) = \tilde{F}(z(n))$  for all  $n$  where  $z(n) > 0$  and, hence,  $f(n) = \tilde{f}(z(n))z'(n)$ . The local Pareto parameter of the ability distribution  $a(n) = nf(n)/(1 - F(n))$  and income distribution  $\tilde{a}(z(n)) = z(n)\tilde{f}(z(n))/(1 - \tilde{F}(z(n)))$  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 (30).

in an individual's total (labor plus capital) income.<sup>27</sup> The parameter  $\beta$ , in turn, determines how quickly welfare weights decline in an individual's ability and shareholdings. I select the value of  $\beta$  such that the average marginal tax rate at the optimal tax system with perfectly competitive labor markets, which is obtained by setting  $\mu = 0$  in equation (14) and imposing  $\psi = 0$ , equals the current rate  $t = 33.1\%$ .

Based on the Pareto weight  $\gamma(n, \sigma)$  for each individual in the sample, I calculate  $\bar{g}(n)$  as the average welfare weight for all individuals with ability at least equal to  $n$ . I subsequently smooth these values and calculate the average welfare weight  $g(n)$  at each point in the ability distribution. The results are shown in Figure 7. To obtain a value for  $\Sigma$ , I calculate the negative covariance between welfare weights and shareholdings, cf. Lemma 1. This gives approximately  $\Sigma = 0.50$ . Incidentally, this value lies midway between  $\Sigma = 0$ , where the government does not value redistributing capital income at all, and  $\Sigma = 1$ , where there is a very strong desire to redistribute capital income because all profits flow back to individuals whose welfare weight equals zero. Because the results turn out to be sensitive to the government's desire to redistribute capital income, in what follows I show results for the entire range of  $\Sigma \in [0, 1]$ .

### 5.3 Optimal tax policy

Based on the calibrated set of primitives and welfare function, I calculate optimal marginal tax rates using equation (14) from Corollary 1. The red, solid line in Figure 2 shows the results. To facilitate the interpretation, the horizontal axis plots labor income at the *current* tax system. Marginal tax rates start out very high at values close to 100%.<sup>28</sup> Intuitively, at low levels of income, marginal tax rates are particularly effective at indirectly taxing profits, cf. Corollary 1. After that, marginal tax rates decrease rapidly and reach a minimum of approximately 26% for individuals whose current income is about \$100,000 and start to increase gradually to about 45% before staying relatively flat for individuals whose current labor income is above \$350,000, the point where the Pareto tail is pasted.

Optimal marginal tax rates with monopsony power look quite different compared to the case where labor markets are competitive, especially at low levels of income. To illustrate this point, the black, dashed line shows the optimal marginal tax rates if the government, based on the *same* set of primitives, wrongfully sets tax policy as if labor markets are competitive. The results are obtained by setting  $\mu = 0$  in equation (14). If labor markets are perfectly competitive, marginal tax rates start at zero, cf. Seade (1977). Intuitively, marginal tax rates at the bottom are not helpful for redistributing labor income and, if labor markets are competitive, not helpful for redistributing capital income as well. 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).

<sup>27</sup>With a linear tax system, an individual's labor income is approximately proportional to  $n^{1+\varepsilon}$ : see equation (30). Then, to make the term  $n^{1+\varepsilon} + \psi\sigma$  in the formula for the welfare weight proportional to an individual's total (labor plus capital) income, as given by  $z(n) + \sigma\bar{\pi}$ , I set  $\psi = \bar{\pi}/((1-t)^\varepsilon(1-\mu/(1+\varepsilon)))$ .

<sup>28</sup>In fact, the low value of the local Pareto parameter  $a(n)$  at the bottom of the ability distribution implies that marginal tax rates are so high that individuals whose current labor earnings are below \$9,186 do not work at the optimal allocation. For these workers, the constraint  $l(n) \geq 0$  in the optimal tax problem is binding.

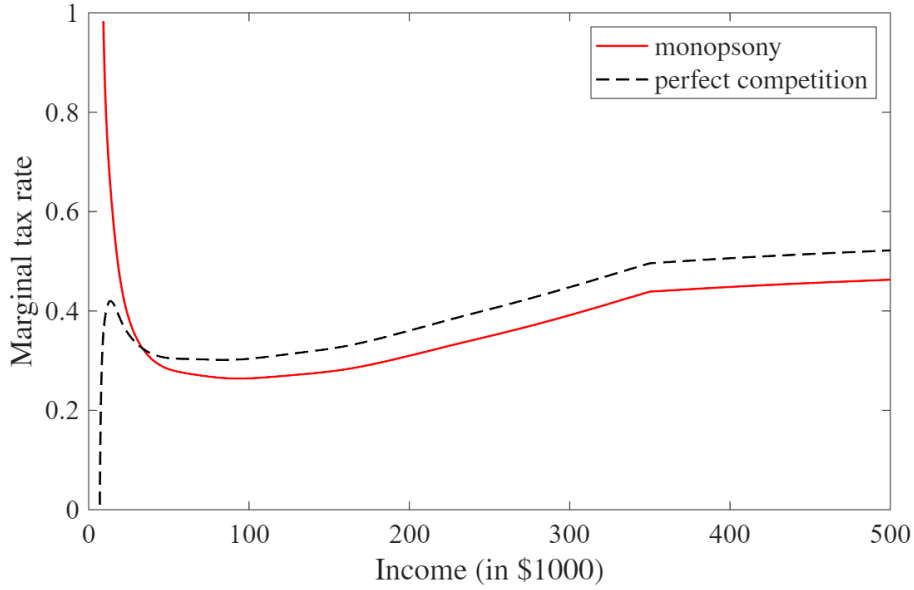


Figure 2: Optimal marginal tax rates

The main insight from Figure 2 is that monopsony power raises (lowers) optimal marginal tax rates at low (middle and high) levels of income. In that sense, monopsony power makes the optimal tax schedule less progressive. For individuals whose current labor income is below \$34,000, marginal tax rates with monopsony power are higher than would be the case if labor markets are perfectly competitive. As explained before, if firms have monopsony power, marginal tax rates on labor income 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: see Corollary 1 and Figure 6. By contrast, for individuals whose current labor income exceeds \$34,000, marginal tax rates with monopsony power are lower than would be the case if labor markets are perfectly competitive. Intuitively, at higher income levels marginal tax rates are primarily aimed at redistributing labor income, and monopsony power makes them less effective at doing so, because part of the tax incidence falls on firms.

Figure 2 plots optimal marginal tax rates with and without monopsony power using the same set of primitives, i.e., without recalibration. As such, the black, dashed line shows the marginal tax rates a government would set if it assumes labor markets are competitive, *despite* having correctly inferred the ability distribution from the underlying income distribution (a mapping that depends on the degree of monopsony power). An arguably more realistic benchmark for optimal marginal tax rates under perfect competition is obtained by first recalibrating the model under the assumption that the degree of monopsony power is  $\mu = 0$ . Figure 8 in Appendix XI shows the results. As in Figure 2, the red, solid and black, dashed lines show optimal marginal tax rates with ( $\mu = 0.28$ ) and without ( $\mu = 0$ ) monopsony power, using the same set of primitives. The blue, dotted line, in turn, shows optimal marginal tax rates under perfect competition after the model is first recalibrated assuming  $\mu = 0$ . As it turns out, recalibration does not have a big effect on optimal marginal tax rates under perfect competition: the black, dashed and blue, dotted lines are very similar.

How monopsony power affects optimal marginal tax rates, especially at the bottom of the income distribution, is sensitive to the preferences for redistributing capital income. To show this, Figure 9 in Appendix XI plots optimal marginal tax rates for different values of  $\Sigma$ , the negative covariance between shareholdings and welfare weights. As in Figure 2, the red, solid line shows optimal marginal tax rates under monopsony in the baseline (with  $\Sigma = 0.50$ ) and the black, dashed line shows optimal marginal tax rates under perfect competition ( $\mu = 0$ ). Without a preference for redistributing capital income ( $\Sigma = 0$ ), optimal marginal tax rates start at zero and always lie below optimal marginal tax rates under perfect competition, see the blue, dotted line. This is because monopsony power makes labor income taxes less effective at redistributing labor income, as part of the incidence falls on firms, and there are no benefits to indirectly taxing profits. Conversely, if  $\Sigma = 1$ , there are large gains from indirectly taxing profits, which significantly raises marginal tax rates especially at the bottom, see the green dash-dotted line. The benchmark with  $\Sigma = 0.50$  lies between these extremes, though, at low income levels, optimal marginal tax rates are much closer to those that are obtained when the government has a very strong preference for redistributing capital income ( $\Sigma = 1$ ). As long as the government has some preference for redistributing capital income, monopsony power significantly raises optimal marginal tax rates at low incomes.

As it turns out, the results of this final exercise are also informative for how the profit tax affects the optimal income tax schedule. The reason is that, whenever firms have a positive degree of monopsony power, optimal marginal tax rates are a function of the composite parameter  $\Sigma(1 - \tau)$ , which determines the negative covariance between welfare weights and *after-tax* capital income, see Corollary 1. Therefore, the blue, dotted line from Figure 9 (with  $\Sigma = 0$ ) would also be obtained under the assumption that profits are taxed at a confiscatory rate,  $\tau = 1$ , since in both cases  $\Sigma(1 - \tau) = 0$ . Conversely, the green dash-dotted line (with  $\Sigma = 1$ ) would also be obtained if capital income is *subsidized* at a rate of approximately 29%, or  $\tau = -0.29$ .<sup>29</sup> Figure 10 in Appendix XI, in turn, shows optimal marginal tax rates for an intermediate case where capital income is subject only to the corporate tax rate of  $\tau = 0.21$  (the black, dashed line) and compares the results to the baseline where profits are taxed at a rate of  $\tau = 0.36$  (the red, solid line). With a lower profit tax, the benefits of indirectly taxing profits are larger. This raises marginal tax rates, though, as can be seen, the effect of moving from  $\tau = 0.36$  to  $\tau = 0.21$  is not large.

#### 5.4 Implications for welfare

To assess the quantitative implications of monopsony power for welfare, 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 red, solid line in Figure 2, with the one that is obtained if the government sets tax policy as if labor markets are competitive, cf. the black, dashed line in Figure 2. Second, I calculate how much the government is willing to pay for getting rid of monopsony power. To that end, I compare the allocation under monopsony power ( $\mu = 0.28$ ) and perfect competition ( $\mu = 0$ ), where in both cases taxes are optimized. The first exercise gives

<sup>29</sup>This figure is calculated as  $1 - (1 - \tau)/\Sigma = 1 - (1 - 0.36)/0.50$ .

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.

Regarding the first exercise, in the baseline economy the welfare costs of ignoring monopsony power when designing tax policy are \$303 in consumption equivalents, or 0.38% of current GDP. Hence, moving from the black, dotted to the red, solid line in Figure 2 generates a welfare gain equivalent to increasing all individuals' consumption by this amount. Regarding the second exercise, getting rid of monopsony power in the baseline economy generates a much larger welfare gain of \$2,707 in consumption equivalents, or 3.41% of current GDP. Starting from a setting with monopsony power ( $\mu = 0.28$ ) where taxes are optimized, this is the welfare gain from moving to a setting with perfectly competitive labor markets ( $\mu = 0$ ) and re-optimizing the tax schedule. It should be emphasized, however, that for a given specification of the welfare function, both figures are clear upper bounds. The reason is that in the calibrated economy *all* economic profits are attributed to monopsony power. If these profits come from other sources as well, the actual degree of monopsony power is smaller, and so are the associated welfare effects.

Not surprisingly, the welfare effects of monopsony power are sensitive to the government's preference for redistributing capital income. To illustrate, Figure 3 plots the results of both exercises for different values of  $\Sigma$  (the vertical, dotted line shows the baseline with  $\Sigma = 0.50$ ). The left axis plots the welfare costs of ignoring monopsony power when designing tax policy, i.e., the costs of "misoptimization". The welfare costs of not taking monopsony power into account range between \$112 and \$982 in consumption equivalents, or between 0.14% and 1.24% of current GDP. The gains are largest if the government has a strong preference for redistributing capital income. The right axis plots the welfare effect of moving from monopsony power ( $\mu = 0.28$ ) to competitive labor markets ( $\mu = 0$ ) and re-optimizing the tax schedule. Recall from Proposition 2 that the welfare effect from changing the degree of monopsony power can be positive or negative. If the government does not value redistributing capital income ( $\Sigma = 0$ ), getting rid of monopsony power generates a welfare *loss* of about \$1,967 in consumption equivalents, or 2.48% of current GDP. By contrast, if the government has a strong desire to redistribute capital income ( $\Sigma = 1$ ), getting rid of monopsony power generates a very significant welfare gain of \$6,981 in consumption equivalents, or 8.80% of current GDP. As mentioned before, however, for a given welfare function these values are clear upper bounds and mostly serve an illustrative purpose, since the calibration attributes all profits to monopsony power.

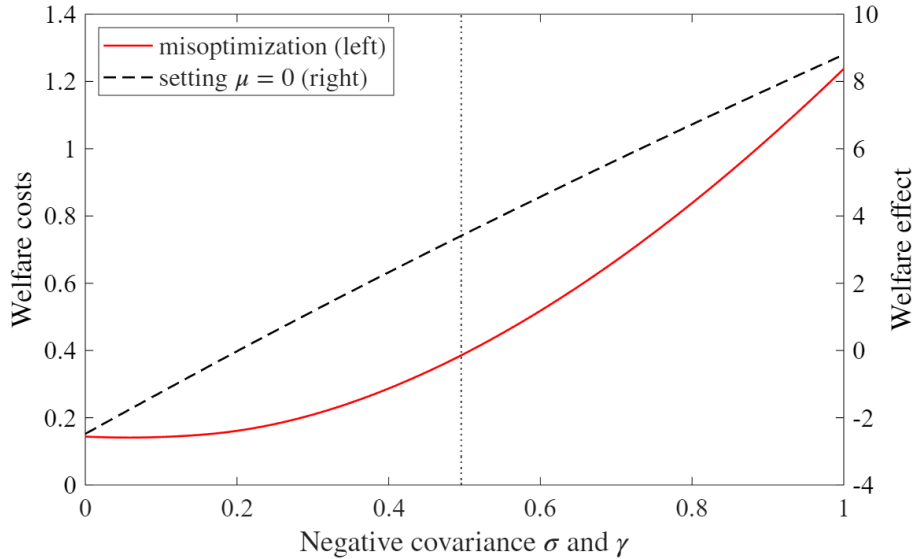


Figure 3: Welfare effect in consumption equivalents (% of GDP)

## 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 version of the model, monopsony power does not generate efficiency losses and profit taxes are non-distortionary. These assumptions are subsequently relaxed. Lastly, I illustrate the findings by calibrating the baseline model without extensions to the US economy.

Monopsony power makes taxes on labor income less effective at redistributing labor income, but more effective at 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. 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) for individuals with low (high) ability 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, the welfare benefits of taking monopsony power into account when designing tax policy are approximately 0.38% of GDP. This figure, however, is likely an upper bound because the calibration attributes *all* economic profits to monopsony power. In reality, profits come from other sources as well.

Monopsony power has an ambiguous effect on welfare. On the one hand, by reducing inequality in labor market payoffs generated by differences in ability, it alleviates the equity-efficiency trade-off that occurs because the government does not observe ability. As such, monopsony power enables the government to exploit the informational advantage of firms. On the other hand, monopsony power exacerbates inequality in capital income generated by differences in shareholdings. In the baseline calibration, the welfare benefits of getting rid of monopsony power are approximately 3.41% of GDP. However, this figure is again likely an

upper bound because monopsony power is not the only source of above-normal profits.

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, as a result, all capital income consists of pure economic profits. Consequently, there is a strong incentive to indirectly tax profits. In reality, capital can raise labor productivity and part of the income it generates are normal returns, e.g., the return for postponing consumption or taking risk. Adding these features likely reduces the optimal tax on capital income cf. the findings from [Atkinson and Stiglitz \(1976\)](#), [Judd \(1985\)](#) and [Chamley \(1986\)](#). However, I conjecture it does not fundamentally alter the insight that (i) taxes on labor income can be used to indirectly tax above-normal profits if firms have monopsony power, and (ii) monopsony power has an ambiguous effect on welfare if firms have an informational advantage compared to the government about their workers' abilities. Second, monopsony power is treated as a single – exogenous and policy-invariant – parameter. However, in reality, firms and workers likely differ in the extent to which they can exert and suffer from monopsony power. It would be interesting to endogenize the degree of monopsony power firms have and account for these differences. I leave this as an extension for future research.

## References

- Ales, L. and C. Sleeter (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 (2023). Optimal fiscal policy in the presence of declining labor share. mimeo, Sabanci 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.
- 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.
- 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.
- 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. (2023). Optimal income taxation with spillovers from employer learning. *American Economic Journal: Economic Policy* 15(2), 82–125.
- 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.
- Eeckhout, J., C. Fu, W. Li, and X. Weng (2021). Optimal taxation and market power. mimeo, UPF Barcelona.
- 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.

- Gürer, E. (2022). Rising markups and optimal redistributive taxation. *International Tax and Public Finance* 29(5), 1227–1259.
- 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.
- Hines, J. and E. Rice (1994). Fiscal paradise: Foreign tax havens and American business. *Quarterly Journal of Economics* 109(1), 149–182.
- Hummel, A. J. (2023). Tax curvature. *Journal of Public Economics* 224, 104921.
- 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.
- 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 (2023). Optimal labor income taxation with the dividend effect. mimeo, Carnegie Mellon University.
- Lipsius, B. (2018). Labor market concentration does not explain the falling labor share. mimeo, University of Michigan.
- Lobel, F. (2024). Who benefits from payroll tax cuts? Market power, tax incidence and efficiency. mimeo, UC Berkeley.
- 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.

- Qiu, Y. and A. Sojourner (2023). Labor-market concentration and labor compensation. *ILR Review* 76(3), 475–503.
- 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 (2024). Tax design with labor market frictions. mimeo, Carnegie Mellon University.
- Stantcheva, S. (2014). Optimal income taxation with adverse selection in the labour market. *Review of Economic Studies* 81(3), 1296–1329.
- Tarasov, A. and R. Zubrickas (2023). Optimal income taxation under monopolistic competition. *Economic theory* 76(2), 495–523.
- Thoresson, A. (2023). Employer concentration and wages for specialized workers. *American Economic Journal: Applied Economics*. Forthcoming.
- 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.

Yeh, C., C. Macaluso, and B. Hershbein (2022). Monopsony in the US labor market. *American Economic Review* 112(7), 2099–2138.

## Appendix

### I Relation $\mu$ and the Nash bargaining weight

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

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

$$\max_{l,z} \Gamma(n) = \mu \log(nl - z) + (1 - \mu) \log(z(1 - t) + q - \phi(l) - (-T(0))), \quad (34)$$

where  $\mu$  is the bargaining power of firms and  $1 - \mu$  the bargaining power of workers. 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

$$l : \frac{\mu}{\pi(n)}n - \frac{1 - \mu}{v(n) + T(0)}\phi'(l(n)) = 0, \quad (35)$$

$$z : -\frac{\mu}{\pi(n)} + \frac{1 - \mu}{v(n) + T(0)}(1 - t) = 0. \quad (36)$$

Combining gives

$$\phi'(l(n)) = n(1 - t), \quad (37)$$

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

$$v'(n) = (1 - t)z'(n) - \phi'(l(n))l'(n) = \frac{\phi'(l(n))}{n} [z'(n) - nl'(n)] = \frac{\phi'(l(n))}{n} [l(n) - \pi'(n)], \quad (38)$$

where I used equation (37) and the definition for profits  $\pi(n) = nl(n) - z(n)$ . Next, rewrite equation (36):

$$\mu(v(n) + T(0)) = (1 - \mu)\pi(n)(1 - t). \quad (39)$$

Differentiate both sides with respect to ability:

$$\mu v'(n) = (1 - \mu)\pi'(n)(1 - t). \quad (40)$$

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

$$\mu v'(n) = (1 - \mu) \frac{\phi'(l(n))}{n} \pi'(n). \quad (41)$$

Equation (35), in turn, implies

$$(v(n) + T(0)) + \frac{\phi'(l(n))}{n} \pi(n) = \frac{1}{\mu} \frac{\phi'(l(n))}{n} \pi(n). \quad (42)$$

Substituting back into equation (41):

$$\mu v'(n) = (1 - \mu) \frac{\phi'(l(n))}{n} \pi'(n). \quad (43)$$

Combining with equation (38) gives:

$$v'(n) = \frac{\phi'(l(n))}{n} (1 - \mu) l(n), \quad (44)$$

$$\pi'(n) = \mu l(n). \quad (45)$$

Integrating equation (45) and imposing the boundary condition  $\pi(n_0) = 0$  gives the following expression for profits:

$$\pi(n) = \mu \int_{n_0}^n l(m) dm, \quad (46)$$

which coincides with equation (7) from the paper. Hence, if the tax function is linear,  $\mu$  corresponds to the firm's bargaining weight from the generalized Nash bargaining solution.<sup>30</sup>

## 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:

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

<sup>30</sup>If the tax function is non-linear,  $\mu$  no longer coincides with the Nash bargaining weight. This can be seen from inspecting equation (36). 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.

$$\begin{aligned}
&= \int_{n_0}^{n_1} g(n)v(n)f(n)dn + (1 - \tau)\bar{\pi} \left( 1 + \underbrace{\int_{\sigma_0}^{\sigma_1} \int_{n_0}^{n_1} (\sigma - 1)(\gamma(n, \sigma) - 1)h(n, \sigma)dnd\sigma}_{= \text{Cov}[\sigma, \gamma] = -\Sigma} \right) \\
&= \int_{n_0}^{n_1} g(n)v(n)f(n)dn + (1 - \tau)(1 - \Sigma) \int_{n_0}^{n_1} \pi(n)f(n)dn, \tag{47}
\end{aligned}$$

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

$$\Sigma = - \int_{\sigma_0}^{\sigma_1} \int_{n_0}^{n_1} (\sigma - 1)(\gamma(n, \sigma) - 1)h(n, \sigma)dnd\sigma = 1 - \int_{\sigma_0}^{\sigma_1} \int_{n_0}^{n_1} \sigma\gamma(n, \sigma)h(n, \sigma)dnd\sigma. \tag{48}$$

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

$$\begin{aligned}
\Sigma &= \int_{\sigma_0}^{\sigma_1} \int_{n_0}^{n_1} (1 - \sigma)\gamma(n, \sigma)h(n, \sigma)dnd\sigma = \int_{\sigma_0}^{\sigma_1} (1 - \sigma) \int_{n_0}^{n_1} \gamma(n, \sigma)h(n, \sigma)dnd\sigma \\
&= \int_{\sigma_0}^{\sigma_1} (1 - \sigma) \underbrace{\left( \frac{\int_{n_0}^{n_1} \gamma(n, \sigma)h(n, \sigma)dn}{\int_{n_0}^{n_1} h(n, \sigma)dn} \right)}_{= \mathbb{E}[\gamma(n, \sigma)|\sigma]} \underbrace{\left( \int_{n_0}^{n_1} h(n, \sigma)dn \right)}_{= \hat{h}(\sigma)} d\sigma. \tag{49}
\end{aligned}$$

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

$$\Sigma = \int_{\sigma_0}^{\sigma_1} (1 - \sigma)\mathbb{E}[\gamma(n, \sigma)|\sigma]\hat{h}(\sigma)d\sigma \geq \int_{\sigma_0}^{\sigma_1} (1 - \sigma)\hat{h}(\sigma)d\sigma = 0. \tag{50}$$

### 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

$$\int_{n_0}^{n_1} nl(n)f(n)dn = \int_{n_0}^{n_1} \left[ v(n) + \phi(l(n)) + (1 - \tau)\pi(n) \right] f(n)dn + G. \tag{51}$$

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

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

$$v'(n) = (1 - T'(z(n)))z'(n) - \phi'(l(n))l'(n). \tag{52}$$

Next, use the first-order condition from the profit maximization problem (4) and the rela-

relationship  $\pi(n) = nl(n) - z(n)$ . Condition (52) can then be written as

$$v'(n) = \frac{\phi'(l(n))}{n} \left[ l(n) - \pi'(n) \right]. \quad (53)$$

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

$$\pi'(n) = \mu l(n). \quad (54)$$

Intuitively, profits increase more rapidly in ability the higher is monopsony power and labor effort. Combining equations (53) and (54) gives

$$v'(n) = (1 - \mu) \frac{\phi'(l(n))l(n)}{n}. \quad (55)$$

From equation (55), the labor market payoff does not vary with ability if firms have full monopsony power (i.e., if  $\mu = 1$ ). 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 (51) and the incentive constraints (54) – (55). As it turns out, it is important to take the non-negativity constraint  $l(n) \geq 0$  explicitly into account. The final restriction that needs to be imposed 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. 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  $\tau$  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

$$\max_{z \geq 0} n\hat{l}(z, v(n)) - z. \quad (56)$$

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

$$\frac{n(1 - T'(z))}{\phi'(\hat{l}(z, v(n)))} - 1 = 0. \quad (57)$$

The second-order condition is strictly satisfied if the left-hand side of equation (57) is strictly decreasing in earnings  $z$ . The latter is true if and only if

$$-\phi''(l) - n^2T''(z) < 0, \quad (58)$$

where I used the first-order condition (57) and substituted out for  $\hat{l}(z, v(n)) = l$ . Because  $\phi(\cdot)$  is strictly convex, condition (58) is satisfied as long as the tax function is not too concave.

To determine how earnings  $z$  vary with ability, rewrite equation (57) and define

$$L(z, n) \equiv n(1 - T'(z)) - \phi'(\hat{l}(z, v(n))) = 0. \quad (59)$$

Next, apply the implicit function theorem and use the first-order condition (57) and the property  $\hat{l}_v = -1/\phi'$  to find

$$z'(n) = -\frac{L_n(z, n)}{L_z(z, n)} = \frac{\phi'(l) + \frac{\phi''(l)}{\phi'(l)}nv'(n)}{\phi''(l) + n^2T''(z)}. \quad (60)$$

From the incentive constraint (55),  $v'(n) \geq 0$  as long as monopsony power is not too quickly increasing in ability (which is assumed throughout). The numerator in (60) 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 (58) is satisfied. Therefore, if the allocation satisfies the monotonicity condition  $z'(n) > 0$ , it follows that the first-order condition for profit maximization (57) is both necessary and sufficient.

## V Lagrangian and first-order conditions

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

$$\begin{aligned}
& \max_{[v(n), \pi(n), l(n)]_{n_0}^{n_1}} \mathcal{W} = \int_{n_0}^{n_1} \left[ g(n)v(n) + (1 - \Sigma)(1 - \tau)\pi(n) \right] f(n)dn, & (61) \\
& \text{s.t.} \quad \int_{n_0}^{n_1} \left[ nl(n) - v(n) - \phi(l(n)) - (1 - \tau)\pi(n) \right] f(n)dn = G, \\
& \forall n : v'(n) = (1 - \mu) \frac{\phi'(l(n))l(n)}{n}, \\
& \forall n : \pi'(n) = \mu l(n), \\
& \forall n : l(n) \geq 0, \\
& \pi(n_0) \geq 0.
\end{aligned}$$

The corresponding Lagrangian is given by

$$\begin{aligned}
\mathcal{L} = & \int_{n_0}^{n_1} \left[ \left( g(n)v(n) + (1 - \Sigma)(1 - \tau)\pi(n) + \eta \left( nl(n) - v(n) - \phi(l(n)) - (1 - \tau)\pi(n) - G \right) \right) f(n) \right. \\
& \left. + \chi(n) \left( 1 - \mu \right) \frac{\phi'(l(n))l(n)}{n} + \chi'(n)v(n) + \lambda(n)\mu l(n) + \lambda'(n)\pi(n) + \psi(n)l(n) \right] dn \\
& + \chi(n_0)v(n_0) - \chi(n_1)v(n_1) + \lambda(n_0)\pi(n_0) - \lambda(n_1)\pi(n_1) + \xi\pi(n_0). & (62)
\end{aligned}$$

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

$$v(n) : (g(n) - \eta) f(n) + \chi'(n) = 0, \quad (63)$$

$$\pi(n) : (1 - \tau)(1 - \Sigma - \eta)f(n) + \lambda'(n) = 0, \quad (64)$$

$$l(n) : \eta (n - \phi') f(n) + \frac{\chi(n)}{n} (1 - \mu)(\phi' + \phi''l(n)) + \lambda(n)\mu + \psi(n) = 0, \quad (65)$$

$$\chi(n) : (1 - \mu) \frac{\phi'(l(n))}{n} - v'(n) = 0, \quad (66)$$

$$\lambda(n) : \mu l(n) - \pi'(n) = 0, \quad (67)$$

$$\eta : \int_{n_0}^{n_1} (nl(n) - v(n) - \phi(l(n)) - (1 - \tau)\pi(n) - G) f(n)dn = 0, \quad (68)$$

$$v(n_0) : \chi(n_0) = 0, \quad (69)$$

$$v(n_1) : -\chi(n_1) = 0, \quad (70)$$

$$\pi(n_0) : \lambda(n_0) + \xi = 0, \quad (71)$$

$$\pi(n_1) : -\lambda(n_1) = 0, \quad (72)$$

$$\psi(n) : \psi(n)l(n) = 0, \quad \psi(n) \geq 0 \text{ and } l(n) \geq 0, \quad (73)$$

$$\xi : \xi\pi(n_0) = 0, \quad \xi \geq 0 \text{ and } \pi(n_0) \geq 0. \quad (74)$$

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 an expression for the optimal marginal tax rate at each point in the income distribution. To that end, it is useful to first derive an expression for the multipliers  $\chi(n)$  and  $\lambda(n)$ . Combining equations (63) and (70) gives

$$\chi(n) = \chi(n_1) - \int_n^{n_1} \chi'(m) dm = - \int_n^{n_1} (\eta - g(m)) f(m) dm. \quad (75)$$

Evaluate equation (75) at  $n = n_0$  and use the transversality condition (69) and the normalization  $\int_{n_0}^{n_1} g(n) f(n) dn = 1$  to find

$$\int_{n_0}^{n_1} (\eta - g(n)) f(n) dn = \eta - 1 = 0. \quad (76)$$

Next, define by

$$\bar{g}(n) = \frac{\int_n^{n_1} g(m) f(m) dm}{1 - F(n)} \quad (77)$$

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 (64):

$$\lambda'(n) = (1 - \tau)\Sigma f(n). \quad (78)$$

Integrating and imposing the transversality condition (72) gives

$$\lambda(n) = -(1 - \tau)\Sigma(1 - F(n)). \quad (79)$$

Hence,  $\lambda(n) \leq 0$ . Equations (71) and (74) 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 (65). 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

$$nf(n) - \mu(1 - \tau)\Sigma(1 - F(n)) < 0, \quad (80)$$

where I imposed  $\eta = 1$  and substituted out for  $\lambda(n)$  using equation (79). The latter is true if the local Pareto parameter

$$a(n) = \frac{nf(n)}{1 - F(n)} < \mu(1 - \tau)\Sigma. \quad (81)$$

Hence, at ability levels where condition (81) holds, optimal labor effort and earnings are zero and the solution is not interior:  $l(n) = 0$  and  $z(n) = nl(n) - \pi(n) = 0$ . At ability levels where condition (81) does not hold, labor effort and earnings are positive and the optimal tax problem has an interior solution. Substituting  $\psi(n) = 0$ ,  $\eta = 1$ , and the first-order condition for profit maximization  $n(1 - T') = \phi'$  in equation (65) gives

$$T'(z(n))nf(n) = -\frac{\chi(n)}{n}(1 - \mu)(\phi' + \phi''l(n)) - \mu\lambda(n). \quad (82)$$

Substituting  $\chi(n)$  and  $\lambda(n)$  from equations (75) and (79), equation (82) can be written as

$$T'(z(n))nf(n) = (1 - \bar{g}(n))(1 - F(n))\frac{\phi'}{n}(1 - \mu) \left(1 + \frac{\phi''l(n)}{\phi'}\right) + \mu(1 - \tau)\Sigma(1 - F(n)). \quad (83)$$

Next, use the condition  $n(1 - T') = \phi'$  and denote by  $\varepsilon(n) = \frac{\phi'}{\phi''l(n)}$  the elasticity of labor supply. Rearranging gives equation (11) from Proposition 1.

## VII Effect of monopsony power on optimal marginal tax rates

To derive an expression for the optimal marginal tax rate if 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:

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

This is a closed-form solution for the optimal marginal tax rate faced by individuals with ability  $n$ .<sup>31</sup> To determine how the latter varies with monopsony power, differentiate equation (84) with respect to  $\mu$  to find

$$\frac{\partial T'(z(n))}{\partial \mu} = \frac{a(n)((1 - \tau)\Sigma - (1 + 1/\varepsilon)(1 - \bar{g}(n))) + (1 - \tau)\Sigma(1 + 1/\varepsilon)(1 - \bar{g}(n))}{(a(n) + (1 - \mu)(1 + 1/\varepsilon)(1 - \bar{g}(n)))^2}. \quad (85)$$

Equation (85) 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 effect 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 (85) by  $a(n)(1 - \tau)\Sigma(1 - \bar{g}(n)) > 0$ . The resulting expression is positive if and only if

$$((1 - \tau)\Sigma)^{-1} < ((1 + 1/\varepsilon)(1 - \bar{g}(n)))^{-1} + a(n)^{-1}. \quad (86)$$

<sup>31</sup>It is worth pointing out that the optimal marginal tax rate according to equation (84) 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 effect on welfare if  $(1 - \tau)\Sigma > 0$ .

## VIII Welfare effect of raising monopsony power

This Appendix analyzes the welfare effect of a proportional increase in monopsony power by  $\alpha$  percent. Hence, after the increase, monopsony power is  $\hat{\mu} = \mu(1 + \alpha)$ . Welfare is then given by

$$\begin{aligned} \mathcal{L}(\alpha) = & \int_{n_0}^{n_1} \left[ \left( (g(n) - \eta)v(n) + (1 - \Sigma - \eta)(1 - \tau)\pi(n) + \eta nl(n) - \phi(l(n)) - G \right) f(n) \right. \\ & + \chi'(n)v(n) + \chi(n)(1 - \mu(1 + \alpha)) \frac{\phi'(l(n))l(n)}{n} + \lambda'(n)\pi(n) + \lambda(n)\mu(1 + \alpha)l(n) \\ & \left. + \psi(n)l(n) \right] dn + \chi(n_0)v(n_0) - \chi(n_1)v(n_1) + \lambda(n_0)\pi(n_0) - \lambda(n_1)\pi(n_1) + \xi\pi(n_0), \end{aligned} \quad (87)$$

which is the optimized Lagrangian (62) evaluated at  $\hat{\mu} = \mu(1 + \alpha)$ . By the envelope theorem, the welfare effect is

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

Next, use equations (75) and (79) to substitute out for  $\chi(n)$  and  $\lambda(n)$ . This leads to

$$\frac{\partial \mathcal{W}(\alpha)}{\partial \alpha} = \int_{n_0}^{n_1} \left[ \frac{\phi'(l(n))}{n} \int_n^{n_1} (1 - g(m)) f(m) dm - (1 - \tau)\Sigma(1 - F(n)) \right] \mu l(n) dn, \quad (89)$$

This result can be rewritten as

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

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

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

The latter can be simplified further after defining

$$\Sigma^v = -\text{Cov}[v, \gamma] \geq 0, \quad (92)$$

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

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  $\sigma(1 - \tau)\bar{\pi}$ , which is proportional to the covariance between shareholdings and welfare weights introduced before. Substituting these terms in equation (91) gives

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

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

$$\mu\Sigma^v > (1 - \mu)\Sigma^k. \quad (95)$$

### Proof Corollary 2

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

$$\begin{aligned} \Sigma^v &= \int_{n_0}^{n_1} (1 - g(n))v(n)f(n)dn = \int_{n_0}^{n_1} (1 - g(n))(z(n) - T(z(n)) - \phi(l(n)))f(n)dn \\ &= \int_{n_0}^{n_1} (1 - g(n))(z(n) - T(z(n)))f(n)dn - \int_{n_0}^{n_1} (1 - g(n))\phi(l(n))f(n)dn \\ &= -\text{Cov}[z - T(z), \gamma] - \int_{n_0}^{n_1} (1 - g(n))\phi(l(n))f(n)dn \\ &= \Sigma^\ell - \int_{n_0}^{n_1} (1 - g(n))\phi(l(n))f(n)dn. \end{aligned} \quad (96)$$

Because  $g(n)$  is weakly decreasing in ability and averages to one, the second term on the last line of equation (96) 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

$$\mu\Sigma^\ell > (1 - \mu)\Sigma^k. \quad (97)$$

Unlike equation (95), 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 (i) the government also chooses a uniform benefit  $b$  paid to non-participants, and (ii) 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 (18).

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  $(n, \sigma, \varphi)$ . The average welfare weight is normalized to one. The welfare function can be derived using similar steps as in Appendix II:

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

$$\begin{aligned}
& + (1 - \tau)\bar{\pi} \int_{n_0}^{n_1} \int_{\sigma_0}^{\sigma_1} \int_{\varphi_0}^{\varphi_1} \gamma(n, \sigma, \varphi) \sigma h(n, \sigma, \varphi) d\varphi d\sigma dn \\
& = \int_{n_0}^{n_1} \left[ \int_{\varphi_0}^{\varphi(n)} g(n, \varphi) (v(n) - \varphi) k(n, \varphi) d\varphi + \int_{\varphi(n)}^{\varphi_1} g(n, \varphi) b k(n, \varphi) d\varphi \right] dn \\
& + (1 - \tau) \int_{n_0}^{n_1} \int_{\varphi_0}^{\varphi(n)} \pi(n) k(n, \varphi) d\varphi dn \int_{n_0}^{n_1} \int_{\sigma_0}^{\sigma_1} \int_{\varphi_0}^{\varphi_1} \gamma(n, \sigma, \varphi) \sigma h(n, \sigma, \varphi) d\varphi d\sigma dn \\
& = \int_{n_0}^{n_1} \int_{\varphi_0}^{\varphi(n)} \left[ g(n, \varphi) (v(n) - \varphi) + (1 - \tau)(1 - \Sigma)\pi(n) \right] k(n, \varphi) d\varphi dn \\
& + \int_{n_0}^{n_1} \int_{\varphi(n)}^{\varphi_1} g(n, \varphi) b k(n, \varphi) d\varphi dn. \tag{98}
\end{aligned}$$

As before,  $\Sigma$  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  $\varphi$ , averaged over shareholdings  $\sigma$ :

$$g(n, \varphi) = \frac{\int_{\sigma_0}^{\sigma_1} \gamma(n, \sigma, \varphi) h(n, \sigma, \varphi) d\sigma}{\int_{\sigma_0}^{\sigma_1} h(n, \sigma, \varphi) d\sigma}, \quad k(n, \varphi) = \int_{\sigma_0}^{\sigma_1} h(n, \sigma, \varphi) d\sigma. \tag{99}$$

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

$$\int_{n_0}^{n_1} \int_{\varphi_0}^{\varphi(n)} (T(z(n)) + \tau\pi(n)) k(n, \varphi) d\varphi dn = \int_{n_0}^{n_1} \int_{\varphi(n)}^{\varphi_1} b k(n, \varphi) d\varphi dn + G. \tag{100}$$

In words, the government collects income taxes  $T(z(n))$  and profit taxes  $\tau$  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  $T(z(n)) = nl(n) - \pi(n) - v(n) - \phi(l(n))$ :

$$\begin{aligned}
& \int_{n_0}^{n_1} \int_{\varphi_0}^{\varphi(n)} (nl(n) - v(n) - \phi(l(n)) - (1 - \tau)\pi(n)) k(n, \varphi) d\varphi dn \\
& = \int_{n_0}^{n_1} \int_{\varphi(n)}^{\varphi_1} b k(n, \varphi) d\varphi dn + G. \tag{101}
\end{aligned}$$

Using the definition for the participation rate (19), the final equation can be written as

$$\int_{n_0}^{n_1} \left[ p(\varphi(n))(nl(n) - v(n) - \phi(l(n)) - (1 - \tau)\pi(n)) - (1 - p(\varphi(n)))b \right] f(n) dn = G. \tag{102}$$

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

$$f(n) = \int_{\varphi_0}^{\varphi_1} k(n, \varphi) d\varphi. \tag{103}$$

The incentive constraints are the same as before and given by equations (54)–(55). This is because firms cannot observe participation costs. Hence, as in the model without a par-

ticipation 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 a benefit  $b$  and  $l(n)$ ,  $v(n)$ , and  $\pi(n)$  for all  $n$  to maximize social welfare (98) subject to the resource constraint (101), incentive constraints (54)–(55), and the requirements  $\pi(n_0) \geq 0$  and  $l(n) \geq 0$  for all  $n$ . The corresponding Lagrangian is

$$\begin{aligned} \mathcal{L} = & \int_{n_0}^{n_1} \left[ \int_{\varphi_0}^{v(n)-b} \left( g(n, \varphi)(b + (v(n) - b) - \varphi) + (1 - \tau)(1 - \Sigma)\pi(n) \right. \right. \\ & + \eta \left[ nl(n) - (v(n) - b) - b - \phi(l(n)) - (1 - \tau)\pi(n) - G \right] k(n, \varphi) d\varphi \\ & + \int_{v(n)-b}^{\varphi_1} \left( g(n, \varphi)b - \eta(b + G) \right) k(n, \varphi) d\varphi + \chi(n)(1 - \mu) \frac{\phi'(l(n))l(n)}{n} \\ & + \chi'(n)(b + (v(n) - b)) + \lambda(n)\mu l(n) + \lambda'(n)\pi(n) + \psi(n)l(n) \left. \right] dn \\ & + \chi(n_0)(b + (v(n_0) - b)) - \chi(n_1)(b + (v(n_1) - b)) + \lambda(n_0)\pi(n_0) - \lambda(n_1)\pi(n_1) + \xi\pi(n_0), \end{aligned} \quad (104)$$

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 (20)

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

$$\int_{n_0}^{n_1} \int_{\varphi_0}^{\varphi_1} (g(n, \varphi) - \eta) k(n, \varphi) d\varphi dn + \int_{n_0}^{n_1} \chi'(n) dn + \chi(n_0) - \chi(n_1) = 0. \quad (105)$$

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  $\eta$  gives the aggregate resource constraint (102):

$$\int_{n_0}^{n_1} \left[ p(nl(n) - v(n) - \phi(l(n)) - (1 - \tau)\pi(n)) - (1 - p)b - G \right] f(n) dn = 0. \quad (106)$$

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

$$\pi(n) : p(1 - \tau)(1 - \Sigma - \eta)f(n) + \lambda'(n) = 0, \quad (107)$$

$$l(n) : p\eta(n - \phi')f(n) + \frac{\chi(n)}{n}(1 - \mu)(\phi' + \phi''l(n)) + \lambda(n)\mu + \psi(n) = 0. \quad (108)$$

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,  $v(n) - b$ ):

$$p \left[ g_p(n) - \eta + \frac{p'}{p} \left( (1 - \tau)(1 - \Sigma)\pi(n) + \eta(nl(n) - v(n) - \phi(l(n)) + b - (1 - \tau)\pi(n)) \right) \right] f(n) + \chi'(n) = 0. \quad (109)$$

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  $\varphi \leq v(n) - b$ :

$$g_p(n) = \frac{\int_{\varphi_0}^{v(n)-b} g(n, \varphi) k(n, \varphi) d\varphi}{\int_{\varphi_0}^{v(n)-b} k(n, \varphi) d\varphi}. \quad (110)$$

The first-order condition with respect to labor effort (108) 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

$$pT'(z(n))nf(n) = -\frac{\chi(n)}{n}(1 - \mu)(\phi' + \phi''l(n)) - \mu\lambda(n). \quad (111)$$

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 (109) and use the transversality condition  $\chi(n_1) = 0$ :

$$\chi(n) = -\int_n^{n_1} p \left[ 1 - g_p(m) - \frac{p'}{p}(T(z(m)) + b + (1 - (1 - \tau)\Sigma)\pi(m)) \right] f(m) dm. \quad (112)$$

To derive an expression for  $\lambda(n)$ , substitute  $\eta = 1$  in equation (107) and rearrange to find

$$\lambda'(n) = p(1 - \tau)\Sigma f(n). \quad (113)$$

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

$$\lambda(n) = -(1 - \tau)\Sigma \int_n^{n_1} pf(m) dm. \quad (114)$$

Substituting the solutions for  $\chi(n)$  and  $\lambda(n)$  in equation (111) gives, after rearranging,

$$\begin{aligned} p(\varphi(n))T'(z(n))nf(n) = & \quad (115) \\ \mu(1 - \tau)\Sigma \int_n^{n_1} p(\varphi(m))f(m) dm + (1 - \mu)(1 - T'(z(n)))(1 + 1/\varepsilon(n)) \\ & \times \int_n^{n_1} p(\varphi(m)) \left[ 1 - g_p(m) - \frac{p'(\varphi(m))}{p(\varphi(m))}(T(z(m)) + b + (1 - (1 - \tau)\Sigma)\pi(m)) \right] f(m) dm. \end{aligned}$$

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 3:

$$T'(z(n)) = \frac{1 - F_p(n)}{nf_p(n)} \left[ \mu(1 - \tau)\Sigma + (1 - \mu)(1 - T'(z(n))) (1 + 1/\varepsilon(n)) \right. \\ \left. \times \mathbb{E} \left[ 1 - g_p(m) - \hat{p}(m)(T(z(m)) + b) - \hat{p}(m)\pi(m)(1 - (1 - \tau)\Sigma) \mid m \geq n \right] \right], \quad (116)$$

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

### Derivation of equation (21)

To derive the welfare effect of a proportional increase in monopsony power from  $\mu$  to  $\mu(1 + \alpha)$ , modify the Lagrangian (104) to

$$\mathcal{L}(\alpha) = \int_{n_0}^{n_1} \left[ \int_{\varphi_0}^{v(n)-b} \left( g(n, \varphi)(b + (v(n) - b) - \varphi) + (1 - \tau)(1 - \Sigma)\pi(n) \right. \right. \\ \left. \left. + \eta \left[ nl(n) - (v(n) - b) - b - \phi(l(n)) - (1 - \tau)\pi(n) - G \right] \right) k(n, \varphi) d\varphi \right. \\ \left. + \int_{v(n)-b}^{\varphi_1} \left( g(n, \varphi)b - \eta(b + G) \right) k(n, \varphi) d\varphi + \chi(n)(1 - \mu(1 + \alpha)) \frac{\phi'(l(n))l(n)}{n} \right. \\ \left. + \chi'(n)(b + (v(n) - b)) + \lambda(n)\mu(1 + \alpha)l(n) + \lambda'(n)\pi(n) + \psi(n)l(n) \right] dn \\ + \chi(n_0)(b + (v(n_0) - b)) - \chi(n_1)(b + (v(n_1) - b)) + \lambda(n_0)\pi(n_0) - \lambda(n_1)\pi(n_1) + \xi\pi(n_0). \quad (117)$$

By the envelope theorem,

$$\frac{\partial \mathcal{W}(\alpha)}{\partial \alpha} = \frac{\partial \mathcal{L}(\alpha)}{\partial \alpha} = \int_{n_0}^{n_1} \left( -\chi(n) \frac{\phi'}{n} + \lambda(n) \right) \mu l(n) dn. \quad (118)$$

Substituting in the solution for  $\chi(n)$  and  $\lambda(n)$  from equations (112) and (114) gives

$$\frac{\partial \mathcal{W}(\alpha)}{\partial \alpha} = \int_{n_0}^{n_1} \mu l(n) \left[ - (1 - \tau)\Sigma \int_n^{n_1} p(\varphi(m)) f(m) dm + \frac{\phi'(l(n))}{n} \right. \\ \left. \times \int_n^{n_1} p(\varphi(m)) \left[ 1 - g_p(m) - \frac{p'(\varphi(m))}{p(\varphi(m))} (T(z(m)) + b + (1 - (1 - \tau)\Sigma)\pi(m)) \right] f(m) dm \right] dn. \quad (119)$$

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

$$\frac{\partial \mathcal{W}(\alpha)}{\partial \alpha} = - (1 - \tau)\Sigma \int_{n_0}^{n_1} \pi(n) f_p(n) dn \\ + \frac{\mu}{1 - \mu} \int_{n_0}^{n_1} v(n) \left[ 1 - g_p(n) - \hat{p}(n)(T(z(n)) + b + (1 - (1 - \tau)\Sigma)\pi(n)) \right] f_p(n) dn \\ = - \Sigma^k - \frac{\mu}{1 - \mu} \int_{n_0}^{n_1} v(n) \hat{p}(n) \pi(n) (1 - (1 - \tau)\Sigma) f_p(n) dn \\ + \frac{\mu}{1 - \mu} \int_{n_0}^{n_1} v(n) \left[ 1 - g_p(n) - \hat{p}(n)(T(z(n)) + b) \right] f_p(n) dn, \quad (120)$$

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 3.

## 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 (23), the social welfare function (9) is modified to

$$\mathcal{W} = \int_{n_0}^{n_1} \left[ g(n)v(n) + (1 - \Sigma)(1 - (1 - s(\tau))\tau - \rho(s(\tau)))\pi(n) \right] f(n)dn, \quad (121)$$

where the term  $(1 - \tau)\pi(n)$  is replaced by  $(1 - (1 - s(\tau))\tau - \rho(s(\tau)))\pi(n)$ : see equation (22). The government budget constraint, in turn, is given by

$$\int_{n_0}^{n_1} \left[ T(z(n)) + \tau(1 - s(\tau))\pi(n) \right] f(n)dn = G, \quad (122)$$

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

$$\int_{n_0}^{n_1} \left[ nl(n) - v(n) - \phi(l(n)) - (1 - (1 - s(\tau))\tau)\pi(n) \right] f(n)dn = G. \quad (123)$$

The incentive constraints are the same as before (see Appendix III). The Lagrangian of the optimal tax problem is

$$\begin{aligned} \mathcal{L} = & \int_{n_0}^{n_1} \left[ \left( g(n)v(n) + (1 - \Sigma)(1 - (1 - s(\tau))\tau - \rho(s(\tau)))\pi(n) \right. \right. \\ & \left. \left. + \eta \left( nl(n) - v(n) - \phi(l(n)) - (1 - (1 - s(\tau))\tau)\pi(n) - G \right) \right) f(n) \right. \\ & \left. + \chi(n)(1 - \mu) \frac{\phi'(l(n))l(n)}{n} + \chi'(n)v(n) + \lambda(n)\mu l(n) + \lambda'(n)\pi(n) + \psi(n)l(n) \right] dn \\ & + \chi(n_0)v(n_0) - \chi(n_1)v(n_1) + \lambda(n_0)\pi(n_0) - \lambda(n_1)\pi(n_1) + \xi\pi(n_0). \end{aligned} \quad (124)$$

### Derivation of equation (24)

The first-order condition with respect to the profit tax  $\tau$  is

$$\int_{n_0}^{n_1} \left[ - (1 - \Sigma)(1 - s(\tau)) + \eta (1 - s(\tau) - s'(\tau)\tau) \right] \pi(n) f(n)dn = 0, \quad (125)$$

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 (125) by aggregate profits  $\int_{n_0}^{n_1} \pi(n) f(n)dn$  and rearrange to find

$$s'(\tau)\tau = \Sigma(1 - s(\tau)). \quad (126)$$

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 (24) from Proposition 4.

### Derivation of equation (25)

The expression for the optimal marginal tax rate is obtained from the first-order condition of the Lagrangian (124) with respect to labor effort:

$$\eta(n - \phi'(l(n)))f(n) + (1 - \mu)\frac{\chi(n)}{n}(\phi'(l(n)) + \phi''(l(n))l(n)) + \mu\lambda(n) = 0. \quad (127)$$

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

$$T'(z(n))nf(n) = -(1 - \mu)(1 - T'(z(n)))(1 + 1/\varepsilon(n))\chi(n) - \mu\lambda(n). \quad (128)$$

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  $\pi(n)$ :

$$\left[ (1 - \Sigma)(1 - (1 - s(\tau))\tau - \rho(s(\tau))) - \eta(1 - (1 - s(\tau))\tau) \right] f(n) + \lambda'(n) = 0. \quad (129)$$

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

$$\lambda(n) = - \left[ (1 - (1 - s(\tau))\tau - \rho(s(\tau)))\Sigma + \rho(s) \right] (1 - F(n)). \quad (130)$$

Substituting the solutions for  $\chi(n)$  and  $\lambda(n)$  in equation (128) and rearranging gives equation (25) from Proposition 4.

### Derivation of equation (26)

The welfare effect of raising monopsony power is:

$$\begin{aligned} \frac{\partial \mathcal{W}(\mu)}{\partial \mu} &= \frac{\partial \mathcal{L}(\mu)}{\partial \mu} = \int_{n_0}^{n_1} \left( -\chi(n) \frac{\phi'(l(n))l(n)}{n} + \lambda(n)l(n) \right) dn \\ &= \int_{n_0}^{n_1} \left( \frac{1}{1 - \mu} \int_n^{n_1} (1 - g(m))f(m)dm \underbrace{(1 - \mu) \frac{\phi'(l(n))l(n)}{n}}_{= v'(n)} \right. \\ &\quad \left. - \frac{1}{\mu} \int_n^{n_1} \left[ (1 - (1 - s(\tau))\tau - \rho(s(\tau)))\Sigma + \rho(s) \right] f(m)dm \underbrace{\mu l(n)}_{= \pi'(n)} \right) dn, \end{aligned} \quad (131)$$

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  $\pi(n_0) = 0$ :

$$\begin{aligned} \frac{\partial \mathcal{W}(\mu)}{\partial \mu} &= \int_{n_0}^{n_1} \left[ \frac{1}{1 - \mu} (1 - g(n))v(n) - \frac{1}{\mu} \left[ (1 - (1 - s(\tau))\tau - \rho(s(\tau)))\Sigma + \rho(s) \right] \pi(n) \right] f(n)dn \\ &= \frac{\Sigma^v}{1 - \mu} - \frac{\Sigma^k}{\mu} - \frac{\kappa(s(\tau))\bar{\pi}}{\mu}, \end{aligned} \quad (132)$$

where  $\Sigma^k$  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 (132) by  $\mu(1 - \mu)$  and set the resulting expression larger than zero. Rearranging gives the final result from Proposition 4.

## XI Additional figures

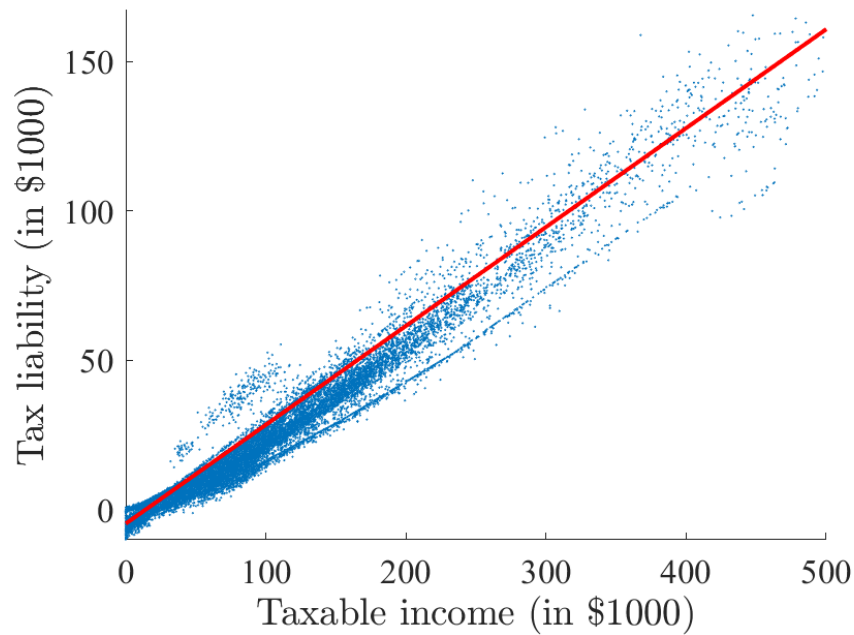


Figure 4: Current tax schedule

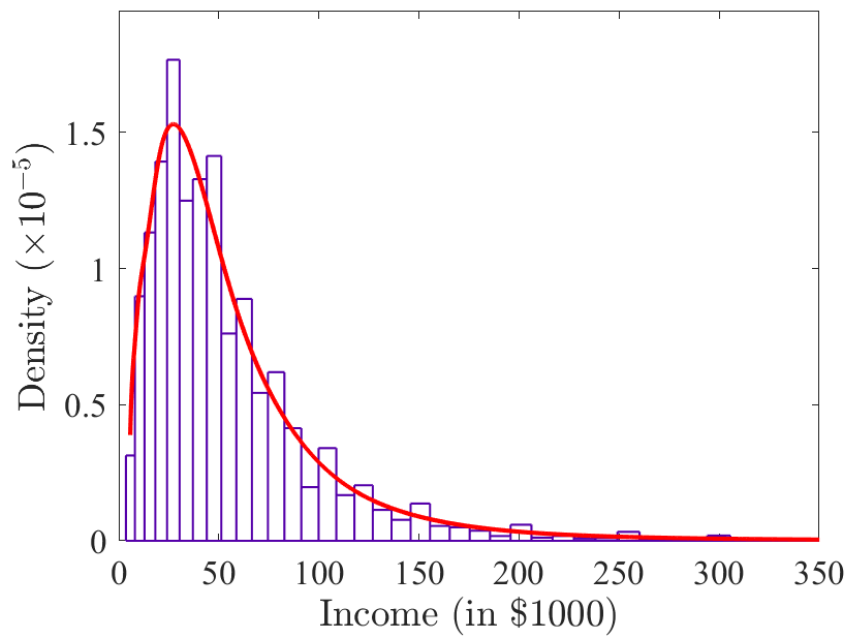
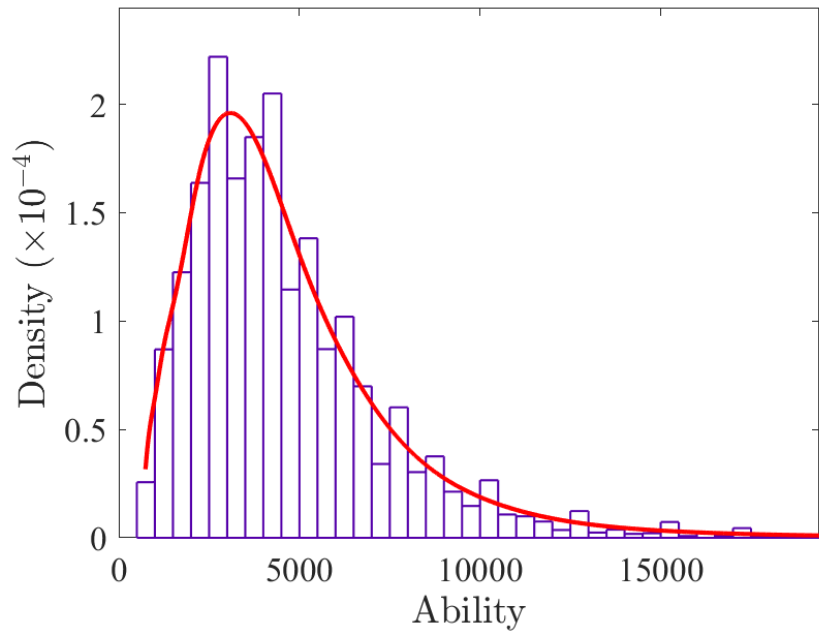


Figure 5: Distribution of ability and income

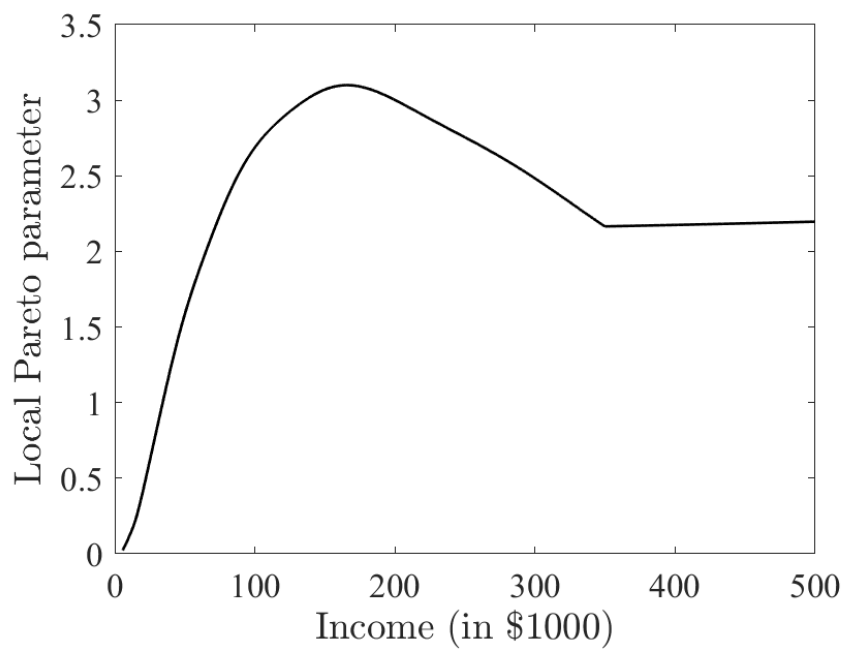


Figure 6: Local Pareto parameter

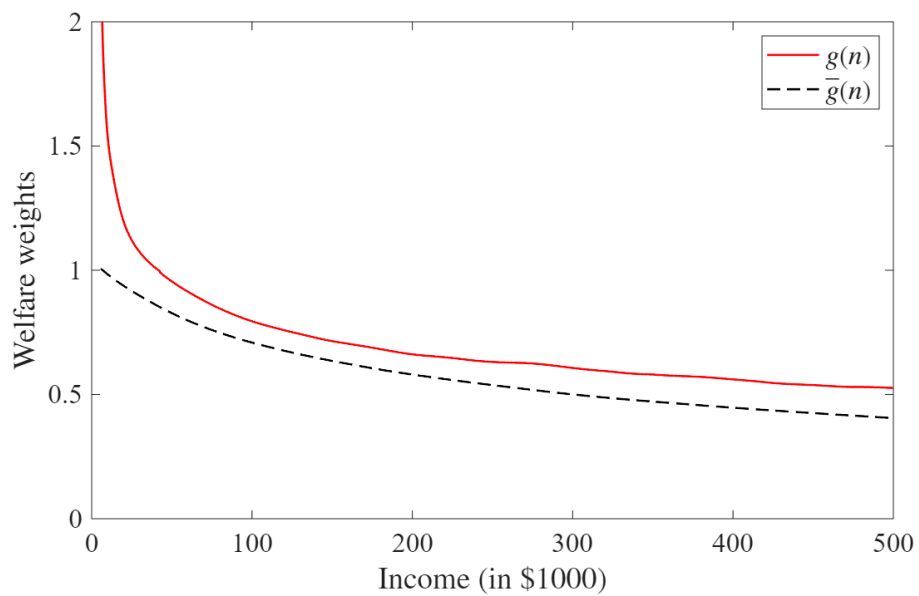


Figure 7: Welfare weights

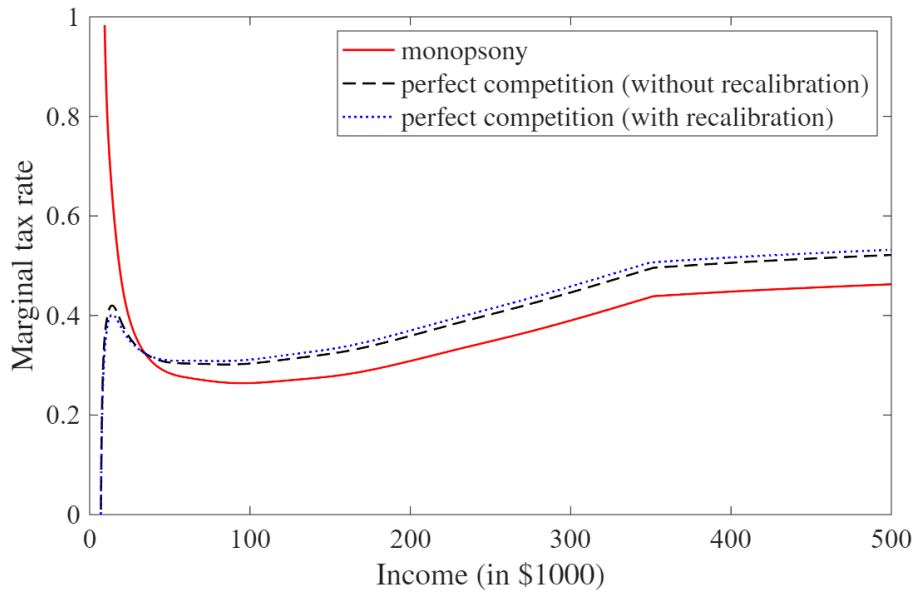


Figure 8: Optimal marginal tax rates (with and without recalibration)

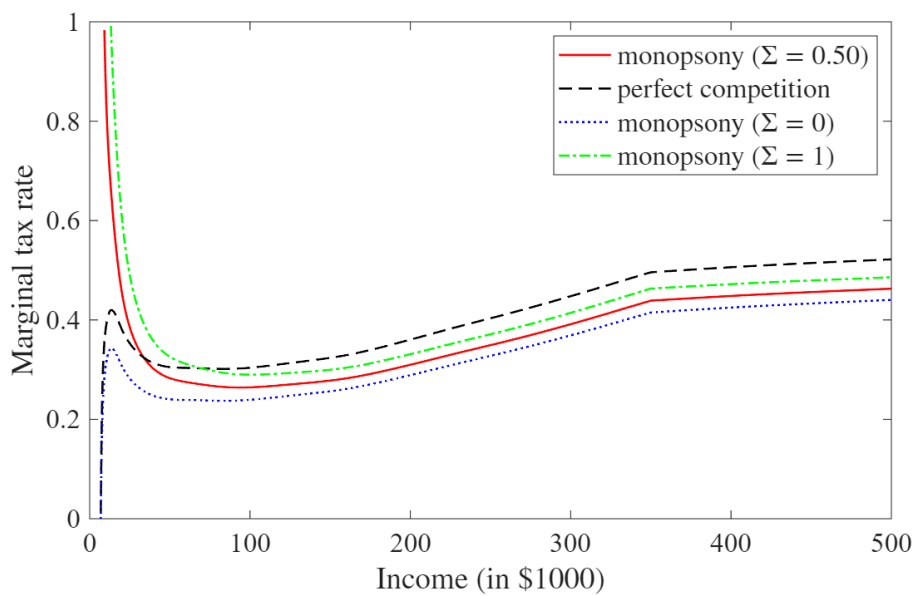


Figure 9: Optimal marginal tax rates (different values of  $\Sigma$ )

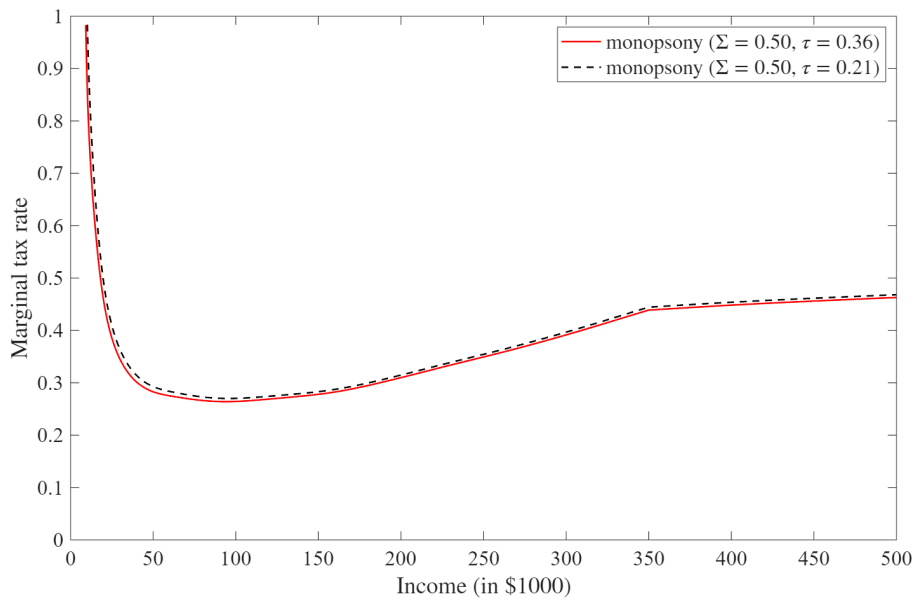


Figure 10: Optimal marginal tax rates (different values of  $\tau$ )