Taxing the Job Creators: Efficient Taxation with Wage Bargaining

Nicholas Lawson*
Aix-Marseille University (Aix-Marseille School of Economics), CNRS & EHESS
nicholas.lawson@univ-amu.fr

JOB MARKET PAPER

October 26, 2015

Abstract

The standard economic view of the personal income tax is that it is a distortionary way of raising revenue which nonetheless has value because of its desirable effects on the distribution of income. However, when wages deviate from marginal product, the laissez-faire (no-tax) equilibrium is inefficient, which can create an independent efficiency rationale for income taxation. I study a setting of wage bargaining within hierarchical teams, and show that the efficiency case for taxing managers depends on a "job-creation" effect: if wages are too low and increased labour supply allows managers to supervise larger teams and thus collect larger rents, they will have an incentive to devote an inefficiently high amount of effort to creating jobs at their firm. It may then be efficient to tax the "job creators" *because* of their job-creation activity. If bargaining compresses the wage distribution for workers, the efficient tax schedule will tend to be V-shaped, and numerical analysis of a calibrated model indeed finds a V-shaped efficient tax schedule with a top marginal rate of more than 25%. For a planner with redistributive motives, wage bargaining also causes optimal marginal tax rates to increase at the top and bottom of the distribution and decrease in the middle.

Keywords: optimal income taxation, efficiency, wage bargaining, hierarchical teams, job creators

*I thank participants of seminars at Princeton University, the Aix-Marseille School of Economics, and the Delhi School of Economics for their input, as well as participants of the National Tax Association 2014 Annual Conference in Santa Fe and the 2015 ZEW Public Finance Conference in Mannheim. I also thank Robin Boadway, Andrew Clark, Cecilia García-Peñalosa, Jacob Goldin, Mikhail Golosov, Nikolaj Harmon, David Lee, Patrick Pintus, Casey Rothschild, Marc Sangnier, Dean Spears and Alain Trannoy for their helpful comments and suggestions, as well as anonymous referees. Any errors or omissions are the responsibility of the author.

1 Introduction

The standard economic view of the personal income tax is that it is a distortionary way of raising revenue which nonetheless has value because of its desirable effects on the distribution of income.[1] This view has shaped the optimal income taxation literature that started with Mirrlees (1971), a literature that focuses on the simple setting of a perfectly competitive labour market despite growing evidence that wages are not generally equal to workers' marginal product.[2] However, if wages deviate from marginal product - due to wage bargaining, for instance - the laissez-faire (no-tax) labour market is inefficient, and this changes the normative consequences of income taxation; it is well known from the Theory of the Second Best that introducing a new distortion into a market that is already distorted has ambiguous welfare effects. In fact, there could well be an efficiency role for taxation, if marginal taxes are used to offset a pre-existing bargaining distortion and return labour supply to the efficient level, and thus the classic tradeoff between equity and efficiency could be altered, and perhaps weakened or even eliminated in some cases.

In this paper, I consider non-linear income taxation in a general equilibrium setting that accommodates the real-world divergence between employees' wages and their marginal product. I focus on production that takes place in hierarchical teams, in which lower-skill workers match with higher-skill managers. This setting represents two essential features of labour markets in developed countries: most individuals are employed in firms with two or more levels, so that workers at the bottom of the hierarchy answer to people higher up, and wages for the lower levels are set by managers at the top of the firm. The latter implies that if wages deviate from marginal product, the managers at the top will also receive returns which deviate from their actual contribution to output, so wage bargaining can generate inefficiency across the entire income distribution.

I begin by investigating these issues within a general model. I use a standard definition of efficiency as the maximization of net output, implying the usual result that the laissez-faire equilibrium with competitive wages is efficient; introducing a model of hierarchical firms

[1] For example, Blomquist, Christiansen, and Micheletto (2010) note that "The common view seems to be that marginal income taxes are purely distortive," and Sandmo (1998) argues that "distortionary effects of taxation...can only be justified from a welfare economics point of view by their positive effects on the distribution of income."

[2] See, for example, Manning (2003) and Manning (2011). The optimal income tax literature is surveyed by both Mankiw, Weinzierl, and Yagan (2009) and Diamond and Saez (2011).

does not change this fundamental fact. An efficiency role for taxation is introduced if wage bargaining causes the allocation to deviate from efficiency: a tax or subsidy can be used to correct workers' incentives. However, the main result is that the efficiency of taxing managers depends critically on the extent to which managers can control the size of their teams: if team size is fixed, there is no efficiency role for taxation of the manager at the top of the firm. If the manager acts as a residual claimant, their incentives are correct once worker labour supply has been set to the efficient value; there is one distorted margin, and thus one tax instrument needed to fix it. However, if team size is increasing in manager effort, this no longer holds, and if wages are below marginal product, efficiency will require a tax on the manager. This result follows from a "job-creation" effect: by working harder, the manager is able to accumulate more workers at the lower level of their firm's hierarchy that they can supervise and exploit for rents; therefore, if workers' wages are too low, the manager's "wage" per unit of labour supply is too high regardless of the level of worker effort. As a result, if wages are inefficiently low, the manager exerts too much effort in creating jobs at their firm, so a positive marginal tax reduces their labour supply towards the efficient level.[3] In other words, contrary to the common argument that taxes at high incomes should be lowered to encourage job-creation,[4] we may want to tax the "job creators" because they want to create *too many* jobs at their firm, and taxation of high-income individuals may be efficient as well as equality-enhancing.

I then provide a general characterization of the efficient tax schedule: if wage bargaining causes wages to be relatively flat with respect to skill, the efficiency-maximizing tax schedule is likely to feature marginal taxes that are regressive for workers and progressive for managers, generating a V-shaped marginal tax schedule. This suggests that a tax system in which low-income workers face high tax-back rates on social benefits, followed by lower marginal taxes at middle incomes and increasing marginal taxes near the top - as found in many developed countries - might actually be justified on efficiency grounds. However, while useful for highlighting the nature of the solution, the results of the general model are dependent on

[3]A management literature presents an alternative reason why managers may attempt to create too many jobs at their firm, as described in Jensen (1986): managers may wish to grow the firm beyond the efficient level in order to maximize the resources under their control and their resulting sense of power. Supporting evidence for such an "empire-building" motive is presented by Hope and Thomas (2008).

[4]See Krugman (November 22, 2011) for a discussion of this point; Krugman points out that this argument is dependent on high-income individuals not fully capturing the benefits that they produce for society. If workers' wages are below marginal product, the opposite may well be true.

a number of quantities for which there is no clear empirical counterpart. To understand whether or not the efficient deviation from zero marginal taxes is economically significant, we need a specific parametric model for numerical analysis.

Therefore, to illustrate the results from the general model, I focus on a specific parametric case of the model in the second half of the paper. I use a model adapted from Antràs, Garicano, and Rossi-Hansberg (2006) which features endogenous hierarchical one-to-many matching, in which lower-skill workers match with higher-skill managers to produce according to the team's ability to overcome problems encountered in production. I introduce labour supply into this model, which makes team size an increasing linear function of the manager's effort, and I examine equilibrium outcomes under competitive wage-setting and a simple form of wage bargaining between workers and managers. Given an underlying skill distribution, wage bargaining compresses the wage distribution for workers, while increasing the dispersion of returns received by managers.

I then consider the effects of taxation in a calibrated version of this model. I demonstrate that efficiency can be restored to a labour market that features wage bargaining using a tax that deviates significantly from zero, with a strong V shape that features rising marginal taxes over most of the income distribution and a top marginal rate of at least 25%. Wage bargaining causes equilibrium wages to deviate significantly from the efficient level, requiring substantial corrective taxes, and significant positive marginal taxes at the top of the income distribution can serve an important efficiency role in offsetting the bargaining power of the highest-skill managers.

The majority of my analysis focuses on the efficiency impact of taxation, because it is an important input into the optimal tax problem, as well as a being a useful baseline in public debates about income taxation when participants may have widely varying tastes for redistribution. However, to connect my results to those from the majority of the optimal income tax literature, in the final section of the paper I evaluate the optimal tax schedule from the perspective of a planner who cares about distribution. Specifically, I assume diminishing marginal utility of income, and I use a perturbation method to derive optimal tax rates for a utilitarian planner as a function of a direct redistribution effect, a distortion effect on the marginal individual, and a new component measuring how taxes shift the wage distribution. With competitive wage-setting, the optimal tax schedule takes an inverted-U shape with near-zero taxes at the top, but with wage bargaining this changes considerably: the optimal marginal tax rate increases significantly at the top and bottom, and drops in the middle of the distribution. In contrast to the standard result in the literature, this result is not sensitive to the magnitude of the labour supply response to taxation; what matters most is the magnitude of the deviation of wages from the efficient level.

My paper contributes to an important growing literature on optimal taxation in non-competitive labour markets. The majority of the optimal income taxation literature has focused on a competitive wage-setting environment; Piketty, Saez, and Stantcheva (2014) note that "There is relatively little work in optimal taxation that uses models where pay differs from marginal product." Varian (1980) is one of the very few early examples that deviates from this setting, considering a case in which variation in income is generated by random luck rather than effort. A more recent literature looks at taxation in the context of search and matching models, starting with several papers which focus on ex-ante identical populations: Boone and Bovenberg (2002) show how a linear wage tax can restore efficiency in a search and matching model, while Robin and Roux (2002) find that progressive taxation of workers can improve welfare by reducing the monopsony power of large firms. An important contribution is made by Hungerbühler, Lehmann, Parmentier, and van der Linden (2006), who examine the effect of taxes on vacancy creation with wage bargaining. They show that progressive taxes can reduce unemployment, with beneficial redistributional effects; however, inefficiency in the laissez-faire equilibrium is assumed away.[5]

Three important recent papers, meanwhile, highlight a role for taxation in settings in which wages do not capture the social return to labour supply.[6] Piketty, Saez, and Stantcheva (2014) argue that most of the responsiveness of income to marginal taxes that has been observed at high incomes comes from changes in rent-seeking rather than labour supply responses, and using rough estimates of those quantities, they find an optimal top tax rate of 83%. Lockwood, Nathanson, and Weyl (2014) focus on the possibility that a few skilled

[5]Also, by focussing on a setting of directed segmented search, in which workers match with vacancies in a continuum of separate labour markets, this paper ignores managers and executives and cannot say anything about income taxation at the upper end of the income distribution. Lehmann, Parmentier, and van der Linden (2011) subsequently extend the model to consider endogenous participation, and other related papers include Jacquet, Lehmann, and van der Linden (2013), who consider both extensive and intensive labour supply responses, and Jacquet, Lehmann, and van der Linden (2014), who consider endogenous participation with Kalai bargaining.

[6]Several other papers have examined taxation in models in which wages are not necessarily equal to marginal product, but these focus on very different settings; for example, Rothschild and Scheuer (2011) examine a labour market with a separate rent-seeking sector, while Stantcheva (2014) considers optimal taxation with adverse selection.

professions may generate important production externalities, and demonstrate that if those professions tend to be concentrated at particular points on the income distribution, non-linear taxation could internalize a portion of the externalities and improve efficiency. Finally, Rothschild and Scheuer (2014) present a very general framework of optimal taxation with multidimensional skill heterogeneity and externalities between workers, and analyze the deviation between the standard Pigouvian tax and the optimal corrective tax in this setting.

My most important contribution to this literature is to consider interactions between workers and managers in a general equilibrium model of production in hierarchical firms. In such a setting, if the wages of some workers deviate from marginal product, the ensuing rents must be collected by other individuals, meaning that the return to effort of the latter also deviates from their contribution to society. As my results highlight, it is important to recognize the bargaining relationships between individuals across the income distribution: my main result that managers should be taxed if they pay inefficiently low wages and can exert effort to increase team size has no parallel in a hierarchy-free model. My use of a general equilibrium model also differs from prior estimates of optimal taxes in this literature, which have focused on special cases or portions of the income distribution; no other study that I am aware of numerically evaluates the efficiency role for taxation in a non-competitive model of the entire income distribution, analogous to the standard Mirrleesian analysis.

Additionally, the results in previous papers in this literature tend to be driven by the introduction of a new margin impacted by income taxes: the results in Piketty, Saez, and Stantcheva (2014) depend on the elasticity of rent-seeking with respect to taxation, while Lockwood, Nathanson, and Weyl (2014) focus on externalities affecting other workers' production. My results present a new argument for efficiency-improving non-linear taxes: even if managers respond to increased taxes *entirely* on the labour supply dimension, non-zero marginal taxes on managers can improve efficiency if that labour supply leads to increased team size. Meanwhile, my numerical analysis fits within the general Rothschild and Scheuer (2014) setting: managers impose negative externalities on each other when wages are too low, as they seek to poach each other's workers to add to their own team. By focussing on models that are more fully specified than the general framework in Rothschild and Scheuer (2014), I am able to derive important new theoretical and numerical results.

I also contribute to a second literature that examines non-linear taxation in settings of occupational choice, beginning with the examination of redistributive taxation in team production with heterogeneous manager ability in Moresi (1998). More recent contributions include Rothschild and Scheuer (2013), which evaluates redistributive taxation in the Roy model; Ales, Kurnaz, and Sleet (2014), which looks at the implications of technical change for tax policy in a task-to-talent assignment model; Boadway and Sato (2014), who examine non-linear taxation in a model of extensive margin choice across occupations and uncertain earnings; and Scheuer (2014), who considers the effects of allowing for differential taxation of workers and managers. However, all of these papers assume efficiency in the labour market, and thus only consider income taxes for the purpose of redistribution. A partial exception is Scheuer (2013), who assumes adverse selection in the credit market used to finance entrepreneurship; he finds that this motivates a tax on entrepreneurs that is less progressive than the tax on workers. I abstract from credit markets and identify a force that acts in the opposite direction: with bargained wages that rise slowly with skill, the corrective tax that restores efficiency will generally be *more* progressive for managers than for workers.

The rest of the paper is organized as follows. Section 2 presents the general model of team production, and characterizes efficient taxes. Section 3 specifies the parametric model and describes the equilibrium under both competitive and bargained wages, and section 4 uses this model to estimate efficient taxes with wage bargaining. Finally, section 5 contains estimates of optimal taxes with diminishing marginal utility, and section 6 concludes the paper.

2 Taxation in a General Model of Team Production

I begin with a general model of production in two-layer hierarchical teams. This setting is intended to represent two essential features of the real-world labour market. First of all, most individuals are employed in firms with two or more levels, so that workers at the bottom of the hierarchy answer to people higher up and, ultimately, to the executives at the top. Second, wages for the lower levels are set, either through bargaining or subject to a competitive labour market, by managers at the top of the firm; therefore, if wages deviate from marginal product, the managers at the top of the firm will also receive returns which deviate from their actual contribution to output.

The model is deliberately very general, in a style analogous to the general model of social insurance in Chetty (2006). I do not specify the production function or the mechanism through which individuals match into teams, leaving both in a general form. The goal of this model is to show how much can be said about the efficiency impact of taxation in a model with minimal assumptions. However, I also present and solve a few simple examples in section 2.3, to demonstrate the applicability of the results from the general model.

I start by describing the model and defining the efficient allocation. If wages are subject to bargaining, the labour market will generally deviate from the efficient allocation, and I analyze the role of marginal taxes in restoring efficiency. My main result is that non-zero taxes on managers can only be justified from an efficiency perspective if team size is increasing in managerial effort, and I illustrate this result in settings of job poaching and matching between unemployed workers and firms; results are altered somewhat when external effects on the productivity of other teams are introduced in an example of a "reserve army" of unemployed individuals.

In the final subsection, I provide a general characterization of the shape of the efficient tax schedule: I show that, if wage bargaining causes the wage schedule for workers to be too flat with respect to skill relative to the efficient schedule, the efficient tax will tend to feature decreasing marginal rates among workers and increasing marginal rates among managers.

2.1 General Model and Efficiency

2.1.1 General Model

I assume that the population consists of a continuum of individuals with skill levels z from some distribution $F(z)$ and associated density $f(z)$; the distribution can be bounded or unbounded. Individuals match into 2-layer hierarchical teams, where $\mathbb{T}$ denotes the set of teams and $\tau \in \mathbb{T}$ indexes a particular team. Each team consists of one (infinitesimally small) manager with skill level z_m^τ, matched with a set of workers of size n^τ and skill level z_p^τ;[7] throughout the paper, I will use subscripts p (for production) and m to refer to workers and managers respectively. Note that the assumption of continuity means that each team is infinitesimally small, and thus both the manager and the set of workers are of measure zero; however, the set of workers is n^τ times as "large" as the manager, so that a mass of such managers of measure μ would be matched with a mass of workers of measure μn^τ.

[7] In the parametric model introduced in section 3, such perfect sorting of workers into teams will necessarily occur in equilibrium. Here, I simply assume that the technology of production ensures such sorting in equilibrium; for example, suppose each skill level of worker requires a different design of the production mechanism, making it most efficient to only use workers of a single type.

Teams are formed according to some general matching mechanism that is determined in equilibrium: $z_m^\tau = m(z_p^\tau)$, which can be a function or a correspondance. The nature of the matching mechanism is deliberately left as general as possible; the critical assumption is simply that production involves interactions between individuals in teams, an empirically reasonable assumption in developed countries.

Output Y^τ (in units of consumption good) of a team depends on the labour supplies and skill levels of the manager and workers: $Y^\tau = Y(L_m^\tau, n^\tau L_p^\tau; z_m^\tau, z_p^\tau)$, where L represents labour supply;[8] in what follows I drop the τ superscript except where necessary to distinguish between teams. I allow for the possibility that the number of workers n may be an increasing function of managerial labour supply: $n'(L_m) \geq 0$, as managers that work harder may be able to increase their span of control and supervise more workers.

This model has been described in a very general way, but encompasses a variety of settings, including the simple examples presented in section 2.3 and the parametric version of the model presented in section 3. The essential features are simply that individuals match in teams of two layers to produce, and that the manager in the top layer may be able to control the size of their team by altering their labour effort.

2.1.2 Defining Efficiency

The optimal tax problem is typically described as a tradeoff between equity and efficiency. In this context, efficiency is explicitly or implicitly defined as the maximization of net output, or output net of labour effort costs,[9] which also defines the first-best allocation with quasi-linear utility $U(C, L) = C - V(L)$, where $V_L, V_{LL} > 0$.

I therefore use the same definition: I assume quasi-linear utility, and define the efficient allocation to be the one that maximizes equally-weighted social welfare W:

$$W = \int_{f(z)>0} f(z)(C(z) - V(L(z))dz = \int_{\mathbb{T}} \big(Y^\tau - V(L_m^\tau) - n^\tau V(L_p^\tau)\big) d\tau.$$

[8] I abstract from physical capital, but the efficiency conditions and efficient tax equations are identical if capital is included in the model.

[9] This definition is often implicit, but see for example Hungerbühler, Lehmann, Parmentier, and van der Linden (2006), who casually refer to "Efficient (i.e. net output-maximizing) values of gross wages, employment, and net output" on page 748. By identifying a single efficient allocation (subject to regularity assumptions on the production function), this is a considerably more restrictive definition than Pareto efficiency; in many models, there is an entire envelope of Pareto efficient allocations that can be achieved by non-linear taxation, and indeed a tax schedule that maximizes any standard social welfare function must be Pareto efficient, as a Pareto improvement would necessarily increase social welfare. This expansiveness of the Pareto efficient space makes it a less useful definition of efficiency in an analysis of the efficiency-equity tradeoff and the specific efficiency role of taxation.

Thus, for now, I completely abstract from any redistributive motive on the part of a social planner (I will consider a desire for redistribution in section 5), and characterize the efficient allocation and the tax system, if any, required to achieve it. This question is of considerable significance given the centrality of the efficiency/equity tradeoff in the optimal income tax problem; a better understanding of the efficiency consequences of taxation is an important input into that problem, as well as a baseline for discussion among individuals who disagree on the socially optimal degree of redistribution.

To solve for the efficient allocation, I must set $\frac{dW}{dL}$ equal to zero for each individual. To abstract from impacts of an individual on the output of people not currently on their team, I make the following assumption:

Assumption 1. *The marginal impact of each individual's labour supply on total output is equal to the partial derivative of their team's output with respect to their own labour. That is, for type* $j = \{p, m\}$ *on team* i, $\frac{d\left(\int_{\mathbb{T}} Y^{\tau} d\tau\right)}{dL_j^i}|_{\bar{L}_j^i} = Y_j^i$, *where* $\bar{L}_j^i$ *represents the labour supply of all other individuals, so that* $\frac{d(\cdot)}{dL_j^i}|_{\bar{L}_j^i}$ *denotes the derivative holding all other labour supplies constant (but allowing the matching allocation to change), and where* $Y_m^i \equiv \frac{\partial Y^i}{\partial L_m^i} > 0$ *and* $Y_p^i \equiv \frac{\partial Y^i}{\partial n^i L_p^i} > 0$ *are the marginal products of manager and worker effort within a team.*

I refer to this as the "no production externalities" assumption; it implies the standard result that, if workers' wages are equal to their marginal product at their firm, as in a competitive market, the labour market achieves the efficient allocation. This assumption allows me to abstract from other potential sources of inefficiency in the labour market, and to focus on the inefficiency introduced by wage bargaining.[10] Assumption 1 is satisfied in many standard models, including the model of section 3 of this paper, and the first two examples in subsection 2.3 (but not in the "reserve army of unemployment" example in

[10] For Assumption 1 to hold, I require that (i) if a worker or manager works harder, they increase their own team's output, but their labour supply has no direct effect on the output of any other team, and (ii) if an individual's marginal change in labour supply triggers a reallocation of individuals across teams, this has no first-order impact on the productivity of the reallocated individuals, implying no net impact on total output. Part (i) is intuitive: since each individual is drawn from a continuum, they are all too small to have a first-order impact on the output of individuals of other teams, and indeed this part of the assumption is already embodied in the definition of the team production function. Part (ii) requires that the act of "creating" a job at one firm does not raise the productivity of the new worker by a finite amount in comparison to their old position, and is also an intuitive consequence of the continuum assumption: if the allocation of individuals across positions is efficient subject to wages, a reallocation should shift workers to positions where they are at most marginally more (or less) productive. Another way of putting this is that if the manager of team 1 works harder and thus is able to attract an additional worker from team 2, this reallocation raises the output of team 1 and decreases the output of team 2, but these reallocation effects cancel out on aggregate, leaving only the direct effect of the manager's effort on team 1 output.

subsection 2.3.3, where a failure of Assumption 1 introduces an additional externality term to the efficient tax equation even if workers' wages are equal to marginal product within their firm).

2.1.3 Solving for the Efficient Allocation and Market Equilibrium

With Assumption 1, I can solve for efficiency team-by-team, which means that I can focus on one representative team (conditional on skill levels). Since the rest of my analysis will be within-team, I now completely drop the τ superscripts to simplify notation. I find the following conditions for efficiency within a given team:

$$V_{L_p} = Y_p \tag{1}$$

$$V_{L_m} = Y_m. \tag{2}$$

Note that these are the standard conditions for allocative efficiency: marginal rate of substitution (here equal to marginal disutility of labour supply, since marginal utility from consumption is unity) must equal the marginal rate of transformation for all individuals. A necessary limitation of the general model is that these conditions are necessary but not sufficient for efficiency; without ruling out nonconvexities in production, I cannot exclude the possibility of (1) and (2) being satisfied at multiple allocations. However, in the examples presented in subsection 2.3, the necessary conditions are uniquely sufficient.

Now consider the market equilibrium for this general model. I assume that the manager is the residual claimant, whereas workers are paid a wage $w(z_p, z_m, \mathbb{L})$ per unit of labour supply, where $\mathbb{L}$ describes the aggregate labour market allocation. The wage may either be exogenously fixed from the perspective of the individual members of any team, or chosen by the manager to maximize profits; all that I require is that changing L_m at the firm level does not directly affect the wage at the firm, and that the wage is bounded away from zero and infinity so that it is possible to use policy to set the after-tax wage to the efficient level. In either case, in equilibrium the wage can be described as a general function of skill levels and aggregate labour market outcomes.

All individuals choose their labour supply to maximize their expected utility $U = C - V(L)$. Within any given team, workers' consumption is $C_p = wL_p$, and they will therefore choose a value of labour supply defined by:

$$w = V_{L_p}. \tag{3}$$

The manager's consumption is $C_m = Y(L_m, n(L_m)L_p) - n(L_m)wL_p$, and so their utility-maximizing choice of labour is defined by:

$$Y_m + L_p n'(L_m)(Y_p - w) = V_{L_m}. \tag{4}$$

By working harder, the manager not only receives their marginal output Y_m; they may also be able to supervise a larger team, which they value if they obtain positive rents from their workers, i.e. if $Y_p > w$. If the wage is also a choice variable for the manager, there will be an additional first-order condition for this choice, but this condition does not enter into the analysis of efficiency, as the wage affects the efficiency of the market equilibrium only through its effects on labour supplies. Thus, I leave the equation describing the equilibrium wage in a general form as $w(z_p, z_m, \mathbb{L})$.

Combining equations (1) through (4), I find the following necessary conditions for the equilibrium to be efficient:

$$w = Y_p \tag{5}$$

$$Y_m + L_p n'(L_m)(Y_p - w) = Y_m. \tag{6}$$

Notice that one sufficient condition to satisfy equation (6) for efficiency of the manager's labour supply is $Y_p = w$, exactly the same as the worker's condition. For efficiency to be satisfied, it is necessary that the worker's wage be equal to marginal product, as it would in a setting of perfect competition; this ensures that both the workers and the manager have the right incentives, and rules out any efficiency role for taxation. However, if the wage is set through some other mechanism, such as some form of bargaining, then the allocation will generally be inefficient, raising the question of whether policy could restore efficiency.

2.2 Efficient Taxation

Suppose that wages are not set competitively, but rather are subject to some bargaining process, and thus are generally not equal to marginal product. Further, suppose that the policy-maker wishes to restore efficiency to the labour market, and has access to marginal income taxes to do so. In principle, other policy instruments could fulfill the same role; for example, if wages are a choice variable for the manager, an output tax combined with a subsidy to the manager's wage bill could replicate the effect of a positive marginal income tax on the manager combined with a negative tax on the worker: both raise the after-tax return to effort for workers, and lower the after-tax return for managers. In other words, other policy instruments may well be identical in practical terms to income taxes, and my analysis would then amount to finding the optimal wedges to impose between before- and after-tax marginal returns. However, those other instruments could be more difficult to implement, if outcomes such as output and the wage bill are harder to measure than individual incomes.[11] I therefore assume that marginal income taxes are the only feasible set of instruments available to the policy-maker, as in the usual optimal income tax analysis.

I then assume that the policy-maker is able to implement marginal income taxes t_p and t_m on the workers and manager within each team,[12] and I ask the question: what taxes must they face in order to achieve the efficient allocation?

To answer this question, I solve for individual labour supply choices given marginal taxes:

$$(1-t_p)w = V_{L_p}$$

$$(1-t_m)[Y_m + L_p n'(L_m)(Y_p - w)] = Y_{L_m}.$$

and combining with (1) and (2), I find the following conditions for efficiency:

$$(1-t_p)w = Y_p \tag{7}$$

$$(1-t_m)[Y_m + L_p n'(L_m)(Y_p - w)] = Y_m. \tag{8}$$

Wages, once again, can either be exogenously fixed or variable from the perspective of the manager, and in equilibrium they may depend on taxes as well as labour market outcomes: $w = w(z_p, z_m, \mathbb{L}, t_p, t_m)$; however, they may be efficient or inefficient, depending on the wage-setting institution. By choosing a worker tax rate t_p to make $(1-t_p)w = Y_p$, the worker's labour supply can be set to the efficient level. Then t_m must be set to ensure efficiency of the manager's labour supply decision, and if firms are of fixed size, then $n'(L_m) = 0$ and (8) simplifies to $(1-t_m)Y_m = Y_m$, and so $t^*_m = 0$. Since w is not an argument in Y, Y_m depends

[11] For example, suppose a manager was able to "outsource" a job to an independent contractor, at least on paper, while in fact continuing to provide supervision to the worker in question; are payments to the contractor included in the wage bill, and does the output of the firm include the output of the contractor, or is it net of the contractor's output? Additionally, if wages were exogenously fixed from the perspective of the manager, a wage bill subsidy would be ineffective at raising the workers' marginal return.

[12] These marginal taxes are allowed to vary across the skill distribution, $t_p(z_p)$ and $t_m(z_m)$, but as before I focus on one team at a time and consider the marginal taxes that managers and workers within a team must face in order to support the efficient allocation as an equilibrium. I abstract from issues of incentive-compatibility and consider the taxes needed to restore the first-best allocation. Without income effects, only the marginal tax rates matter for labour market allocations.

on the wage only through the worker's labour supply L_p, and therefore if t_p is set to restore L_p to the efficient level, L_m will also be efficient, and there is no need for a non-zero t_m. Only one margin of choice is distorted, and so only one tax instrument is required to correct it.

If on the other hand $n'(L_m) > 0$, then the manager can acquire additional workers by exerting more effort, and thus the manager's level of labour supply depends directly on the wedge between the worker's wage and marginal product. If $w < Y_p$ in equilibrium, then the manager earns positive rents from their workers, and even if worker labour supply is corrected using a subsidy, managerial labour supply will be inefficiently high: each manager works too hard in order to accumulate additional workers in their firm, who they can then exploit for rents. As a result, the efficient marginal tax t_m^* on the manager is positive, and vice-versa if the wage is higher than marginal product. This result is summarized in the following proposition.

Proposition 1. *With one-manager/n-worker teams, where the manager is residual claimant and Assumption 1 ("no production externalities") holds:*
(i) if n is fixed, then conditional on an efficient effort choice by the workers, the manager's effort choice is efficient;
(ii) if n is fixed, and the wage-setting mechanism is such that a tax or subsidy on the worker can achieve the efficient worker labour supply, the efficient marginal tax faced by the manager is zero;
(iii) if teams are not of fixed size, with n increasing in the manager's labour supply, the efficient marginal tax faced by the manager takes the same sign as $Y_p - w$.

Proposition 1 is the central result of the paper: it tells us that wage bargaining between workers and managers does not by itself provide an efficiency argument for positive tax rates at the top of the income distribution. Suppose that wages are below marginal product (results are inverted if $w > Y_p$): workers must be subsidized in order to restore efficiency, but since the manager is assumed to be the residual claimant, their goal is to maximize surplus, so once their workers exert the efficient labour supply, the manager does as well if they have no power over team size. Non-zero marginal tax rates on the managers at the top of the distribution can only be justified from an efficiency perspective by a "job-creation" effect: because the manager receives rents from each worker they supervise, and because they can acquire more workers by working harder, their "wage" or private return to effort is too high even if worker labour supply is efficient, and they exert too much effort in "creating" jobs at their firm. A positive marginal tax reduces the manager's labour supply towards the efficient level.

A common argument among many politicians and in the media is that the tax system should reward job-creation, but in fact my results suggest the opposite conclusion in the presence of inefficiently low wages: we may want to tax high-income individuals or "job creators" *because* of their (excessive) desire to create jobs at their firm. Note further that this analysis could be extended to a multi-layer setting, and that this result does not depend on where the extra workers on the team are drawn from; if I allow some teams to be of size one and to feature a worker matched with themselves, the model could incorporate self-employment (where the worker/team produces output) and unemployment (where labour supply is search effort, and the output is the increase in future output from finding a job). What matters is that, for the competitive allocation to be sustained as an equilibrium, individuals' incentives must be aligned with their effect on output: individuals who receive a return to effort that is lower than their marginal product should be subsidized, while individuals who seek the rents from individuals working below them in their firm should be taxed.

This surprisingly strong conclusion is drawn only from the minimal assumptions of a very general model. It can be further illustrated with the use of simple and specific examples, which I pursue in the next subsection before returning to the general model for further analysis.

2.3 Three Examples of Efficient Taxation

To further illustrate the general model, this subsection presents three simplified examples. I show that the general results hold in both a setting of full employment, where increased manager labour supply only attracts additional workers away from other teams, and in a search and matching setting with unemployment, where teams of one worker and one manager are formed according to search effort from both parties. In both cases, efficiency requires both a subsidy to workers' wages and a tax on managers in the presence of inefficiently low wages. I conclude with a third example that demonstrates the limits of the general result: if there is a large "reserve army" of unemployed workers, and workers can only provide effort if plucked out of unemployment by managers, manager effort provides a positive externality to unemployed workers that must also be taken into account.

2.3.1 Example 1: Job Poaching

Suppose, for simplicity, that the economy consists of a mass of measure 1 of managers, and a mass of measure 1 of workers. Within a team of one manager and a set of size n of workers, output is $Y = L_m + nL_p$, and team size is also determined by managerial effort: $n = \frac{L_m}{E(L_m)}$, where $E(L_m)$ is the average value of L_m among all managers. That is, workers are shared among managers in proportion to their effort L_m, and as a result $n'(L_m) = \frac{1}{E(L_m)} > 0$.[13] I also assume that utility takes the simple quasi-linear form $U(C, L) = C - \frac{1}{2}L^2$, and therefore the efficient allocation is simple: $L_m = L_p = 1$.

Consider the market equilibrium when workers face a marginal tax rate of t_p and managers a rate of t_m. Assume for simplicity that the worker's wage is fixed at w, and then the worker's labour supply is equal to $L_p = w(1 - t_p)$. Meanwhile, the manager chooses L_m to maximize $U = (1 - t_m)(L_m + n(L_m)(1 - w)L_p) - \frac{1}{2}L_m^2$, with a result of:

$$(1 - t_m)\left(1 + \frac{(1 - w)L_p}{E(L_m)}\right) = L_m \tag{9}$$

and in equilibrium, all managers choose the same value of L_m, so the $E(L_m)$ in (9) can be replaced with L_m.

In this case, the efficient tax applied to workers is simple: t_p must be set to make $w(1-t_p)$ equal to one, and therefore $t_p = \frac{w-1}{w}$. If the wage is set lower than the marginal product of 1, the worker should receive a wage subsidy to encourage them to increase their labour supply to 1. The tax applied to managers, t_m, can be found as the value that sets $L_m = 1$ when $L_p = 1$; solving (9), this leads to the expression $t_m = \frac{1-w}{2-w}$, and as for the worker, this depends on whether or not the wage is equal to 1. If $w < 1$, so that the wage is inefficiently low, restoring efficiency requires a subsidy to workers and a tax on managers, exactly as in the general model. It is also easy to show that, if team size were fixed at 1, the manager's effort choice would be $L_m = (1 - t_m)$ regardless of worker effort, and no tax on managers would be required.

In both cases, the findings of the general model apply directly to this model. Here, the "job-creation" by managers is only at the level of their firm; on net, no new jobs are created,

[13] I thank Casey Rothschild for suggesting a model of this type; the current example is a simplified version of the model he suggested.

as any increase in effort by one manager simply poaches workers from his competitors. In the end, the allocation of workers across managers is the same as in the efficient allocation - one worker per manager - but managers exert excessive effort and would all be better off reducing their labour supply. But the intuition is the same: managers receive rents from their workers, and thus work too hard in attempting to accumulate additional workers at their firm in the absence of taxation.

2.3.2 Example 2: Search and Matching

As an alternative setting, consider a case of search and matching between prospective managers and unemployed workers. This setting has previously been studied by Boone and Bovenberg (2002); I will demonstrate that their model is a special case of my analysis.

Once again, assume that each group consists of a mass of measure 1. The model is a one-shot game, with all individuals starting out unmatched and searching for a match on the other side of the market. Each unemployed worker exerts effort L_p towards job search, and managers similarly undertake search L_m, with both groups receiving quasi-linear utility $U(C, L) = C - \frac{1}{\gamma} L^{\gamma}$. Matches are formed according to a standard Cobb-Douglas matching function: $M = L_m^{\alpha} L_p^{1-\alpha}$, and each match produces an output equal to 1.

Net output is $W = M - \frac{1}{\gamma}\left(L_m^{\gamma} + L_p^{\gamma}\right)$, and so the conditions for efficiency are simple; defining matching rates as $\theta_m \equiv \frac{M}{L_m}$ and $\theta_p \equiv \frac{M}{L_p}$, they can be written as:

$$(1-\alpha)\theta_p = L_p^{\gamma-1}$$

$$\alpha\theta_m = L_m^{\gamma-1}.$$

With a continuum of individuals, each one is unable to affect the economy-wide average values of L_p and L_m. Thus, workers face a fixed θ_p and their probability of finding employment is $\theta_p L_p$; therefore, if a worker receives a fixed wage w if they succeed in finding employment, and zero otherwise, their expected income is $w\theta_p L_p$. If this income is subjected to a tax t_p, their search effort choice will be given by $L_p^{\gamma-1} = (1 - t_p)w\theta_p$. Meanwhile, each manager obtains a match with a worker with probability $\theta_m L_m$, and receives $1 - w$ if a match is formed; if the manager faces a tax rate t_m, their search effort will be given by $L_m^{\gamma-1} = (1 - t_m)(1 - w)\theta_m$.

The efficient tax system is then straightforward to calculate: for the worker, the efficient tax is $t_p = \frac{w-(1-\alpha)}{w}$, and for the manager it is $t_m = \frac{1-\alpha-w}{1-w}$. Again, the critical value of w

is defined by the marginal product of worker labour: if $w < 1 - \alpha$, the worker should be subsidized and the manager taxed. This result is related to the "Hosios condition": Hosios (1990) identified the condition on worker bargaining power that would need to be satisfied for search equilibrium to be efficient, which in this case is $w = 1 - \alpha$, and subsequent papers have examined optimal tax policy when the Hosios condition is not satisfied. A prominent example is Boone and Bovenberg (2002), who find that a tax on firms and a subsidy to workers is efficient if the workers' bargaining power is too low, as in that case there would be both insufficient search by workers and excessive vacancy creation by firms.[14] My analysis demonstrates that this is in fact a special case of a more general result: in any setting in which individuals are matched together at different levels of a productive enterprise, a "job-creation" effect of one individual's effort, combined with wages that deviate from efficiency, is sufficient to require corrective taxes on multiple levels of the enterprise.

2.3.3 Example 3: "Reserve Army" of Unemployed

I conclude this series of examples with one that does not fit the general model, to demonstrate what is required for results to deviate from those presented in section 2.2. Suppose that, as before, there is a mass of measure 1 of managers, but the mass of prospective workers is very large, large enough that they cannot all be employed in equilibrium. A worker that is matched with a manager chooses a labour supply L_p and produces L_p units of output, but unemployed workers produce only an amount of home production which I normalize to zero. Each manager chooses their labour L_m, and while this has no direct effect on output, their team will be of size $n = L_m$; thus, by working harder, a manager plucks additional workers out of unemployment and into productive employment.

The utility function is $U(C, L) = C - \frac{1}{\gamma}L^\gamma$ as above, and thus the efficient allocation is, as usual, the one that maximizes net output:

$$W = L_m L_p - L_m \frac{1}{\gamma} L_p^\gamma - \frac{1}{\gamma} L_m^\gamma.$$

This gives the following expressions for the efficient allocation:

$$L_p = 1$$

$$L_m^{\gamma-1} = L_p - \frac{1}{\gamma} L_p^\gamma$$

[14] A similar result is found by Cahuc and Laroque (2014), who consider a monopsonistic labour market with only an extensive margin.

and applying $L_p = 1$ in the manager's condition gives the result that $L_m = \left(\frac{\gamma-1}{\gamma}\right)^{\frac{1}{\gamma-1}}$ must be true at the efficient allocation.

Assume as before that workers and managers face taxes of t_p and t_m. As in the previous examples, I assume that the wage is fixed at w for simplicity, and then each worker will set $L_p^{\gamma-1} = w(1-t_p)$, and thus the efficient tax will be $t_p = \frac{w-1}{w}$. The manager sets their labour supply such that $L_m^{\gamma-1} = (1-t_m)(1-w)L_p$, and thus if L_p is set to 1 by the efficient worker tax, the efficient manager tax must satisfy:

$$t_m = \frac{\frac{1}{\gamma} - w}{1-w}.$$

Therefore, the condition for efficient manager labour supply is $w = \frac{1}{\gamma}$, which is not the same as the worker's efficiency condition of $w = 1$ if $\gamma > 1$; the manager should only be taxed if $w < \frac{1}{\gamma}$. In fact, for wages in between $\frac{1}{\gamma}$ and 1, both the worker and the manager will need to be subsidized.

The way to understand this is to evaluate equations (7) and (8). The worker condition for efficiency is exactly as in (7), given that $Y_p = 1$ for any worker that actually exists in the labour market. For the manager, there is a bit of flexibility in how the terms are reconstructed, but the simplest is to say that $Y_m = 0$, as the manager's labour effort does not have any direct effect on output. Then, the left-hand side of (8), for the manager's choice of L_m, follows immediately since $n'(L_m) = 1$ and $Y_p - w = 1 - w$. But efficiency requires that this be set to $\frac{\gamma-1}{\gamma}$, not $Y_m = 0$, and the reason is that the manager's actions generate a positive production externality for someone outside the team: each additional worker hired by a manager is given a utility bonus of $\frac{\gamma-1}{\gamma}$ just due to their existence in the labour market. However, the manager, having hired the worker, cannot extract that surplus with only a wage at their disposal.[15]

This demonstrates the limit of applicability of the general model: production externalities between teams will alter the equation for optimal taxation. The competitive laissez-faire outcome with worker wage equal to their marginal product is not efficient; an additional motive for corrective taxation is introduced above and beyond the wage bargaining motive. Comparison of this result to the one from the search and matching case illuminates a common

[15]However, a competitive wage of $w = 1$ combined with a lump-sum fee of $\frac{\gamma-1}{\gamma}$ paid by any worker to their manager would achieve the efficient equilibrium; and in the presence of such a lump-sum job-finding fee, a bargained wage below $w = 1$ would once again motivate a negative tax on workers and a positive tax on managers.

public debate about the top tax rate: if job-creation at a particular firm takes individuals from other productive tasks, such as searching for a job (or from employment elsewhere), then managerial labour supply expended in attempting to accumulate workers and their rents is wasteful, motivating taxes on managers. On the other hand, if potential workers are made discretely more productive through the actions of an entrepreneur, and if that entrepreneur has no way of extracting that surplus from the worker, then this introduces a new motive for subsidizing job-creation - in that case, the competitive equilibrium is already inefficient. My analysis abstracts from the latter mechanism and focuses on the former, which fits many existing models, including the two previous examples and the parametric model to come in section 3.

2.4 Efficient Taxation Over the Skill Distribution

The idea that taxes can be used to offset pre-existing distortions in the labour market is related to the Theory of the Second Best: introducing a new distortion may well improve welfare when the market is already distorted. However, if I return to the general model with a few added assumptions, I can make some additional statements about the shape of the efficient tax schedule.

First of all, I assume that the equilibrium must feature what I call perfect positive assortative matching: the matching process will take the form of a function, i.e. a one-to-one mapping between worker skill level z_p and manager skill level z_m, with less-skilled agents (with z below some cutoff z^*) becoming workers and more-skilled agents becoming managers, and with the skill of a manager monotonically increasing in the skill of the workers in their team.[16] In the model presented in section 3, I prove that matching equilibrium must necessarily take this form.

I also make a few standard assumptions about the team production function: I assume diminishing marginal returns to both forms of labour, $Y_{pp} < 0$ and $Y_{mm} < 0$, and complementarity of manager and worker labour supply, $Y_{mp} > 0$. Then, I can consider the tax

[16] Spanos (2013) finds evidence in favour of these assumptions in French data: he finds support for the hypotheses that higher-ability workers tend to work in higher layers of the firm hierarchy, and that there is positive assortative matching between layers. These findings rely on the Abowd, Kramarz, and Margolis (1999) framework for estimating worker ability using fixed effects, and are subject to criticisms of this approach in Andrews, Gill, Schank, and Upward (2008) and Eeckhout and Kircher (2011). However, Andrews, Gill, Schank, and Upward (2012) also find evidence in favour of positive assortative matching using German social security records.

schedule over the set of workers, and separately over the set of managers.

For workers, the efficient tax must satisfy:

$$(1 - t_p)w = Y_p$$

and so, differentiating with respect to z_p:

$$\frac{dt_p}{dz_p} = \frac{1}{w}\left[(1 - t_p)\frac{dw}{dz_p} - \frac{dY_p}{dz_p}\right].$$

The sign of this derivative depends on the nature of the wage bargain. In many contexts, including the model of section 3, wage bargaining will tend to compress the wage distribution of workers, making wages rise slowly with skill. An extreme case consists of workers' wages being completely flat across the distribution, but wages are likely to rise more slowly than marginal product in a variety of models and bargaining environments, if workers receive some given share of the surplus but their actual relative contribution to that surplus rises with skill. For example, appendix B proves analytically that Nash bargaining must flatten wages for workers in a simplified version of the model from section 3.

Let me therefore assume that wages rise slowly with skill, so that $\frac{dw}{dz_p}$ will be relatively small; then L_p will rise slowly with skill and L_m will rise faster, as the managers are obtaining increasing rents. Given my assumptions about the production function, this will tend to make the marginal product of worker labour rise more quickly with skill, i.e. $\frac{dY_p}{dz_p}$ will be large. Therefore, with wages rising slowly with skill, $\frac{dt_p}{dz_p}$ will tend to be more negative, so the marginal tax schedule will be more likely to be downward sloping with respect to skill (and therefore presumably with respect to income).

Meanwhile, for managers, the efficient tax must satisfy:

$$(1 - t_m)[Y_m + L_p n'(L_m)(Y_p - w)] = Y_m$$

and if I use $\rho = Y_p - w$ to denote the rents obtained by the manager, this can be rewritten as:

$$t_m = \frac{\rho L_p n'(L_m)}{Y_m + \rho L_p n'(L_m)}$$

Differentiating with respect to z_m, I obtain:

$$\frac{dt_m}{dz_m} = \frac{\left[Y_m\left(\frac{d\rho}{dz_m}L_p n'(L_m) + \rho\frac{dL_p}{dz_m}n'(L_m) + \rho L_p n''(L_m)\frac{dL_m}{dz_m}\right) - \rho L_p n'(L_m)\frac{dY_m}{dz_m}\right]}{[Y_m + \rho L_p n'(L_m)]^2}.$$

Once again, the sign of this derivative depends on the nature of the wage bargain. First, I simplify by assuming $n''(L_m) = 0$, which holds in the case of the model studied in section 3. Then $\frac{dt_m}{dz_m} > 0$ will follow if and only if:

$$Y_m \left(\frac{d\rho}{dz_m} L_p + \rho \frac{dL_p}{dz_m} \right) > \rho L_p \frac{dY_m}{dz_m}$$

This also appears more likely to be satisfied if wages of workers increase slowly with skill; in that case, managers' rents increase quickly with skill and $\frac{d\rho}{dz_m}$ will be large, and while $\frac{dL_p}{dz_m}$ will be small, so will $\frac{dY_m}{dz_m}$, because L_p will rise slowly with skill and L_m will rise faster, reducing the marginal returns to manager effort.

These results tell us something about the efficient tax schedule: a bargained wage that rises slowly with skill will likely lead to a non-monotonic efficient tax schedule, with declining marginal rates among workers and increasing rates among managers. If the wage rises slowly, lower-skill workers will tend to be overpaid, and higher-skill workers will tend to be overpaid. This result will be mirrored for managers: the lowest-skill managers are matched with low-skill workers who are overpaid, making the managers' returns inefficiently low, whereas the highest-skill managers extract rents from their underpaid high-skill workers. Wage bargaining thus tends to compress the bottom of the income distribution, while raising the slope of incomes with respect to skill at the top. The gap between wage and marginal product will therefore be V-shaped, and as a result the efficient tax is also V-shaped: individuals at the bottom and top of the overall distribution are overpaid and need to be taxed, while those in the middle should receive a subsidy.

This suggests that a system in which low-income workers face high tax-back rates on social benefits, followed by lower marginal taxes at middle incomes and increasing marginal taxes near the top, similar to that found in many developed countries, might actually be justified on efficiency grounds. However, there is a limit to what we can learn from this general analysis; the equations depend on a variety of values on which there is no good empirical evidence. To obtain further insight and to see how far efficient taxes might deviate from zero, a parametric model of the labour market is required, and this is the focus of the next section of the paper.

3 Parametric Model of Hierarchical Teams

To provide further illustration of the general results presented in the previous section, I now consider a specific parametric case of the general model. In particular, I use a model of production in hierarchical teams adapted from Antràs, Garicano, and Rossi-Hansberg (2006). I will begin by presenting and explaining the model, and then I will solve for the equilibrium under both competitive wage-setting and a form of wage bargaining, and highlight the inefficiencies generated by wage bargaining. I will conclude the section with a presentation of the calibration of the model to the U.S. economy.

3.1 Model Setup

The model features a continuum of agents with skill levels $z \in [0, 1]$, distributed according to a continuous function $F(z)$ and associated probability density function $f(z)$, who match in teams of one manager and a set of n workers (of infinitesimal size as before).[17] The workers specialize in production, while the managers supervise the production process, and as before I use subscripts p and m to denote quantities attached to workers and managers respectively. The matching process is endogenous, but any equilibrium must feature perfect positive assortative matching; the proof of this result is presented in appendix A.

After agents form teams, wages are set, either competitively or as the result of a bargaining process, and workers and managers choose their labour supply, $L_p(z_p)$ and $L_m(z_m)$; from now on, I omit the z arguments from labour supply to simplify the notation. Then, during production, each worker faces a problem of difficulty d drawn from a uniform distribution over $[0, 1]$, and can solve any problem with difficulty less than or equal to their own skill level z_p. If the worker can't solve the problem, they communicate it to the manager, subject to a managerial time cost $h(z_m) > 0$, where $h'(z) \leq 0$; for the proof of uniqueness of equilibrium in appendix A, I also need to assume $h < 1$ in the competitive case, though that assumption is not required in the wage bargaining case.[18] The manager must spend $h(z_m)L_p$ units of time on each problem that is forwarded to them,[19] and can solve problems with $d \leq z_m$. In

[17] As in Antràs, Garicano, and Rossi-Hansberg (2006), I assume that self-employment is not feasible for reasons outside the model. For example, suppose that, even in the case of quasi-linear utility, individuals receive negative infinite utility when consumption is zero; this would ensure that nobody would choose self-employment, given the risk of receiving an unsolvable problem and producing no output.

[18] There is also an upper bound on h to ensure existence of equilibrium in the competitive case, as described in appendix A, which will be verified in numerical analysis.

[19] Workers who supply more labour are assumed to work on more aspects of production, and thus the

Antràs, Garicano, and Rossi-Hansberg (2006), the communication cost h is constant, but allowing higher-skill individuals to be able to solve problems faster, as well as a larger set of them, will later allow me to better calibrate the model to the long right tail of the real-world income distribution.

If a particular worker's problem is solved, L_p units of output are produced by that worker, whereas workers with unsolved problems produce nothing. The manager therefore spends $h(z_m)L_p(1-z_p)$ units of time in expectation on each worker, and given that the manager faces a continuum of workers of size n, the manager faces no uncertainty and has a managerial time constraint of $nh(z_m)L_p(1-z_p) = L_m$, and the team's total output is nL_pz_m. Thus, the equation for team size is $n = \frac{L_m}{h(z_m)L_p(1-z_p}$, and so each manager's team size is increasing linearly in their labour supply.

Individuals receive utility from consumption C and disutility from labour supply according to a utility function $U(C, L)$. A utility function with no income effect is common in the optimal income tax literature;[20] this implies that labour supply depends only on the marginal after-tax wage rate. To allow for diminishing marginal utility of income or a social taste for redistribution, which will be relevant in section 5, I specify utility as $U(C, L) = \frac{\left(C-\frac{1}{\gamma}L^{\gamma}\right)^{(1-\theta)}}{1-\theta}$, where θ controls how fast marginal utility declines with income. In the efficiency analysis to come later in this section and in the efficient tax analysis of section 4, I will assume $\theta = 0$ to focus on quasi-linear utility as before.

Workers choose their labour supply L_p to maximize utility, so a worker receiving a wage $w(z_p)$ will set $L_p = w(z_p)^{\frac{1}{\gamma-1}}$. The manager chooses the skill level of worker z_p he wishes to hire, which must be consistent with the equilibrium matching function, as well as labour supply L_m. The manager receives total consumption of $C(L_m) = nL_p(z_m - w(z_p)) = L_m\frac{z_m-w(z_p)}{h(z_m)(1-z_p)}$, and thus sets $L_m = \left(\frac{z_m-w(z_p)}{h(z_m)(1-z_p)}\right)^{\frac{1}{\gamma-1}}$, so that $r(z_m; z_p) \equiv C'(L_m) = \frac{z_m-w(z_p)}{h(z_m)(1-z_p)}$, which is the manager's return per unit of time, can be thought of as the manager's "wage."

I can then solve for the matching function given a particular wage function $w(z)$; since each agent represents an infinitesimally small space on the skill distribution, both functions will be continuous. If I denote z^* for the cutoff skill level at which individuals are indifferent between being a worker or manager, and $m(z)$ as the skill level of the manager who supervises

problem they face takes longer for the manager to study.

[20]The same assumption is made in Diamond (1998) and Persson and Sandmo (2005), and is described as a standard assumption by Lehmann, Parmentier, and van der Linden (2011).

workers of skill z, equilibrium in the labour market requires:

$$\int_0^{z_p} f(z)dz = \int_{m(0)}^{m(z_p)} n(m^{-1}(z))f(z)dz \quad \forall z_p \leq z^*.$$

Since $n = \frac{L_m}{h(z_m)L_p(1-z_p)}$, this can be rewritten as:

$$\int_0^{z_p} f(z)dz = \int_{m(0)}^{m(z_p)} \left[\frac{1}{h(z)(1-m^{-1}(z))}\right]^{\frac{\gamma}{\gamma-1}} \left[\frac{z - w(m^{-1}(z))}{w(m^{-1}(z))}\right]^{\frac{1}{\gamma-1}} f(z)dz \quad \forall z_p \leq z^*$$

and differentiating with respect to z_p and rearranging, I find that the matching function is defined by:

$$m'(z) = \left[\frac{(h(m(z))(1-z))^{\gamma}w(z)}{m(z)-w(z)}\right]^{\frac{1}{\gamma-1}} \frac{f(z)}{f(m(z))}. \tag{10}$$

Although this differential equation has no simple analytical solution, there is a fairly simple intuition behind it. $m'(z)$ is increasing in both $h(m(z))$ and $w(z)$: higher wages mean higher worker labour supply, and both that and higher communication costs mean a larger unit of managers is required to supervise a unit of workers. Meanwhile, $m'(z)$ is decreasing in z and $m(z)$: higher-skill workers require less supervision, and higher-skill managers prefer to work harder, so a smaller unit of managers is required per unit of workers. Finally, m' is increasing in worker density and decreasing in manager density: this is a simple mechanical effect, as higher density at a point means more individuals to be matched.

To solve for the equilibrium, the wage-setting mechanism must be described, and the following subsection presents and compares the two alternatives that I consider.

3.2 Competitive vs Bargaining Equilibrium

I now close the model by explaining the wage-setting mechanisms. I begin with competitive wages, and then describe the alternative of wage bargaining. Finally, I will contrast the results in the case of a uniform skill distribution, to illustrate the implications of the model; I show that, as expected, wage bargaining generates a wage function that is flatter for workers than in the efficient competitive allocation, and returns that are steeper with respect to skill for managers.

3.2.1 Competitive Wage Setting

The equilibrium will consist of two differential equations, one for the matching function and one for the wage function $w(z)$. As described above, in equilibrium, the manager's choice of z_p must be consistent with the matching function, so in the competitive case I assume that the manager faces a wage function $w(z)$ and must choose their preferred z_p. Thus, I differentiate the manager's rents $C = L_m \frac{z_m - w(z_p)}{h(z_m)(1-z_p)}$ with respect to z_p and set the derivative equal to zero,[21] solving for:

$$w'(z_p) = \frac{z_m - w(z_p)}{1 - z_p}.$$

Therefore, the equilibrium is defined by (10) and the equation describing the wage function:

$$w'(z) = \frac{m(z) - w(z)}{1 - z}$$

along with the boundary conditions $m(0) = z^*$, $m(z^*) = 1$, and $C(L_m(z^*)) - \frac{1}{\gamma} L_m(z^*)^\gamma = w(z^*)L_p(z^*) - \frac{1}{\gamma} L_p(z^*)^\gamma$, which ensures that individuals at z^* are indifferent between being a worker or a manager, and which simplifies to $w(z^*) = \frac{z^* - w(0)}{h(z^*)}$. The equation for $w'(z)$ has a simple intuition: if a manager chooses a more skilled set of workers, they save on supervision time and can supervise more workers, and the right-hand side expresses this gain, with the surplus $m(z) - w(z)$ scaled by one minus the worker skill level (with better workers, the gain from saved time in terms of additional workers that can be hired is proportionately larger). Meanwhile, the left-hand side captures the cost of increasing worker skill, in the form of higher wages.

3.2.2 Wage Bargaining

Alternatively, instead of a perfectly competitive labour market, I can also consider a simple form of bargaining over the wage per unit of worker labour supply.[22] Specifically, I assume sharing of expected output between the worker and manager, where bargaining takes place before production so that $w = \beta z_m$, and where β is set in equilibrium to clear the aggregate labour market.[23] Thus, bargaining takes place at the firm level, between the manager and each worker, but using a worker bargaining power β that is equal across the population and

[21] Essentially, the manager's first-order condition tells us what the slope of the wage function must be for $w(z)$ to be an equilibrium.

[22] I assume, as is common in search and matching models for instance, that the institution of wage bargaining is an unchangeable reality of the labour market, a social equilibrium that mandates managers and workers to provide and accept the outcome of such a bargain. Thus, policy cannot simply replace this institution with a competitive market.

[23] Bargaining thus takes place before workers choose their labour supply; managers can observe worker labour supply, and thus can offer a wage based on the actual worker effort, eliminating any role for conditioning wages on stochastic output.

determined in equilibrium: the worker's share of the output is the value that makes the individual at the cutoff z^* indifferent between becoming a worker or a manager.

In the absence of taxes, this is also the outcome of a Nash bargain over the surplus when the worker and manager fallback values are set to zero. However, I will focus on the simple output-sharing specification, for two reasons. The first is that the Nash bargaining solution becomes far more convoluted when non-zero taxes are introduced, as wages shift with the marginal and average tax rates faced by both the worker and the manager; computation of optimal taxes in this setting proved difficult.[24] More importantly, I actually want to abstract from a response of bargained wages to taxes, to show that my results do not depend on this feature; Piketty, Saez, and Stantcheva (2014) have already demonstrated that optimal taxes can be raised by a strong response of rent-seeking to taxation - which is analogous to a change in a bargained wage - but I show that efficient taxes can be non-zero even if the response of incomes to taxes is driven by changes in labour supply.

With this output-sharing rule, the matching function simplifies to:

$$m'(z) = \left[\frac{\beta(h(m(z))(1-z))^{\gamma}}{1-\beta}\right]^{\frac{1}{\gamma-1}} \frac{f(z)}{f(m(z))}. \tag{11}$$

This equation defines the equilibrium, along with the wage equation $w(z) = \beta m(z)$, the boundary conditions $m(0) = z^*$ and $m(z^*) = 1$, and the condition of indifference at z^*, which simplifies to $\beta = \frac{z^*}{h(z^*)+z^*}$. Thus, β adjusts in equilibrium to ensure that the marginal individual is indifferent between being a worker and a manager.

3.2.3 The Inefficiency of Wage Bargaining

In this model, the actions of the members of any particular team have no impact on the productivity of individuals on other teams, as there is full employment and all individuals are too small to affect equilibrium wages. Therefore, the laissez-faire equilibrium in the competitive case is efficient. Wages in the bargaining case, however, will generally deviate from the competitive wages, and thus it is clear that the allocation with wage bargaining will be inefficient. With wages distorted from their efficient levels, labour supplies are also distorted, with some workers (those with wages that are too low) supplying too little effort,

[24] Efficient taxes with Nash bargaining will deviate less from zero than with simple output-sharing, because taxes shift after-tax wages faster; in the uniform-distribution case, the efficient top tax rate is 40.14% with Nash bargaining, as opposed to 60.55% with output-sharing. However, optimal taxes with diminishing marginal utility from income will tend to be more progressive with Nash bargaining, as progressive taxes induce indirect redistribution by increasing wages at lower incomes.

and some supplying too much, working beyond the point when their contribution to society equals the marginal utility cost from effort.

A simple illustration can be provided by simulating the model with a uniform skill distribution and a constant value of $h = 0.5$, which is in the middle of the permissable values.[25] Here and for the remainder of the main numerical analysis of the paper, I assume a compensated elasticity of taxable income equal to 0.25,[26] implying that $\gamma = 5$, and I use a population with 10001 mass points at $\{0, 0.0001, ..., 1\}$ as an approximation to a continuous distribution.

Figure 1 presents the results for both the competitive and bargaining cases. The wage functions in (a) display workers' wages $w(z)$ up to z^*, and then the managers' "wages" $r(z_m)$ for values of z above z^*; meanwhile, the figure for the matching function in (b) presents $m(z)$ up to z^* and then the inverse matching function $m^{-1}(z)$ beyond that point. Starting with panel (b), z^* takes a value of about 0.8 in both cases, so that individuals with z above that value become managers and individuals below z^* become workers.[27] One important characteristic of the matching functions is that they flatten out as z approaches z^*, indicating that higher-skill managers are able to supervise more workers because both L_m and z_p are higher, the latter meaning that each worker can solve more problems and bothers the manager less frequently. Meanwhile, both wage functions exhibit a kink at z^*: the wage rises more rapidly to the right of z^*. This confirms that it is an equilibrium for individuals below z^* to become workers and those above z^* to become managers: for a given skill level, the "wage" earned as a manager is higher than that earned as a worker for $z > z^*$ and lower for $z < z^*$.

The matching functions look quite similar in the two cases, but we can see from panel (a) that the same is not true of the wage functions: in the bargaining case, wages are flatter with respect to skill for workers, and steeper for managers. As expected from the previous

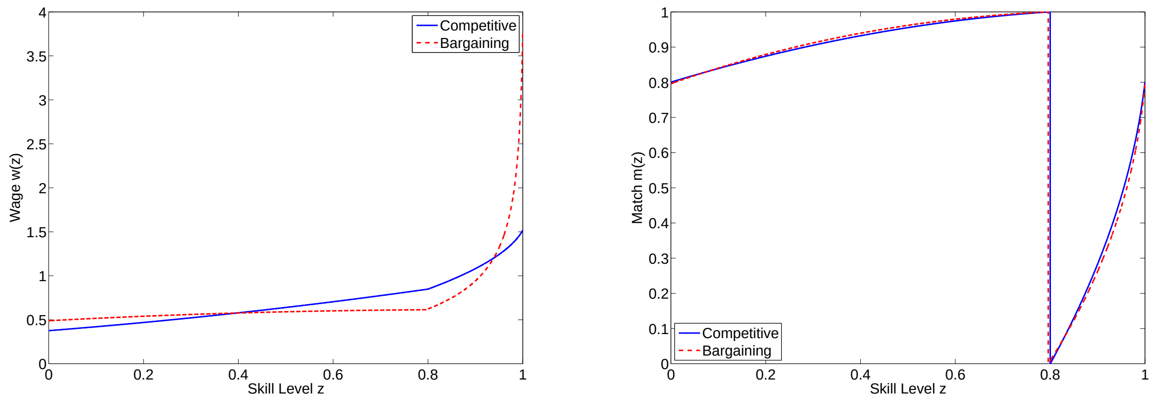
[25]When I calibrate a flexible h (along with the skill distribution) in the following subsection, I end up with h that ranges from 0.20 to 0.61 in the competitive case and from 0.33 to 1.10 in the bargaining case, suggesting that $h = 0.5$ is a reasonable compromise for this illustration, in which my goal is to compare outcomes from competition and bargaining in the same parameterized model.

[26]An elasticity of 0.25 is suggested by Saez (2001), and Saez, Slemrod, and Giertz (2012) select it as the approximate midpoint of a range of plausible estimates from 0.12 to 0.4. I perform a sensitivity analysis on my efficient and optimal tax results in appendix F.2, using a value of 1 for the elasticity of taxable income, and show that my conclusions are robust to this modification.

[27]Along with labour supplies being distorted, it is also possible that there are too many workers, or too many managers; that is, that z^* is too high or too low. In this case, the deviation of z^* in the bargaining case from its efficient value is small, because some managers work too hard while others work too little, with the same also true for workers; the efficient z^* is 0.800, whereas in the bargaining case the value is 0.796. Therefore, there are slightly too many managers in the bargaining case.

Figure 1: Wage and Matching Functions with Uniform Distribution & Constant h

(a) Wage Function $w(z)$ (b) Matching Function $m(z)$



section, wage bargaining generates a deviation between the market return per hour of labour supply and the marginal product that is V-shaped over the skill distribution. Because wages are too flat for workers, the lowest-skill workers are overpaid, while the highest-skill workers - those with z just below z^* - are underpaid. With perfect positive assortative matching, the opposite is true for the managers. As a result, labour supplies also deviate from the efficient levels, with middle-skill individuals working too little, and low-skill workers and high-skill managers working too hard. This misallocation of labour can have significant efficiency consequences; in this case, if utility is quasi-linear in consumption (i.e. $\theta = 0$), the bargaining equilibrium features average utility that is 0.65% of mean consumption lower than the first-best.

As indicated in section 2.4, this is a general result of many wage bargaining specifications. Indeed, I prove analytically in appendix B that, in a simplified version of the current model, my specification of simplified Nash bargaining *must* flatten the wage function for workers. The intuition is as follows: complementarity of worker and manager effort makes the worker's marginal product, and thus the competitive wage function, convex with respect to worker skill, whereas the bargained wage function is concave in skill due to a concave matching function. Essentially, wage bargaining implies that workers receive some given share of the surplus, whereas their actual relative contribution to that surplus rises with skill.

This simple analysis has used a uniform skill distribution and a constant value of the managerial supervision cost h; however, such a parameterization, while useful for a first illustration of the model, does not provide a good fit to reality. In the next subsection, I describe my calibration of the model to the U.S. income distribution.

3.3 Calibration of Model

In what follows, I present the calibration of the model; I begin by calibrating the wage bargaining case of my model to the U.S. income distribution. As described in appendix C, I use data from the 2013 March CPS to estimate the real-world income distribution, and for any set of parameters I can evaluate the model's income distribution. My goal is to match the model's distribution as closely as possible to that from the CPS, where both distributions are smoothed using a kernel density.[28] I assume that the baseline tax system is represented by the approximation to the U.S. income tax used by Jacquet, Lehmann, and van der Linden (2013): a linear tax at rate 27.9% and a lump-sum transfer of $4024.90, which they argue is a good approximation to the real tax schedule of singles without dependent children according to the OECD tax database.

I calibrate both the skill distribution $f(z)$ and the managerial time cost $h(z)$ in a flexible way. Specifically, I allow the skill density to be a cubic spline across 11 nodes at $\{0, 0.1, ..., 1\}$, whereas h takes a functional form with three parameters: $h(z) = h_1 \exp\left(-h_2(z^{h_3})\right)$. I then search over the values of the density at the 11 nodes, and the three parameters of $h(z)$, to find the parameters that minimizes the sum of squares across the income distribution of a function defined as the difference between the resulting kernel income densities multiplied by income. The level of h, as determined primarily by h_1, is identified from how spread-out the distribution is, as a lower h means managers can hire more workers and receive larger returns, while the curvature of h as the top, as determined by h_2 and h_3, helps to fit the long right tail to the income distribution. Meanwhile, a flexible skill distribution helps to match the shape of the income distribution at a closer level.

Upon finding the combination of parameters that best match the bargaining case of the model to the income distribution, I then perform the same calibration on the competitive version of the model, but in this case I use the distribution from the calibrated bargaining

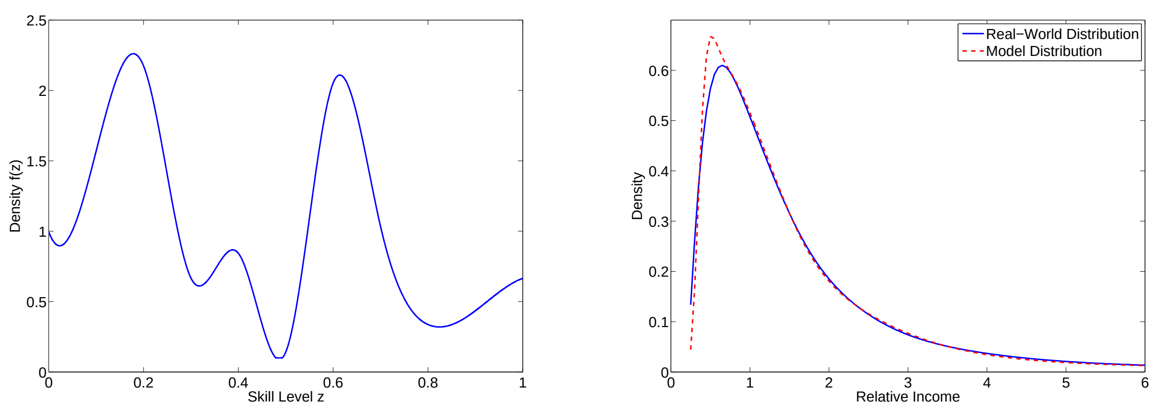
[28] I use a Gaussian kernel over log income to smooth the distributions. The CPS data surely contains measurement error, and the model can only be a first-order approximation to reality, so it is not realistic to attempt to match the distributions at a precise micro level; instead, I use kernel bandwidths that smooth out all obvious spikes and irregularities in either distribution, which are 0.2 for the CPS data and 0.25 for the model, and look for a good match for the overall shape of the income distribution. An alternative would be to try to match income levels at specific quantiles of the distribution, which would not require any kernel smoothing; results available upon request show that the match of my calibrated model to the data is also quite good on this dimension.

case as the target; since I will compare optimal tax schedules in the two cases later in the paper, it is especially important that they provide a close fit to each other.

The procedure of calibration is described in further detail in appendix C. In the wage bargaining case, the resulting function for h is $1.100 \exp\left(-1.195(z^{80.4})\right)$,[29] and the skill distribution and the fit to the real-world kernel density are displayed in Figure 2; the fit is very good over most of the distribution, but the model generates slightly too few people at very low incomes and too many at slightly higher incomes. Meanwhile, in the competitive case, the calibration produces $h(z) = 0.607 \exp\left(-1.094(z^{25.5})\right)$, and the skill distribution and fit to the bargaining case displayed in Figure 3.

Figure 2: Calibration Results with Wage Bargaining

(a) Calibrated Skill Density $f(z)$ (b) Fit of Income Kernel Density



4 Efficient Taxation

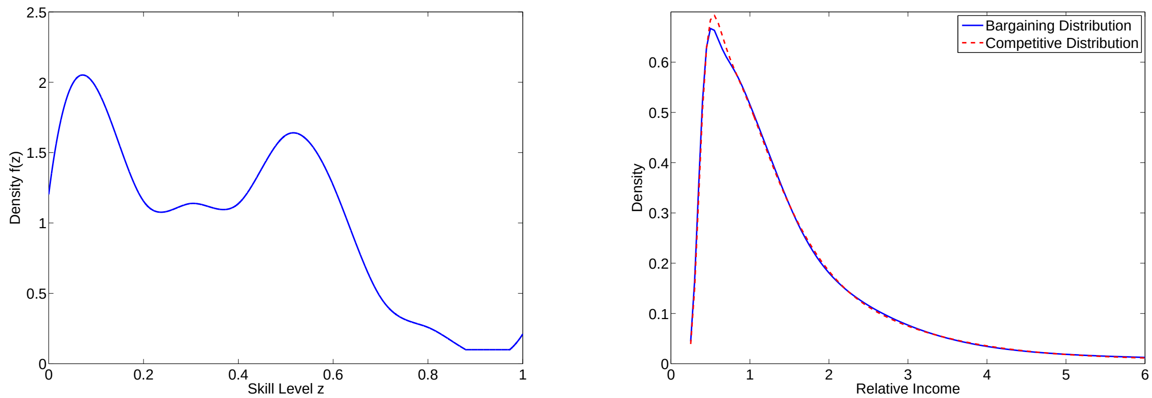
The analysis of the parametric model so far is consistent with the discussion in section 2.4: in the simple uniform version of the model, Figure 1 demonstrated that with wage bargaining, wages for workers tend to be too flat with respect to skill, whereas the returns for managers rise too quickly, generating a V-shaped pattern for the gap between wage and marginal product.

I can show that the exact same result holds in the calibrated version of the model. Using the parameters of the bargaining calibration, I compute the bargaining equilibrium *and* the

[29] As described in section 4, 1.1 is imposed as an upper bound for h_1, to ensure that a competitive allocation can be calculated as a target for efficient taxation.

Figure 3: Calibration Results with Competitive Wages

(a) Calibrated Skill Density $f(z)$ (b) Fit of Income Kernel Density



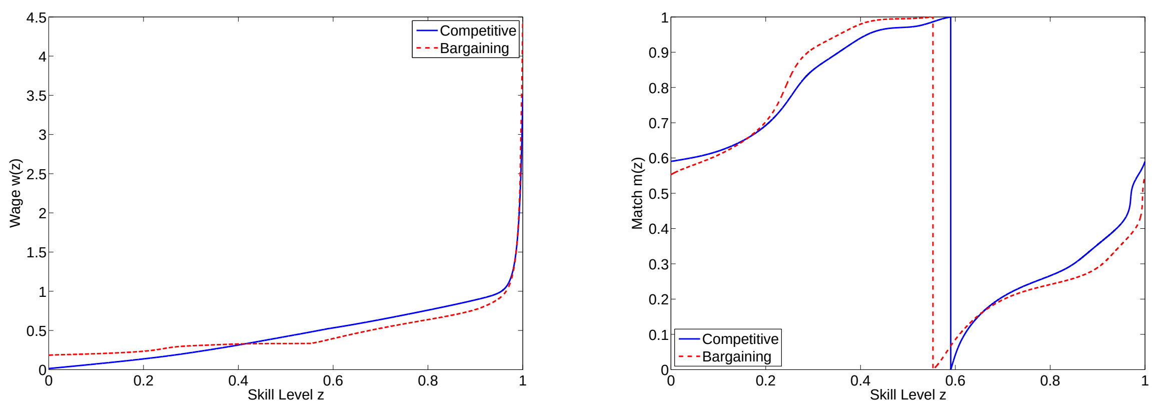
counterfactual efficient competitive allocation. As demonstrated by the proof in appendix A, while an equilibrium exists for any h with wage bargaining, a sufficiently high value of h at z^* can cause the competitive equilibrium to break down, by making the slope of the wage function higher below z^* than above, and this is true in the case of the parameters from the bargained calibration. However, in this case, I can still compute what I call the restricted competitive allocation, which is the equilibrium if individuals are forced to match in a perfect positive assortative allocation. This allocation remains the efficient allocation with wage bargaining: recall that the only possible equilibrium with wage bargaining is a perfect positive assortative allocation - the proof in appendix A.2 is unaffected by taxation - and all individuals are paid their marginal product in the restricted competitive allocation.[30] Figure 4 shows that, once again, wages are too high in the bargaining equilibrium for low-skill workers and the most high-skilled managers, and too low in the middle. In this case, there are also noticeably too few managers in the bargaining case, because many managers face inefficiently low returns and work too little.

In such a setting, rather than being distortionary, marginal taxes can serve an efficiency purpose: if a tax schedule can be chosen that sets each individual's after-tax wage to the efficient value, then individuals will all choose the efficient labour supply and the first-best

[30]The constraint on h ensuring existence and uniqueness of equilibrium in the competitive case is not binding with wage bargaining, but I restrict the maximum value of h to be 1.1 in the bargaining calibration, as this ensures that a restricted competitive allocation is feasible. I have further confirmed using the method of calculating optimal taxes in the following section that the restricted competitive allocation is efficient with wage bargaining - that is, that it does maximize quasi-linear social welfare with wage bargaining.

Figure 4: Wage and Matching Functions with Bargaining Calibration

(a) Wage Function $w(z)$ (b) Matching Function $m(z)$



will be attained. Therefore, in this section, I introduce marginal income taxes, to examine their quantitative role in maximizing efficiency in the current model.

Because the utility function exhibits zero income effects, each individual's labour supply depends only on their after-tax wage: $L_p(z) = [(1-t(z))w(z)]^{\frac{1}{\gamma-1}}$ and $L_m(z) = [(1-t(z))r(z)]^{\frac{1}{\gamma-1}}$, where $t(z)$ is the marginal tax rate faced by an individual with skill z. As a result, the matching function with wage bargaining is defined by:

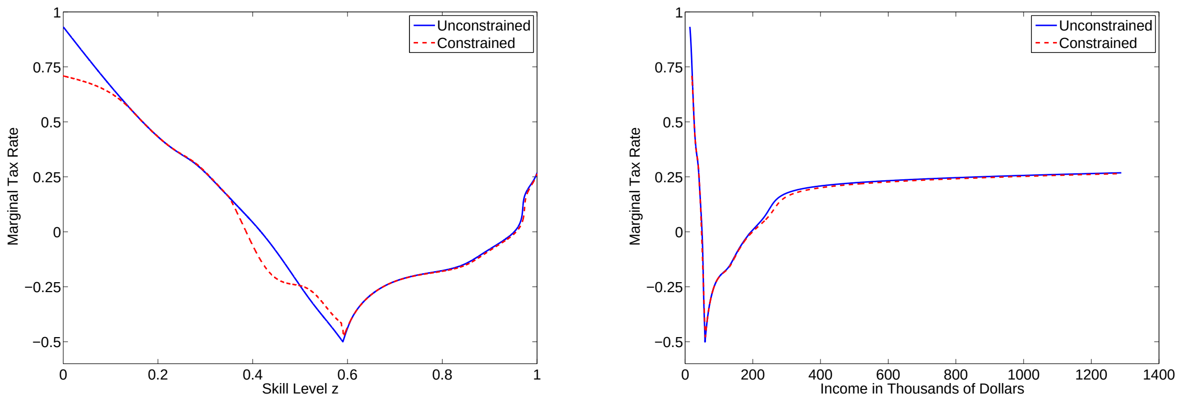
$$m'(z) = \left[\frac{\beta(h(m(z))(1-z))^{\gamma}}{1-\beta}\left(\frac{1-t(z)}{1-t(m(z))}\right)\right]^{\frac{1}{\gamma-1}}\frac{f(z)}{f(m(z))}. \tag{12}$$

I first solve for the marginal tax rates applying to each individual which would restore efficiency to the labour market; that is, the tax as a function of skill level that would cause the wage bargaining allocation to be identical to the competitive one. For the moment, I ignore the question of incentive-compatibility, and assume that the government is able to tax individuals of different skill levels directly. I use a simple procedure in which I iterate between choosing the marginal taxes that match labour supply to the competitive value at each point along the skill distribution, and re-solving for equilibrium at the new taxes. The results are displayed in Figure 5 as the blue solid lines in both panels (a) and (b).

Panel (a) presents efficient taxes as a function of skill z, while panel (b) displays the tax schedule with income on the x-axis. The efficient taxes deviate from zero by a large amount, with large positive taxes at the bottom of the distribution, negative marginal rates in the middle, and rising tax rates at the top that reach more than 25%. This non-monotonicity was predicted in section 2.4, and follows immediately from the pattern of distortions to wages

Figure 5: Efficient Taxes with Wage Bargaining and Calibrated Model

(a) As Function of Skill (b) As Function of Income



illustrated in Figure 4: wages rise slowly with skill, requiring a marginal tax schedule that is declining for workers but progressive for managers. If the manager of a team is overpaid, the workers are underpaid, so the pattern of efficient taxes for workers will be mirrored for managers: if taxes on workers need to be positive at the bottom and negative at the top to offset inefficiently flat wages, the opposite must be true for the managers. Appendix B demonstrates the generality of this result by proving that, at least in a simplified version of the current model, my simplified Nash bargain necessarily makes wages too flat relative to the efficient allocation.

The blue lines in Figure 5 show the first-best efficient tax schedule if individualized marginal tax rates could be assigned to each individual. In practice, this is generally regarded as infeasible: the standard assumption in the optimal taxation literature is that the government cannot observe skill levels of individuals, meaning that taxes can only be levied based on income. Therefore, the next question is: how close to efficiency can we come with only a non-linear income tax? Analogous to the standard Mirrleesian analysis, an income tax is equivalent to assigning marginal tax rates to particular skill levels conditional on two constraints: an identification constraint requires that income increases with skill so that the government can identify skill levels from observed income and impose the tax, and an incentive-compatibility constraint requires that when presented with the tax schedule as a function of income, each individual must prefer the labour supply and thus the income that they would have chosen if faced only with the flat marginal rate assigned to them under the skill tax.[31] These constraints are presented algebraically in appendix D.1.

However, unlike in the examples in section 2, when I check to see if the efficient skill tax satisfies the constraints, I find that it fails the incentive-compatibility constraint over a portion of the distribution; tax rates decline too fast at certain points of lower and middle incomes, encouraging excessive labour supply. I therefore need to adjust the tax schedule to fit inside the incentive-compatibility constraint. Given that I have a well-defined first-best to aim for, I use a simple procedure to find the second-best, efficiency-maximizing tax within an intuitive class of tax schedules: on every iteration I solve for the tax at each skill level that returns labour supply to the value in the competitive allocation; then, I define a value $\hat{z}$, and to ensure that the constraints are satisfied, I hold $t(\hat{z})$ fixed at the value that sets that individual's labour supply to the first-best value, and move left and right from there, adjusting the tax to fit inside the boundaries imposed by the constraints. Therefore, each $\hat{z}$ defines a central fixed point in the adjustment of the efficient tax to satisfy the constraints, and each $\hat{z}$ thus defines a candidate tax schedule. I then search over $\hat{z}$ to find the value and accompanying tax schedule that maximizes quasi-linear welfare; further detail is provided in appendix D.1. The main alternative would have been to use the perturbation method presented in section 5.1 to find the optimal tax schedule with $U_C = 1$;[32] however, the current method is much simpler computationally.[33]

The resulting second-best tax schedule is displayed as the red dashed lines in Figure 5. As can be seen, the resulting tax schedule looks very similar to the first-best skill tax, with minor differences concentrated at lower and middle incomes. With quasi-linear utility, the welfare gain from moving from a 27.9% flat tax to this second-best efficient income tax is equivalent to a substantial 3.60% of mean consumption, or nearly 99% of the efficiency gains which would have been attained had the first-best tax been feasible.

This analysis confirms that non-zero marginal income taxes, for both workers and man-

[31]The incentive-compatibility constraint is analogous to that in the standard Mirrleesian analysis, in which it must be the case that no individual wishes to "imitate" another worker and deviate from their prescribed income. In such a setting, income increasing with skill is a necessary condition for optimal taxation; see Mirrlees (1971).

[32]Using equation (13), it is easy to see that $\frac{dW}{dt_i} = 0$ at the competitive laissez-faire: the gain from redistribution is zero with quasi-linear utility; $t_i = 0$, so the distortion term is zero; finally, the integral of all $L_q \frac{dw_q}{dt_i}$ must be equal to zero because the competitive laissez-faire maximizes the integral of $wL - \frac{1}{\gamma}L^{\gamma}$. The latter implies that the integral of $L_q \frac{dw_q}{dt_i} + w_q \frac{dL_q}{dt_i} - L_q^{\gamma-1}\frac{dL_q}{dt_i}$ must equal zero, and the latter two terms cancel out via the individual first-order condition.

[33]Additionally, as described in section 5, the perturbation approach can only lead to an approximation to the optimal tax schedule because of the complexity of the problem.

agers, can improve efficiency. In particular, the efficient tax schedule takes exactly the V shape predicted in section 2.4, and the deviations from zero are quantitatively large. In appendix F, additional numerical analyses demonstrate the robustness of these results: efficient taxes are V-shaped and quantitately large both when the simple uniform parameterization is used, and when the model is calibrated to a much higher elasticity of taxable income of one. In fact, in both cases, the efficient top tax rate is even higher than in the current analysis. I therefore conclude that the efficiency role of marginal income taxation in a setting with wage bargaining in hierarchical teams is indeed economically significant.

5 Optimal Taxation with Diminishing Marginal Utility

My analysis to this point has focussed on efficiency, and the taxes needed to reach the efficient allocation. As described earlier, this is because the efficiency impact of taxation is an important and understudied input into the optimal tax problem. Efficiency is also a useful baseline in public debates about income taxation, where the participants may have widely varying tastes for redistribution.

However, it is likely that a policy-maker will also care about distribution, and thus will be interested in the optimal tax schedule when the utility function $U(C, L)$ exhibits diminishing marginal utility from income. Therefore, in this final section of the paper, I consider a utilitarian planner searching for the optimal non-linear tax schedule, and present a method for approximating the optimal tax schedule. I will show that the insights from the previous section are robust to such an analysis: the effect of wage bargaining on optimal taxes is V-shaped, with the optimal marginal tax rate rising at the top and bottom, and falling in the middle of the distribution.

One standard method for solving for an optimal income tax schedule is the Mirrleesian method, in which all constraints are specified and the optimal allocation is characterized subject to these constraints. In the current setting, with matching between workers at different points in the distribution and general equilibrium adjustments, the analysis is more complicated. Rothschild and Scheuer (2014) show how to evaluate the optimal tax schedule in a complex setting with multidimensional skills and externalities, and our model fits their general setting; however, they do not perform a numerical analysis, and in the context of my model, their method is computationally infeasible. Rothschild and Scheuer present their solution in two steps: the "outer problem" consists of finding the optimal vector of aggregate efforts, as this allows the wage function to be solved; then, the "inner problem" solves for the optimal allocation given that vector (and thus given a wage function). In my context, the vector of aggregate efforts is replaced by the matching function $m(z)$, as in both the competitive and bargaining cases knowledge of $m(z)$ is sufficient to allow a solution for the wage function. Given a matching function, I can indeed solve the inner problem for the optimal allocation of labour supplies and utilities (with a constraint requiring labour supply to be consistent with the matching function), which would allow me to back out a tax schedule; however, the outer problem is infinite-dimensional, requiring a search over continuous matching functions. Even in practical computational terms with a skill distribution with 10001 mass points as in this paper, the problem is infeasible without parametric or other ad-hoc assumptions on the matching function.

Therefore, I will instead use the other common method for solving for an optimal income tax schedule: the perturbation method, as in Saez (2001). In this method, I consider a small change to the tax schedule at one point; at the optimum, this must have no first-order impact on welfare. I can calculate the derivative of social welfare with respect to taxes at each point on the distribution, as a function of a set of sufficient statistics. Then, by simulating the model and finding the values of the sufficient statistics generated by the model, I can find the tax schedule that comes closest to setting all the welfare derivatives to zero, subject to the constraints.[34] As will be discussed below, this approach necessarily involves some simplifying assumptions as well, and thus my results can only be approximations, but the necessary assumptions seem more sensible and transparent than those that would be required in the Mirrleesian method. My goal, therefore, is to use the perturbation method to provide the best possible approximation to the optimal tax schedule.

In appendix G, I also present an optimal income tax analysis of a simple 2-type matching model, where the complications discussed above no longer exist and the Mirrleesian method is feasible. The numerical results there are completely consistent with the intuition from the results to come in section 5.2: the optimal marginal tax rate (in the sense of a wedge

[34] A perturbation method is based on small changes to the tax schedule at a particular point on the *income* distribution; however, in a general equilibrium model the income distribution is not necessarily isomorphic to the *skill* distribution. Thus, while I can calculate an expression for the welfare impact of a small change to taxes at a point on the income distribution, I can only practically implement a small change to taxes at a point on the skill distribution, and thus constraints become relevant because tax changes facing a particular skill level may not be feasible or incentive-compatible.

between marginal rate of substitution and marginal return to labour supply) faced by lower-skill workers increases with the difference between their wage and the efficient wage, while the marginal rate faced by higher-skill managers decreases with the same gap. This model also further clarifies the difficulties of using the Mirrleesian method when the matching function is not exogenous to tax rates.

5.1 Optimal Taxation Using Perturbation Method

I can solve for the derivative of social welfare with respect to taxes at each point on the distribution in a very general setting, and thus my results will be applicable not only to the model presented in section 3, but to a far more general class of models. To simplify the algebra I consider a population of Q individual mass points, denoted by $q = \{1, ..., Q\}$, with mass $f(z_q)$ at skill levels $z_q = \{z_1, z_2, ..., z_Q\}$; if z is bounded within $[0, \bar{z}]$, a perfectly continuous case is the limit as $Q \to \infty$ and the gap between mass points goes to zero. I assume that the government chooses a tax schedule $T(y)$ that is piecewise linear, consisting of Q marginal tax rates, one for each individual mass point, where the first applies to income up to and including the lowest-skill individual's income y_1, and each subsequent tax rate t_q applies to the income between y_{q-1} and y_q.[35] Given a tax schedule over income, individual labour supply choices will ensure that income is increasing in skill. In the limit as $Q \to \infty$, this will approach a continuous tax schedule.

Managers' "wages" $C'(L_m)$ will be denoted w just like those of production workers, to ensure that my results will apply generally to a wide variety of models, including traditional models with wages set competitively. Then I consider the effect on individuals across the distribution when the government changes one of the tax rates t_i; all individuals will receive a change in the lump-sum transfer, and a change in wages as the labour market equilibrium adjusts, while individuals at and above i will also pay higher taxes on $\Delta y_i \equiv y_i - y_{i-1}$. I combine these effects in appendix E, and additionally make the standard assumption that there are no income effects, with a utility function of $U = U\left(C - \frac{1}{\gamma}L^{\gamma}\right)$, to arrive at the result summarized by the following proposition.

[35] To be precise, I apply each tax rate t_q to $(y_{q-1} + \epsilon, y_q + \epsilon]$, where ϵ is very small, so that I can evaluate the derivative $\frac{dy_q}{dt_i}$ without having to be concerned about behavioural changes shifting an individual into a different tax bracket.

Proposition 2. *The welfare gain from raising t_i is given by:*

$$\frac{dW}{dt_i} = \Delta y_i Q_i \left[E(U_{Cq}) - E(U_{Cq}|q \geq i)\right] - f(z_i)\frac{y_i}{(\gamma-1)}\frac{t_i}{1-t_i}E(U_{Cq}) + \sum_{q=1}^{Q} f(z_q)L_q\left[\frac{\gamma}{\gamma-1}t_q E(U_{Cq}) + (1-t_q)U_{Cq}\right]\frac{dw_q}{dt_i} \tag{13}$$

where $Q_i \equiv \sum_{q=i}^{Q} f(z_q)$.

Proof. See appendix E. □

This equation can easily be understood as the sum of three effects. The first term in (13) is the redistribution effect: the total tax revenues collected are multiplied by the marginal welfare gain from taxing high incomes and redistributing to everyone through a lump-sum transfer. The second term is the distortionary effect of the tax, the lost tax revenues from the reduced labour supply of individual i. These first two terms represent the standard tradeoff in optimal taxation between redistribution and distortionary effects. However, the final term is a new component, a wage-shifting effect: the effect of the tax t_i on wages is valued both for its redistribution effect, where it is weighted by each U_{Cq}, and for its efficiency effect, where multiplied by $E(U_{Cq})$. This term provides an alternative way of thinking about the distortion-offsetting effects of taxation: if a particular individual's wage is too high, then taxing them will tend to increase average wages by shifting the matching function in an efficiency-enhancing direction.[36] In particular, notice that if all taxes are initially zero and utility is quasi-linear, the final term simplifies to $\sum_{q=1}^{Q} f(z_q)L_q\frac{dw_q}{dt_i}$, which is simply the effect of any wage changes on total output; effects through labour supply cancel out of utility via the envelope theorem, except for the effect for individual i which shows up in the distortionary term. Increases in total output that occur holding labour supplies fixed clearly represent increases in efficiency. It is also easy to generalize (13) to a continuous distribution, as demonstrated in appendix E.

If I write $R_i = \Delta y_i Q_i \left[E(U_{Cq}) - E(U_{Cq}|q \geq i)\right]$ for the redistribution term and $S_i = \sum_{q=1}^{Q} f(z_q)L_q\left[\frac{\gamma}{\gamma-1}t_q E(U_{Cq}) + (1-t_q)U_{Cq}\right]\frac{dw_q}{dt_i}$ for the wage-shifting effect, at the optimal

[36] A simple thought experiment shows why taxes must shift wages if they are not equal to marginal product, even in a fixed-team-size setting: a manager's "wage" is the sum of their actual contribution to society plus the rents they collect from workers divided by their labour supply. If a tax is imposed on the workers, they will work less and thus provide fewer rents to the manager, changing the hourly return the latter receives. Meanwhile, in a variable-team-size model, the manager can hire additional workers to achieve the same overall worker labour input, but reallocation of workers across teams will change who works for whom, thus changing the rents each manager obtains.

tax rate t_i it must be true that:

$$R_i + S_i = f(z_i)\frac{y_i}{\gamma - 1}\frac{t_i}{1 - t_i}E(U_{Cq})$$

and rearranging, this gives:

$$t_i = \frac{(\gamma - 1)(R_i + S_i)}{f(z_i)y_i E(U_{Cq}) + (\gamma - 1)(R_i + S_i)}. \tag{14}$$

Equations (13) and (14) look like a new set of "sufficient statistics" conditions for welfare analysis of taxation, a generalization of the results in Saez (2001) and Diamond and Saez (2011). Note that, if I make the same assumptions as in the analysis of the optimal top tax rate in Diamond and Saez (2011), which are $S_i = 0$, $Q_i = f(z_i)$, $E(U_{Cq}|q \geq i) = 0$, and the Pareto tail assumption which implies that $y_i = \frac{\alpha y_{i-1}}{\alpha - 1}$, where α is the Pareto parameter, my expression simplifies exactly to theirs: $t_i = \frac{1}{1+e\alpha}$ for the top tax rate, where $e \equiv \frac{1}{\gamma - 1}$ is the elasticity of taxable income. When S_i is not equal to zero, there is an additional term on the top and bottom, and the optimal tax rate t_i is increasing in S_i: a more positive effect on average wages (weighted by labour supply and marginal utilities as stated in the formula for S_i) raises the optimal tax rate.

However, while (13) and (14) can be applied in a wide variety of situations, beyond the team production setting studied in this paper, their practical applicability is limited by the fact that they require us to be able to measure not only marginal utilities, but also individual wages w_q and changes in those wages with taxation. Observation of wages is generally ruled out in analyses of optimal taxation; in the usual competitive labour market setting, wages are equivalent to skill levels, and thus observation of equilibrium wages would make redistributional lump-sum taxes feasible. However, (13) and (14) can be used with any specific model, regardless of the wage-setting mechanism; by simulating the model, we can calculate the sufficient statistics and plug them into (13) to obtain the effect of changing t_i on social welfare. Therefore, they can be used with the parametric model from section 3.

5.2 Approximate Optimal Tax Schedules

I can now present the optimal income tax schedules with diminishing marginal utility of income. Specifically, I assume $\theta = 1$, which implies $U(C, L) = \ln\left(C - \frac{1}{\gamma}L^\gamma\right)$. I will present results using the skill distributions and $h(z)$ from section 3.3, which have been calibrated to the U.S. economy; results with a uniform distribution can be found in appendix F.1.

The government's objective is to maximize average utility, by choosing a non-linear continuous income tax $T(y)$ to finance lump-sum transfers and a quantity of required government spending denoted by G; the value used for G is \$12207, as this is the amount that balances the government budget given a baseline marginal tax of 27.9% and a minimum income of \$4024.90. Therefore, the government's problem is:

$$\max_{T(y)} W = \int_0^1 U(C(z), L(z)) f(z) dz \;\; s.t. \int_0^1 T(y(z)) f(z) dz = G$$

where I allow $T(0)$ to represent any lump-sum transfer or tax.

As discussed before, labour supply depends only on the after-tax wage, and the matching function is altered to account for this, with (12) giving the matching function in the wage bargaining case, and the following equation for the competitive case:

$$m'(z) = \left[\frac{(h(1-z))^{\gamma} w(z)}{m(z) - w(z)} \left(\frac{1 - t(y(z))}{1 - t(y(m(z)))} \right) \right]^{\frac{1}{\gamma - 1}} \frac{f(z)}{f(m(z))}.$$

To solve for optimal taxes in this setting, I need to evaluate the welfare impact of changing taxes at each point in the distribution, using equations (13) and (14). I use an iterative procedure in which I start at a baseline tax schedule and use (14) and my model to gradually approach the optimal schedule.[37] At the baseline tax schedule, I solve my model for the values of R and S across the distribution, and input these into (14) to find the implied optimal tax rate t_i^* for each individual (where I continue to use a population of 10001 individual mass points at skill levels $\{0, 0.0001, ..., 1\}$). Then I calculate the upper and lower bounds $\overline{t_i}$ and $\underline{t_i}$ for each individual that ensure that the identification and incentive-compatibility constraints are satisfied; finally, I assign each individual their constrained optimal tax rate $t_i^{**} = \min\{\overline{t_i}, \max\{\underline{t_i}, t_i^*\}\}$. In practice, the constraints are often binding, so the tax schedule adjusts slowly towards the optimum; on the next iteration, the new tax schedule is treated as the baseline, and the process is repeated. The iterative procedure is described in further detail in appendix D.2.

This procedure can be path-dependent, as each shift in the tax schedule at one point affects outcomes and thus estimated optimal tax rates across the rest of the distribution;

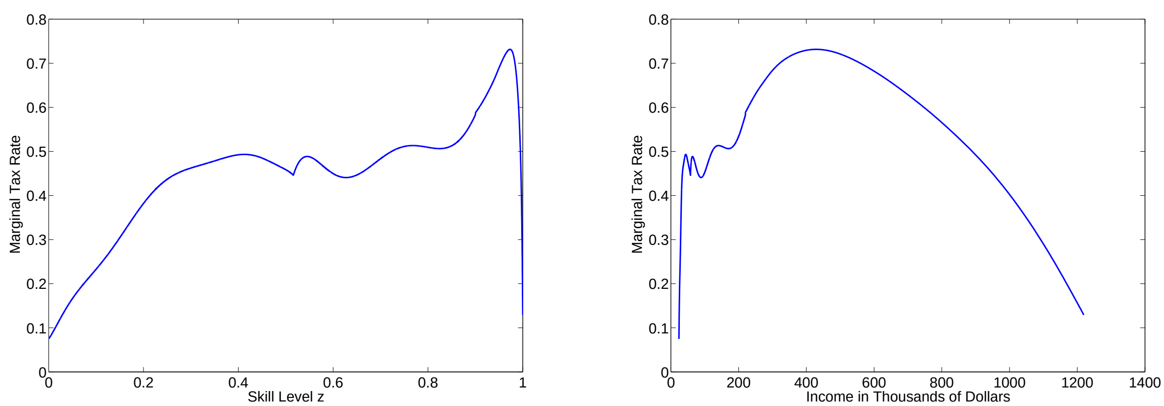
[37] An alternative is to parameterize the tax schedule, as a cubic spline for instance, and perform a grid search for the optimum. However, this approach is computationally time-consuming, and I encountered severe difficulties in ensuring convergence: the tax can either be defined over income, in which case the model must be solved iteratively for the income distribution and marginal tax rates faced by each individual, or defined over skill and then translated into an equivalent income tax. The former is very difficult in its iterative solution, whereas the latter presents the difficulty of trying to search for a global optimum over a high-dimensional space defined by the constraints.

therefore, adjusting a tax rate to satisfy a constraint at one point on the distribution can shift the constraints and the estimated optimal tax at another point. To ensure the robustness of my results, in each case I start from 4 different baseline tax schedules: flat marginal taxes of zero and of 27.9%, a concave quadratic tax schedule given by $t = z - z^2$, and the convex schedule given by 0.25 minus the concave schedule. I solve for the best approximation to the optimal tax schedule from each starting schedule, and select the one that provides the highest average utility, although all four are generally similar. This approach provides the best approximation to the optimal tax schedule that I am capable of, given the complexity of the matching model. As described earlier, appendix G solves for optimal taxes using a version of the Mirrleesian method in a simple 2-type model; results are qualitatively similar to those in this section, but the same method is not computationally feasible with the full general equilibrium matching model.

The competitive results are presented in Figure 6;[38] the optimal marginal tax rates roughly follow an inverted-U shape. Moving from the baseline flat tax to the optimum produces welfare gains equivalent to a 4.04% increase in each individual's consumption, due to gains from redistribution.[39]

Figure 6: Optimal Tax Schedule with Log Utility and Competitive Wages

(a) As Function of Skill (b) As Function of Income



The reason for the inverted-U shape of the optimal tax schedule is simple: it is primarily

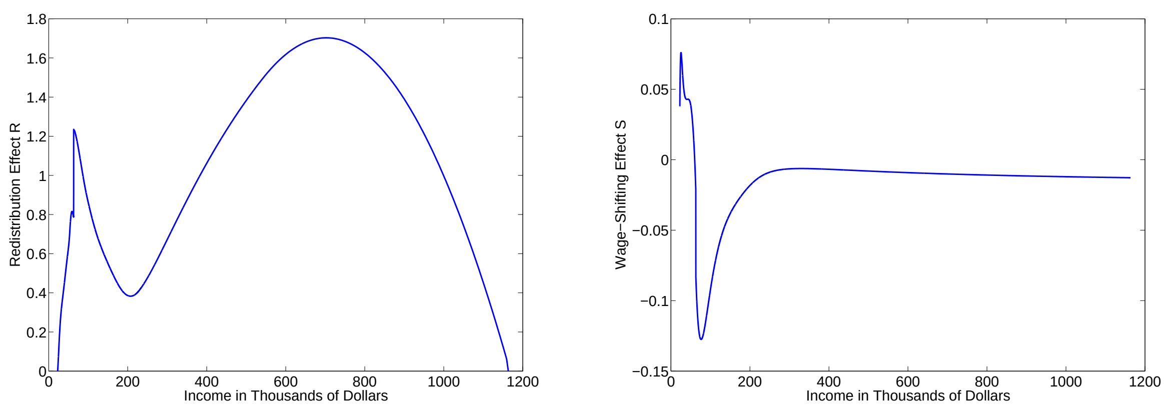
[38]In this figure, and in all optimal tax figures in the paper, the cutoff between workers and managers is at the visible point of non-differentiability of the tax schedule.

[39]That is, if W_1 is welfare at the optimum and $\{C_{10}, C_{20}, ...\}$ is the vector of baseline consumption for all individuals, then holding each individual's labour supply constant, I find the value r that sets $W_1 = W(C_{10}(1+r), C_{20}(1+r), ...)$.

driven by a redistribution effect that is generally in the shape of an inverted-U itself, as can be seen in Figure 7, which displays the values of R and S at baseline taxes. The gains from redistribution are zero at the top and bottom of the income distribution, because a marginal tax at the top raises no revenue and a tax at the bottom cannot perform any redistribution; however, gains from redistribution are positive in between. There is a sharp spike upwards in the gains from redistribution at the cutoff skill level z^*, because the income distribution becomes thinner at that point, but this is largely offset by the wage-shifting effect, which is positive at low incomes but drops abruptly to a large negative value above z^*. The latter occurs because a positive tax at any point in the distribution reduces labour supply at that point, shifting the matching function accordingly; thus, a tax on workers below z^* has beneficial effects on welfare because it shifts workers to higher-skill managers and increases their wages, while a tax on managers has the opposite effect.

Figure 7: Values of R and S for Competitive Wages and Baseline Taxes

(a) Value of R (b) Value of S

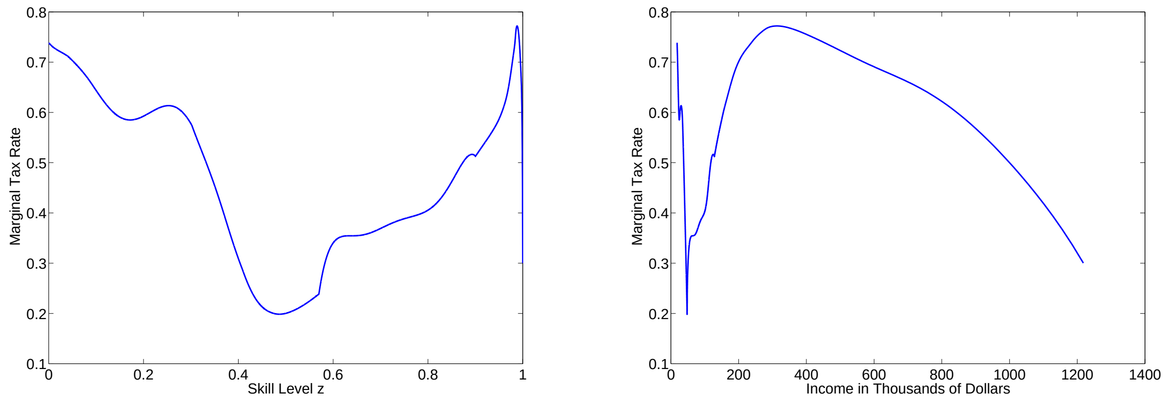


The results with wage bargaining can be found in Figure 8. The results are now a cross between the efficient tax with wage bargaining found in Figure 5 and the competitive results with log utility above: for the thick part of the income distribution, at low-to-moderate incomes, optimal taxes are still generally V-shaped, but after rising to nearly 80% at a fairly high income, the optimal marginal rate declines to about 30% at the top. The resulting welfare gains from moving to the optimal tax are equivalent to a 4.62% increase in consumption.

The reason for this roughly S-shaped result can be found in the forms of R and S displayed in Figure 9. The gains from direct redistribution are positive but small for workers

Figure 8: Optimal Tax Schedule with Log Utility and Wage Bargaining

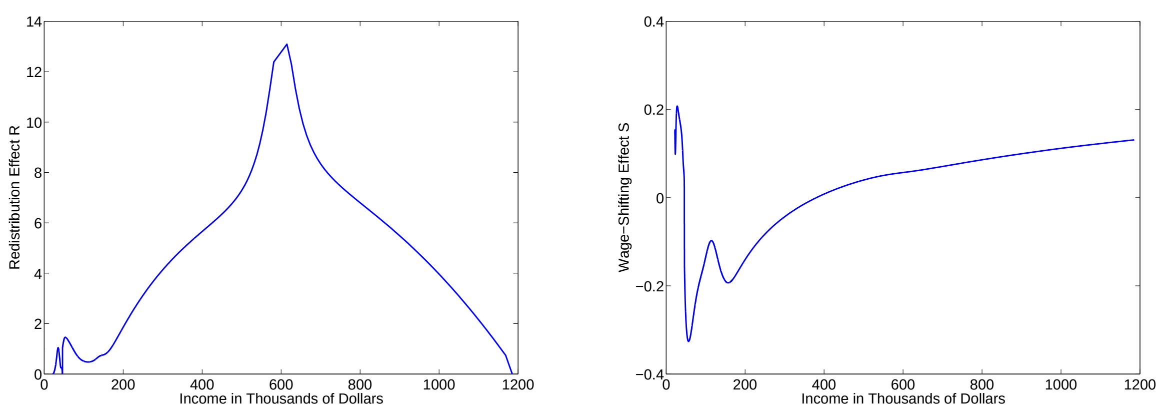
(a) As Function of Skill (b) As Function of Income



below z^* (who occupy a very small space on the income distribution), but large and generally inverted-U-shaped for managers, justifying high taxes at relatively high incomes but declining rates at the very top. Meanwhile, the wage-shifting effect takes a U-shape over most of the distribution, and this explains why marginal taxes do not go to zero at the top: high taxes at the very top of the distribution, by offsetting the bargaining power held by those highest-skill managers, improve efficiency.

Figure 9: Values of R and S for Wage Bargaining and Baseline Taxes

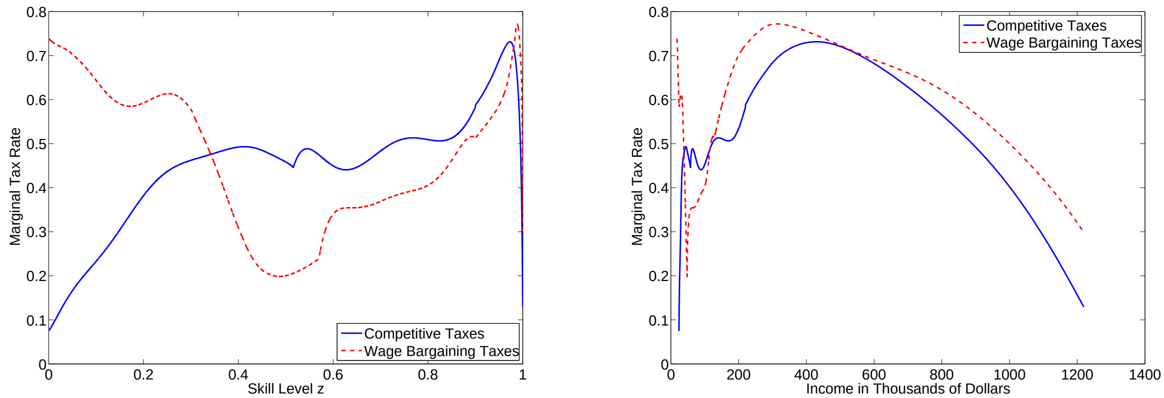
(a) Value of R (b) Value of S



To highlight the effect of wage bargaining on the optimal tax schedule, the two optimal schedules can be presented in the same figure, specifically in Figure 10. Given the findings of section 4 on efficient taxes, the results are exactly as one would expect: the optimal tax schedule in a setting of wage bargaining features considerably larger tax rates at the bottom of the distribution, lower in the middle, and higher at the top.

Figure 10: Optimal Tax Schedules with Log Utility

(a) As Function of Skill (b) As Function of Income



The results in this section demonstrate that the effect of wage bargaining on optimal income taxes is substantial. In particular, the results on efficient taxes in the previous sections of the paper apply robustly here as well: because wage bargaining tends to flatten wages for workers and increase the slope of returns for managers, it introduces a motive for taxes that are regressive among workers and progressive among managers. This is true regardless of whether the planner has a desire for redistribution.

To further confirm the robustness of my results, appendix F presents an additional set of numerical results. First of all, appendix F.1 presents results with a uniform skill distribution and constant h, and finds very similar results, with even higher efficient and optimal tax rates at the top in the wage bargaining case.

My analysis to this point has used an elasticity of taxable income of 0.25, which is a standard estimate from the empirical literature. However, earlier studies often found considerably higher elasticities; Feldstein (1995), for example, finds a value of at least 1, and Weber (2014) argues that most previous estimators of this elasticity are inconsistent due to mean reversion of income, finding an elasticity of 0.858 with a new estimator.[40] Therefore, in appendix F.2, I redo the analysis with a considerably higher elasticity of taxable income of

[40] Additionally, recent research has indicated that macro-level long-run elasticities of labour supply could be larger than the standard micro-level estimates; Keane and Rogerson (2012) find that micro estimates could understate the true preference parameter by a factor of about two.

1, and I find that my conclusions are largely unaltered: efficient taxes may actually be larger at the top end, and while the optimal competitive taxes drop considerably, the effect on optimal taxes with wage bargaining is relatively modest, leaving the qualitative conclusions about the impact of wage bargaining unaffected.

One notable feature of the results presented above is the strong decline of the optimal marginal tax rate above about $400 thousand. This is primarily due to the fact that there is a finite top to the income distribution, a necessary component of my model given that $z = 1$ means that an individual can solve all potential production problems. A finite top to the income distribution eliminates the redistributional role of marginal taxes at the top of the distribution, because a marginal tax at the top raises no revenue. In appendix F.3, I perform a robustness check in which I assume that there is a Pareto tail of individuals at very high incomes outside my model; in this case, there is still a strong redistributive role for marginal taxes at the top of my skill distribution, and the optimal tax rates remain around 70% near the top of the distribution. However, the tax schedules are otherwise very similar to those in Figure 10, and the effect of wage bargaining on the optimal tax schedule remains V-shaped.

Finally, in appendix F.4, I present optimal bracketed taxes, with thresholds set at approximately the levels facing a single taxpayer in the U.S. both in the baseline calibrated setting and in the alternative calibration with an elasticity of taxable income of 1. While the results are coarser, the general results of this section are confirmed, as wage bargaining causes the optimal tax rates to rise in the bottom and top brackets, and drop in the middle.

6 Conclusion

In this paper, I have studied the efficiency role of taxation in a context of wage bargaining within teams. Using a general model of two-layer team production, I show that non-zero marginal taxes on high-skill managers can only be justified from an efficiency perspective if team size increases in manager effort; in other words, a "job-creation" effect is required, in which high-skilled managers exert effort in trying to accumulate workers and the rents that come with them. I show that efficient marginal taxes are likely to be decreasing for workers and increasing for managers.

I then turn to a specific parametric model of team production in general equilibrium.

Using the latter, I find that a highly right-skewed income distribution can be generated without a skewed skill distribution when rents from workers are captured by high-skill team managers. I demonstrate that in this setting, the deviation between wages and marginal product is V-shaped with wage bargaining, and thus the efficient marginal tax schedule is also V-shaped with a top rate of at least 25%. Marginal taxes that deviate significantly from zero can play an important role in improving efficiency. Finally, I apply an optimal income tax analysis to the model, and show that wage bargaining dramatically alters the optimal tax schedule to feature reductions in tax rates at middle incomes and increased tax rates near the top and bottom of the distribution.

Given the small number of papers which attempt to address issues of the use of income taxes to offset labour market distortions, I believe this subject holds the promise of numerous important future contributions to our understanding of the welfare consequences of income tax policy. My analysis indicates that there are few general results when it comes to efficient or optimal taxation in labour markets affected by wage bargaining: specific parametric models are required, and so future work could include testing alternative models to further our understanding of the relationship between wages and marginal product. Additionally, previous research, both theoretical and empirical, has commonly assumed that top tax rates have no effect on individuals further down in the income distribution. My analysis shows that this may be incorrect, which implies that groups that are not directly affected by tax changes may not be good control groups when estimating elasticities of taxable income; future empirical work on this question would fill an important gap.

References

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

Ales, L., M. Kurnaz, and C. Sleet (2014): "Technical Change, Wage Inequality and Taxes," *American Economic Review*, forthcoming.

Andrews, M., L. Gill, T. Schank, and R. Upward (2008): "High Wage Workers and Low Wage Firms: Negative Assortative Matching or Limited Mobility Bias?," *Journal of the Royal Statistical Society*, 171(3), 673–697.

——— (2012): "High Wage Workers Match With High Wage Firms: Clear Evidence of the Effects of Limited Mobility Bias," *Economics Letters*, 117(3), 824–827.

Antràs, P., L. Garicano, and E. Rossi-Hansberg (2005): "Offshoring in a Knowledge Economy," Working Paper no. 11094, NBER, Cambridge, MA.

——— (2006): "Offshoring in a Knowledge Economy," *Quarterly Journal of Economics*, 121(1), 31–77.

Blomquist, S., V. Christiansen, and L. Micheletto (2010): "Public Provision of Private Goods and Nondistortionary Marginal Tax Rates," *American Economic Journal: Economic Policy*, 2(2), 1–27.

Boadway, R., and M. Sato (2014): "Optimal Income Taxation and Risk: The Extensive-Margin Case," *Annals of Economics and Statistics*, 113-114, 159–184.

Boone, J., and L. Bovenberg (2002): "Optimal Labour Taxation and Search," *Journal of Public Economics*, 85(1), 53–97.

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.

Chetty, R. (2006): "A General Formula for the Optimal Level of Social Insurance," *Journal of Public Economics*, 90(10-11), 1879–1901.

Diamond, P., and E. Saez (2011): "The Case for a Progressive Tax: From Basic Research to Policy Recommendations," *Journal of Economic Perspectives*, 25(4), 165–190.

Diamond, P. A. (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., and P. Kircher (2011): "Identifying Sorting – In Theory," *Review of Economic Studies*, 78(3), 872–906.

Feldstein, M. (1995): "The Effect of Marginal Tax Rates on Taxable Income: A Panel Study of the 1986 Tax Reform Act," *Journal of Political Economy*, 103(3), 551–572.

Gruber, J., and E. Saez (2002): "The Elasticity of Taxable Income: Evidence and Implications," *Journal of Public Economics*, 84(1), 1–32.

Hope, O.-K., and W. B. Thomas (2008): "Managerial Empire Building and Firm Disclosure," *Journal of Accounting Research*, 46(3), 591–626.

Hosios, A. J. (1990): "On the Efficiency of Matching and Related Models of Search and Unemployment," *Review of Economic Studies*, 57(2), 279–298.

Hungerbühler, M., E. Lehmann, A. Parmentier, and B. van der Linden (2006): "Optimal Redistributive Taxation in a Search Equilibrium Model," *Review of Economic Studies*, 73(3), 743–767.

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.

——— (2014): "Optimal Income Taxation with Kalai Wage Bargaining and Endogenous Participation," *Social Choice and Welfare*, 42(2), 381–402.

Jensen, M. C. (1986): "Agency Costs of Free Cash Flow, Corporate Finance, and Takeovers," *American Economic Review*, 76(2), 323–329.

Keane, M., and R. Rogerson (2012): "Micro and Macro Labor Supply Elasticities: A Reassessment of Conventional Wisdom," *Journal of Economic Literature*, 50(2), 464–476.

Krugman, P. (November 22, 2011): "Taxing Job Creators," http://krugman.blogs.nytimes.com/2011/11/22/taxing-job-creators/, New York Times.

Lehmann, E., A. Parmentier, and B. van der Linden (2011): "Optimal Income Taxation with Endogenous Participation and Search Unemployment," *Journal of Public Economics*, 95(11-12), 1523–1537.

Lockwood, B. B., C. G. Nathanson, and E. G. Weyl (2014): "Taxation and the Allocation of Talent," Unpublished Paper.

Mankiw, N. G., M. Weinzierl, and D. Yagan (2009): "Optimal Taxation in Theory and Practice," *Journal of Economic Perspectives*, 23(4), 147–174.

Manning, A. (2003): *Monopsony in Motion: Imperfect Competition in Labor Markets*. Princeton University Press, Princeton, NJ.

——— (2011): "Imperfect Competition in the Labor Market," in *Handbook of Labor Economics*, ed. by D. Card, and O. Ashenfelter, vol. 4b, Elsevier, pp. 973–1041.

Mirrlees, J. (1971): "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, 38(2), 175–208.

Moresi, S. (1998): "Optimal Taxation and Firm Formation: A Model of Asymmetric Information," *European Economic Review*, 42(8), 1525–1551.

Persson, M., and A. Sandmo (2005): "Taxation and Tournaments," *Journal of Public Economic Theory*, 7(4), 543–559.

Piketty, T., E. Saez, and S. Stantcheva (2014): "Optimal Taxation of Top Labor Incomes: A Tale of Three Elasticities," *American Economic Journal: Economic Policy*, 6(1), 230–271.

Robin, J.-M., and S. Roux (2002): "An Equilibrium Model of the Labour Market with Endogenous Capital and Two-Sided Search," *Annales d'Economie et de Statistique*, 67/68, 257–307.

Rothschild, C., and F. Scheuer (2011): "Optimal Taxation with Rent-Seeking," Working Paper no. 17035, NBER, Cambridge, MA.

——— (2013): "Redistributive Taxation in the Roy Model," *Quarterly Journal of Economics*, 128(2), 623–668.

——— (2014): "A Theory of Income Taxation Under Multidimensional Skill Heterogeneity," Working Paper no. 19822, NBER, Cambridge, MA.

Saez, E. (2001): "Using Elasticities to Derive Optimal Income Tax Rates," *Review of Economic Studies*, 68(1), 205–229.

Saez, E., J. Slemrod, and S. H. Giertz (2012): "The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review," *Journal of Economic Literature*, 50(1), 3–50.

Sandmo, A. (1998): "Redistribution and the Marginal Cost of Public Funds," *Journal of Public Economics*, 70(3), 365–382.

Scheuer, F. (2013): "Adverse Selection in Credit Markets and Regressive Profit Taxation," *Journal of Economic Theory*, 148(4), 1333–1360.

——— (2014): "Entrepreneurial Taxation with Endogenous Entry," *American Economic Journal: Economic Policy*, 6(2), 126–163.

Spanos, G. (2013): "Sorting Within and Across French Production Hierarchies," Unpublished Paper, University of Toronto.

Stantcheva, S. (2014): "Optimal Income Taxation with Adverse Selection in the Labor Market," *Review of Economic Studies*, 81(3), 1296–1329.

Stiglitz, J. (1982): "Self-Selection and Pareto Efficient Taxation," *Journal of Public Economics*, 17(2), 213–240.

Varian, H. (1980): "Redistributive Taxation as Social Insurance," *Journal of Public Economics*, 14(1), 49–68.

Weber, C. E. (2014): "Toward Obtaining a Consistent Estimate of the Elasticity of Taxable Income Using Difference-in-Differences," *Journal of Public Economics*, 117, 90–103.

A Proof of Perfect Positive Assortative Matching

In this appendix, I present the proof that any equilibrium of my model from section 3 must feature perfect positive assortative matching, which I define as a one-to-one mapping between worker and manager skill with a single cutoff skill level z^* between lower-skill workers and higher-skill managers, with manager skill monotonically increasing in worker skill. I also show that such an equilibrium must exist in the wage bargaining case, and exists in the competitive case as long as communication costs are not too high. I will start with competitive wage-setting, followed by the proof with wage bargaining.

A.1 Proof with Competitive Labour Market

In the competitive labour market, much of the proof is identical or very similar to that in Antràs, Garicano, and Rossi-Hansberg (2005); thus, I present my deviations from that proof here, and for any sections of the proof that are completely identical to Antràs, Garicano, and Rossi-Hansberg (2005), I will simply refer to the relevant section of the proof in that paper.

I first need to prove that the mapping between worker skill z_p and manager skill z_m will be one-to-one, so that I can use a matching function. This proof is presented in appendix B of Antràs, Garicano, and Rossi-Hansberg (2005), where the first step in Lemma B.1 is identical to mine: a particular manager can not be matched with an interval of workers, since each manager is infinitesimally small. Then I need to prove that a particular worker type cannot be matched with several different manager types; the analysis proceeds as on page 46 of Antràs, Garicano, and Rossi-Hansberg (2005). The problem of a manager of type z_m who hires I different types of workers is:

$$U = \max_{n,z} \sum_{i=1}^{I} n_i L_i(z_m - w(z_i)) - \frac{1}{\gamma} L_m^{\gamma} + \lambda[L_m - h(z_m) \sum_{i=1}^{I} n_i L_i(1 - z_i)]$$

and therefore the first-order conditions for each i are:

$$L_i(z_m - w(z_i)) - \lambda h(z_m) L_i(1 - z_i) = 0$$
$$-n_i L_i w'(z_i) + \lambda h(z_m) n_i L_i = 0.$$

Combining these expressions, I find that:

$$w'(z_i) = \frac{z_m - w(z_i)}{1 - z_i}$$

as in Antràs, Garicano, and Rossi-Hansberg (2005). And as in Lemma B.2 of the latter paper, this implies that a worker type z_p matched with two different manager types z_m and $z'_m > z_m$ must satisfy:

$$w'(z_p) = \frac{z_m - w(z_p)}{1 - z_p} = \frac{z'_m - w(z_p)}{1 - z_p}$$

but $z'_m > z_m$, which contradicts the final equality. Therefore, any given worker type can only be matched with one manager type, and a matching correspondance $m(z)$ must exist. As in Lemma B.3 of Antràs, Garicano, and Rossi-Hansberg (2005), this correspondance must be continuous, since maximization takes places on a continuous function over a compact set.

To complete the proof that $m(z)$ is a one-to-one function, all that is needed is that each manager hires only one worker type. The proof for this is very similar to that for Lemma B.4 in Antràs, Garicano, and Rossi-Hansberg (2005); we know from the first of the first-order conditions above that:

$$z_m = w(z_i)) + \lambda h(z_m)(1 - z_i).$$

Therefore, for any manager z_m that hired two worker types z_1 and $z_2 > z_1$, it must be true that:

$$w(z_2) - w(z_1) = \lambda h(z_m)(z_2 - z_1).$$

Thus, this manager z_m would also be willing to hire any individual $z \in [z_1, z_2]$ if:

$$w(z) = \tilde{w}(z) \equiv w(z_1) + \lambda h(z_m)(z - z_1).$$

Since manager z_m only hires z_1 and z_2, it must be the case that $w(z) > \tilde{w}(z)$ for all $z \in [z_1, z_2]$. This implies that there must be some interval H between z_1 and z_2 for which the slope of the wage function is greater than $\lambda h(z_m)$, and some interval L for which it is lower. Take a manager z_m^H hiring workers from point z_p^H in interval H, where the slope of the wage function is $s > \lambda h(z_m)$. By definition of the fact that this manager hires workers at this point, they must be willing to hire workers of other abilities z if:

$$w(z) = \tilde{w}^H(z) = w(z_p^H) + s(z - z_p^H)$$

and this implies that $\tilde{w}^H(z_2) > w(z_2)$, and thus manager z_m^H would like to hire workers at z_2 and pay them $w(z_2) + \epsilon$ for a small enough ϵ, contradicting the assumption that manager z_m hires workers at both z_1 and z_2. Therefore, managers will only hire one worker type, and I can use a matching function $z_m = m(z_p)$ to describe the matching equilibrium.

Next, I must prove that $m(z)$ exhibits positive assortative matching; that is, that $m'(z) > 0$. The proof follows the same form as that for Theorem 1 in Antràs, Garicano, and Rossi-Hansberg (2005), but is slightly more complicated. Managers choose their workers optimally to maximize their "wage" $r(z_m; z_p) = \frac{z_m - w(z_p)}{h(z_m)(1-z_p)}$, and so we know that in equilibrium $\frac{\partial r(z_m;z_p)}{\partial z_p} = 0$. Differentiating this expression, I find that:

$$\frac{\partial z_m}{\partial z_p} = -\frac{\frac{\partial^2 r(z_m;z_p)}{\partial z_p^2}}{\frac{\partial^2 r(z_m;z_p)}{\partial z_p \partial z_m}}.$$

The manager's second-order condition tells us that the numerator is negative, so for $m'(z) > 0$, we need the denominator to be positive. This can be shown as follows:

$$\begin{aligned}
\frac{\partial^2 r(z_m; z_p)}{\partial z_p \partial z_m} &= \frac{\partial}{\partial z_p}\left(\frac{h(z_m) - h'(z_m)(z_m - w(z_p))}{h(z_m)^2(1-z_p)}\right) \\
&= \frac{h'(z_m)w'(z_p)(1-z_p) + h(z_m) - h'(z_m)(z_m - w(z_p))}{h(z_m)^2(1-z_p)^2} \\
&= \frac{1}{h(z_m)(1-z_p)^2} > 0.
\end{aligned}$$

Therefore, the matching function is increasing: $m'(z) > 0$.

All that remains to be proved is that the equilibrium consists of a single cutoff z^* above which individuals become managers, and below which individuals become workers. That is, I need to rule out consecutive disjoint sets of workers and managers. Suppose that the latter was true: suppose that individuals from z_1 to z_2 and z_3 to z_4 were workers, while individuals from z_2 to z_3 and z_4 to z_5 were managers, where $z_5 > z_4 > z_3 > z_2 > z_1$. We know that $m(z_1) = z_2$ and $m(z_2) = z_3$, and so forth. For this to be an equilibrium, several necessary conditions must be satisfied. First, defining the wage functions for workers and managers as $w_{13}(z)$ and $r_{13}(z)$ between z_1 and z_3, and $w_{35}(z)$ and $r_{35}(z)$ between z_3 and z_5, I need $r_{13}(z_3) = w_{35}(z_3)$; that is, it must be the case that individuals at z_3 are indifferent between being workers and managers.

I also need $\lim_{z\uparrow z_3} \frac{\partial r_{13}(z)}{\partial z} < \lim_{z\downarrow z_3} \frac{\partial w_{35}(z)}{\partial z}$, as otherwise a manager at z_4 who is currently matched with a worker at z_3 would like to hire someone at $z_3 - \epsilon$ at a better wage than the latter currently makes as a manager. I find that the partial derivative on the left-hand side takes the following form:

$$\frac{\partial r_{13}(z)}{\partial z} = \frac{h(z_m) - h'(z_m)(z_m - w(z_p))}{h(z_m)^2(1-z_p)} \geq \frac{1}{h(z_m)(1-z_p)} > 1$$

where the first greater-than sign follows if $h'(z) \leq 0$, and the second follows if $h < 1$. As in Antràs, Garicano, and Rossi-Hansberg (2005), consider two different cases for the derivative of w_{35}: one in which $w_{13}(z_2) \geq z_2$, and one in which $w_{13}(z_2) < z_2$. In the first case, since $\frac{\partial r_{13}(z)}{\partial z} > 1$, it must be the case that $r_{13}(z_3) = w_{35}(z_3) > z_3$. Then I find that:

$$\frac{\partial w_{35}(z)}{\partial z} = \frac{z_4 - w_{35}(z_3)}{1 - z_3} < 1$$

because $z_4 < 1$ and $w_{35}(z_3) > z_3$. Thus, if $w_{13}(z_2) \geq z_2$, I have $\lim_{z \uparrow z_3} \frac{\partial r_{13}(z)}{\partial z} > 1 > \lim_{z \downarrow z_3} \frac{\partial w_{35}(z)}{\partial z}$, which contradicts the assumption of disjoint sets of workers and managers.

If, on the other hand, $w_{13}(z_2) < z_2$, we know that

$$\lim_{z \downarrow z_3} \frac{\partial w_{35}(z)}{\partial z} = \frac{z_4 - w_{35}(z_3)}{1 - z_3} = \frac{z_4 h(z_3)(1 - z_2) - z_3 + w_{13}(z_2)}{(1 - z_3)h(z_3)(1 - z_2)}$$

and thus, to prove that such an equilibrium cannot exist, I need that:

$$\frac{1}{h(z_3)(1 - z_2)} > \frac{z_4 h(z_3)(1 - z_2) - z_3 + w_{13}(z_2)}{(1 - z_3)h(z_3)(1 - z_2)}$$

which simplifies to:

$$1 > \frac{z_4 h(z_3)(1 - z_2) - z_3 + w_{13}(z_2)}{(1 - z_3)}.$$

The latter follows if $z_4 h(z_3)(1 - z_2) + w_{13}(z_2) < 1$, which can rearranged to form:

$$z_4 < \frac{1 - w_{13}(z_2)}{h(z_3)(1 - z_2)}$$

and since $w_{13}(z_2) < z_2$, $h(z_3) < 1$, and $z_4 < 1$, the left-hand side is less than one while the right-hand side is greater than one, which completes the proof. Therefore, if $h(z) < 1 \ \forall z$, an equilibrium with multiple sets of workers and managers is impossible; the only possible equilibrium is one with a single cutoff z^* between workers and managers.

To prove that such an equilibrium exists is simple: we need $r'(z^*) > w'(z^*)$, and the latter is equal to $\frac{1 - w(z^*)}{1 - z^*}$, while $r'(z^*) > \frac{1}{h(z^*)}$ as proven above. Using $w(z^*) = r(z^*) = \frac{z^* - w(0)}{h(z^*)}$ to substitute for $w(z^*)$ in the expression for $w'(z^*)$, I find that the following is a sufficient condition for this equilibrium to exist:

$$1 > \frac{h(z^*) - z^* + w(0)}{1 - z^*}$$

which is satisfied if $h(z^*) + w(0) < 1$, and given that $w(0) < 1$, this is satisfied as long as h is small enough. In practice, I find that an equilibrium exists in the uniform case with constant h for $h < 0.916$. If h is larger, communication is sufficiently costly that no matching equilibrium can exist.

I have confirmed that, if $h(z) < 1$, any competitive equilibrium must feature a matching function with $m'(z) > 0$, with a single cutoff z^* between lower-skill workers and higher-skill managers, and I have proved that such an equilibrium must exist if $h(z^*)$ is small enough. Therefore, I can conclude that a competitive matching equilibrium in my model must feature perfect positive assortative matching.

A.2 Proof with Wage Bargaining

The proof with wage bargaining is somewhat different. With a fixed output-sharing rule of $w = \beta z_m$, all workers want to work for the highest-skill manager available; and the manager gets the same rent of $(1-\beta)z_m$ per unit of worker time no matter what their skill level, but must spend more time $h(z_m)(1-z_p)$ on lower-skill workers, so they also strictly prefer the highest-skill worker possible. Therefore, the equilibrium must feature a one-to-one mapping between z_p and z_m, with positive assortative matching between individuals in the set of workers and those in the set of managers, as otherwise a profitable deviation would exist. In the interior of the set of workers, the matching function follows equation (11).

Additionally, there must be workers at the bottom of the distribution; otherwise, some point in the distribution features workers matched with a zero-skill manager and receiving zero income, and those workers would be strictly better off matching amongst themselves. Similarly, there must be managers at the top of the distribution; otherwise, the top individual could hire someone with $z = 1-\epsilon$ and receive arbitrarily large rents as ϵ goes to zero. Therefore, the equilibrium must either feature perfect positive assortative matching with a single cutoff z^*, or consecutive disjoint sets of workers and managers.

To prove that the latter is impossible, consider a situation in which individuals with skill up to z_1 are workers, those between z_1 and z_2 are managers, those between z_2 and z_3 are workers, and so on with any number of alternating blocks of workers and managers. Denote wages as $w_{01}(z)$ on the first segment and $w_{23}(z)$ on the third segment, with the manager's wage denoted as $r_{12}(z)$ on the segment in between. If this is an equilibrium, it must be true that:

$$
\begin{aligned}
w_{01}(z_1) &= r_{12}(z_1) \\
\lim_{z\uparrow z_1} \frac{\partial w_{01}(z)}{\partial z} &< \lim_{z\downarrow z_1} \frac{\partial r_{12}(z)}{\partial z} \\
\lim_{z\uparrow z_2} \frac{\partial r_{12}(z)}{\partial z} &< \lim_{z\downarrow z_2} \frac{\partial w_{23}(z)}{\partial z}.
\end{aligned}
$$

The first equation requires that individuals at z_1 are indifferent between being workers and managers, and can be written simply as $\beta z_2 = \frac{(1-\beta)z_1}{h(z_1)}$, or, more useful for what follows, $\frac{\beta h(z_1)}{1-\beta} = \frac{z_1}{z_2}$. The two inequalities ensure that individuals marginally above and below the relevant cutoffs become workers or managers as required: if the first condition is not satisfied, a manager at z_2 would like to hire a manager at $z_1+\epsilon$ to become their worker, while if the second is not satisfied, a manager at z_3 would like to hire a worker at $z_2-\epsilon$.

Assume that the equality is true, and consider the two inequalities. The first important point to notice is that $\frac{\partial w_{01}(z)}{\partial z} = \frac{\partial w_{23}(z)}{\partial z} = 0$; from the perspective of a fixed manager z_m, the wage that must be offered to workers of different skill levels is constant at βz_m. This simplifies the problem considerably; we simply need to test whether or not $\lim_{z\downarrow z_1} \frac{\partial r_{12}(z)}{\partial z} > 0$ and $\lim_{z\uparrow z_2} \frac{\partial r_{12}(z)}{\partial z} < 0$. It is easy to show that the first condition is always satisfied while the second is never satisfied, because the partial derivative of r is always positive:

$$
\frac{\partial r_{12}(z)}{\partial z} = (1-\beta)\frac{h(z)-h'(z)z}{h(z)^2(1-m^{-1}(z))} \geq \frac{1-\beta}{h(z)(1-m^{-1}(z))} > 0.
$$

This result implies that no equilibrium with multiple alternating blocks of workers and managers can exist, because any point where individuals to the left become managers and individuals to the right become workers is infeasible: a manager at a higher skill level would always like to hire a worker from the pool of managers slightly below the cutoff. Therefore, the only possible equilibrium features perfect positive assortative matching with a single cutoff z^*. And such an equilibrium must exist, because the proof of the first inequality applies to any z_1, including the z^* in the perfect positive assortative matching equilibrium.

B Proof of Flat Bargained Wages in Simplified Model

The full model from section 3 is more complicated than that from Antràs, Garicano, and Rossi-Hansberg (2006), as it includes a labour supply dimension, which also makes a proof of the effect of bargaining on the slope of wages difficult because the matching function is not constant. Therefore, in this appendix I will consider a simplified version of the model, corresponding to the original model in Antràs, Garicano, and Rossi-Hansberg (2006), but with wage bargaining as well as competitive wages, to prove that, at least in this simplified setting, Nash bargaining necessarily makes the wage function flatter with respect to skill.

In this simplified model, the labour supply of each individual is exogenously fixed at 1, and I assume that h is a constant across skill levels, along with a uniform skill distribution. Therefore, the matching function does not depend on the wage-setting mechanism, and since $n = \frac{1}{h(1-z_p)}$ when labour supplies are set to 1, I can solve as follows:

$$\int_0^{z_p} f(z)dz = \int_{m(0)}^{m(z_p)} \frac{1}{h(1-m^{-1}(z))} f(z)dz \quad \forall z_p \leq z^*.$$

Differentiating with respect to z_p and rearranging, the matching function is defined by:

$$m'(z) = h(1-z)$$

along with the boundary conditions. This can be solved for a closed-form solution for the matching function:

$$m(z) = hz - \frac{h}{2}z^2 + z^*$$

where the cutoff z^* is:

$$z^* = \frac{h+1-\sqrt{h^2+1}}{h}.$$

Next, I can solve for the wage functions; for clarity, I will denote the bargained wage function as $w_m(z)$, and the competitive wage function as $w_c(z)$. In the bargaining case, the wage function is the simplified Nash bargain from section 3.2.2: $w_m(z) = \beta m(z)$, where $\beta = \frac{z^*}{h+z^*}$. The competitive wage also takes the same differential equation as before: $w_c'(z) = \frac{m(z)-w_c(z)}{1-z}$. I want to prove that $w_m'(z) < w_c'(z)\ \forall z \in [0, z^*]$.

First, let me solve for the second derivative of each wage function. Given that $m'(z) = h(1-z)$, I know that $w_m'(z) = \beta h(1-z)$, and therefore $w_m''(z) = -\beta h$. Meanwhile, the second derivative of the competitive wage function is:

$$w_c''(z) = \frac{(1-z)(m'(z)-w'(z)) + m(z) - w(z)}{(1-z)^2} = \frac{m'(z)}{1-z} = h.$$

Therefore, the competitive wage function must be convex with respect to skill, while the bargained wage must be concave.

Finally, consider the starting value of the competitive wage function, $w_c(0)$. There are three possible cases: $w_c(0) > w_m(0)$, $w_c(0) = w_m(0)$, and $w_c(0) < w_m(0)$. Start with the equality case; then $w_c'(0) = (1-\beta)m(0) = (1-\beta)z^*$, and since $1-\beta = \frac{h}{h+z^*}$, this means that $w_c'(0) = \frac{hz^*}{h+z^*}$. Meanwhile, $w_m'(0) = \beta h = \frac{hz^*}{h+z^*} = w_c'(0)$. Therefore, if the starting wage at $z = 0$ is equal for the competitive and the bargained wage functions, the slopes of the wage functions are also equal at that point, but the second derivative is higher everywhere in the competitive case. However, this is impossible, because it implies that $w_c(z^*) > w_m(z^*)$, and yet we know that $w_m(z^*)$ is the wage that makes a person at z^* indifferent between being a worker or manager given a starting wage of $w_m(0) = w_c(0)$, so the labour market does not clear in the competitive case: too many people want to be workers.

This suggests that the starting wage must be lower in the competitive case, but for completeness let me first consider the case when $w_c(0) > w_m(0)$. In this case, the starting slope of the wage function would be lower in the competitive case, but it is easy to show that this case is infeasible as well. Consider the wage slopes at any given value of $z \leq z^*$: $w_m'(z) = \beta m'(z) = \frac{h(1-z)z^*}{h+z^*}$, and $w_c'(z) = \frac{m(z)-w_c(z)}{1-z}$. Let me replace $w_c(z)$ with $w_m(z)+\chi$, where χ represents the amount by which the competitive wage is larger than the bargained wage at z, which can be positive or negative. Then $w_c'(z) = \frac{(1-\beta)m(z)-\chi}{1-z}$, which is larger than $w_m'(z)$ if and only if:

$$\frac{h}{h+z^*}\frac{hz - \frac{h}{2}z^2 + z^*}{1-z} - \frac{\chi}{1-z} > \frac{h(1-z)z^*}{h+z^*}$$

$$\frac{h}{h+z^*}\left[hz - \frac{h}{2}z^2 + z^* - z^*(1-z)^2\right] > \chi.$$

Because $h > 0$ and $z^* \in (0,1)$, the left-hand side is strictly positive: $\frac{h}{h+z^*} > 0$, and both $hz > \frac{h}{2}z^2$ and $z^* > z^*(1-z)^2$ because $z < 1$. Therefore, as long as $\chi = w_c(z) - w_m(z)$ is less than some positive number, $w_c'(z) > w_m'(z)$. This means that, by continuity, if $w_c(\hat{z}) > w_m(\hat{z})$ for some value $\hat{z} > 0$, the competitive wage can never drop below the bargained wage for any $z > \hat{z}$, because as $w_c(z)$ approaches $w_m(z)$ from above, it must enter a region where $w_c'(z) > w_m'(z)$, forcing the gap between the two wages to expand once again. Therefore, if $w_c(0) > w_m(0)$, then it must also be true that $w_c(z^*) > w_m(z^*)$; however, since the starting competitive wage is assumed to be higher than the starting bargained wage, the lowest-skill manager receives a lower wage than $w_m(z^*)$ with competitive wages, and thus the labour market does not clear in this case either.

Therefore, the competitive equilibrium must feature a wage function that starts at a lower value than the starting bargained wage: $w_c(0) < w_m(0)$. Then:

$$w_c'(0) = m(0) - w_c(0) > (1-\beta)m(0) = \frac{hz^*}{h+z^*} = w_m'(0)$$

and because $w_c''(z) > w_m''(z)$ for all $z \leq z^*$, this implies that $w_m'(z) < w_c'(z)\ \forall z \in [0, z^*]$, and the desired result has been proved. The competitive wage function must start at a lower level than the bargained wage, and is convex because of complementarity in manager and worker labour inputs. Meanwhile, the bargained wage is concave because total surplus depends on manager's skill, which is concave in worker skill according to the matching function.

C Calibration of Skill Distributions

For the calibration, I use data from the 2013 March CPS; I use the earnings of full-year, full-time individuals aged 25 to 64 who report earning at least a full year's federal minimum wages in 2012, and I exclude government employees, whose "firms" are not expected to fit the description of the model. Individuals are weighted using the March supplement weights. The data file is the 2013 ASEC Public Use file from the Census Bureau website, in which incomes above the topcode value are swapped rather than censored at the topcode, allowing for a reasonable approximation of the right tail of the income distribution.

To calibrate the wage bargaining case, I start at $h(z) = 0.5\exp\left(-0.5(h^{40})\right)$ and a flat skill distribution. Then, I test small changes in the parameters of h and in the density at 11 points along the skill distribution, $\{0, 0.1, ..., 1\}$, where the distribution is defined as a cubic spline across these points, and adjust the parameters of h and the spline nodes in the direction that reduces an objective function. Taking kernel densities of the real-world and model income distributions, using a Gaussian kernel over log income with a bandwidth of 0.2 for the data and 0.25 for the model, I define $q(y)$ as income y times the difference between the kernel densities at y; I then define the objective function as the sum of squares of $q(y)$ evaluated at evenly spaced points an interval of 5% of mean income apart, from 25% to 26 times mean income. To prevent very low densities, which could cause my matching algorithm to function poorly, I impose a minimum value of $f(z) = 0.1$ for the distribution. I also impose the restriction that $h(z) \leq 1.1$ in the wage bargaining calibration, to ensure that a restricted competitive allocation (as defined in section 4) can be computed for the purposes of calculating the efficient tax schedule.

In the competitive case, the same procedure is followed, except that the income distribution is calibrated to the wage bargaining case, to provide as close a match as possible between the two. No restrictions on the maximum value of h are required, as an unrestricted calibration always selects parameters for h that ensure existence and uniqueness of equilibrium. Kernel bandwidths of 0.25 are used for both distributions.

D Solution Methods

In this appendix, I describe the method used to solve for efficient taxes within the class described in section 4, as well as the more complex method used to solve for optimal taxes in all settings.

D.1 Solution Method for Efficient Taxes

As stated in section 4, an income tax is equivalent to assigning marginal tax rates directly to skill levels conditional on an identification constraint and an incentive-compatibility constraint. These constraints can be expressed algebraically as follows, where for simplicity the managers' "wage" $r(z)$ is written $w(z)$ to correspond with that of the workers. First, let income at skill level z be written as $y(z) \equiv w(z)L(z)$, and write the tax function as $T(z)$; then the identification constraint requires that $y'(z) > 0$, which simplifies to:

$$T''(z) < \frac{\gamma(1 - T'(z))w'(z)}{w(z)}.$$

Meanwhile, the incentive-compatibility constraint requires that utility is maximized at the $L(z)$ chosen under the skill tax; in the simulations, I check this globally, but the local condition is that the second derivative $\frac{d^2U}{dL^2} < 0$, or:

$$T''(z) > \frac{-(\gamma - 1)L^{\gamma-2}}{w^2}.$$

Therefore, the marginal tax rates cannot increase so fast as to make individuals choose income levels that decrease with skill, but they must also not decrease so fast as to encourage workers to exert excessive effort.

In the efficient taxation analysis, the first-best taxes are not incentive-compatible, as the marginal tax rates decline too quickly below z^*. Therefore, at each iteration, I apply an adjustment in which taxes are made to fit inside the boundaries imposed by the constraints. This procedure provides an approximation to the efficiency-maximizing tax schedule; the alternative is to use equation (13) to calculate the optimal tax with $\theta = 0$, but as described in section 5, this procedure is very complicated and itself is only capable of providing an approximation to the optimal tax. The method described in this appendix is a simple way of providing a close approximation.

In each case, I begin by choosing a value $\hat{z}$ for a central fixed point in the tax schedule; after each iteration, the tax rate at this point is held at the value that sets the labour supply of $\hat{z}$ to the value in the first-best allocation, and I then move both left and right from there, taking any tax rates that do not satisfy the constraints and assigning the maximum (to the left of $\hat{z}$) or minimum (to the right) tax rate that satisfies the constraint. The adjustment is itself performed iteratively: after each round of adjustments, I re-solve for equilibrium and test the constraints again, adjusting any that are still not satisfied.

For any selected value of $\hat{z}$, there is a resulting candidate tax schedule $t_c(z; \hat{z})$. I perform a numerical search for the value of $\hat{z}$ that provides the highest welfare $W(t_c(z; \hat{z}))$, where W is the average quasi-linear utility; the optimal values of $\hat{z}$ are 0.67 in the uniform case, 0.50 in the main calibration, and 0.46 with an elasticity of taxable income equal to one. This method finds the efficiency-maximizing tax schedule within a class of tax schedules that set labour supplies as close as possible to their efficient values subject to the incentive-compatibility and identification constraints. In the uniform case, the welfare gains are nearly identical to those from the first-best skill tax, while they are nearly 99% of the first-best welfare gains in the calibrated case. With an elasticity of taxable income equal to one, the first-best taxes deviate further from the constraints, but the second-best efficiency gains are still at least two-thirds of the first-best gains.

D.2 Solution Method for Optimal Taxes

To solve for the optimal tax schedule, I use equation (14) following an iterative procedure. Each round of iteration involves the following steps:

- I go through each individual one at a time and find the optimal tax rate within the bounds imposed by the identification and incentive-compatibility constraints, as described in section 5.2;
- I then re-solve for equilibrium, and the constraints change slightly;
- I then check the constraints, and adjust the tax rates to fit within the updated constraints, starting the adjustment from $0.99z^*$ ($0.9z^*$ in the flat-distribution case, as well as in the bargaining case when ETI $= 1$), and moving left and right from there;
- I repeat this process until the resulting tax schedule is entirely within the constraints.

In practice, the tax schedule adjusts very gradually towards the optimum, because the bounds imposed by the constraints are narrow but shift with the tax schedule.

Solving for the wage-shifting term S is the most time-consuming step, so I perform some number q of rounds of iteration at one time, and only re-solve the S term after all q rounds in order to save time; additionally, I solve S at 101 points, $z = \{0, 0.01, ..., 1\}$, and use 7th-order smoothed polynomials on each side of the threshold skill level z^* to approximate the function (with an elasticity of taxable income equal to one, I instead use one 13th-order smoothed polynomial across the whole distribution to prevent spikes).

At the end of each block of q rounds, before solving for the values of S for the next round, I also smooth the tax schedule using a polynomial best fit, as otherwise S becomes unstable and poorly behaved; I use a 13th-order polynomial below z^*, a 7th-order polynomial between z^* and $z = 0.9$, and finally a 7th-order polynomial *defined over income and not skill* above $z = 0.9$, to allow for considerable flexibility.[41] Then I go through the tax rates and adjust them as needed to ensure that they fit inside the bounds imposed by the constraints. I allow q to decline over time as the tax schedule converges, and the process concludes when q reaches 5 and the squared sum of shifts in the tax schedule over 5 iterations drops below 0.005 (which corresponds to a shift of about 0.0007 per individual).

E Derivation of Optimal Tax Equation

In this appendix, I consider the effect on individuals across the distribution when the government changes one of the tax rates t_i in the general model from section 5.1; I will separately consider the impact on individuals in $\{1, ..., i-1\}$ and $\{i, ..., Q\}$.

Impact of t_i on Individuals $q = \{1, ..., i-1\}$:

When the government raises the tax rate on individual i, there are only two effects on individuals at lower skill (and income) levels: there will be a change in the lump-sum transfer $T(0)$, and their wages may change as the labour market equilibrium adjusts. Utility of individual q is:

$$U_q = U\left(w_q L_q - T(w_q L_q), L_q\right)$$

so if I denote $b = T(0)$, the effect of a change in t_i is:

$$\frac{dU_q}{dt_i} = U_{Cq}\left[\frac{db}{dt_i} + (1-t_q)\left(L_q\frac{dw_q}{dt_i} + w_q\frac{dL_q}{dt_i}\right)\right] + U_{Lq}\frac{dL_q}{dt_i}.$$

Using the individual's first order condition $-U_{Lq} = U_{Cq}w_q(1-t_q)$, this simplifies to:

$$\frac{dU_q}{dt_i} = U_{Cq}\left[\frac{db}{dt_i} + (1-t_q)L_q\frac{dw_q}{dt_i}\right]$$

where the two terms in square brackets directly reflect the change in consumption due to changes in the lump-sum transfer b and in the wage w_q.

Impact of t_i on Individuals $q = \{i, ..., Q\}$:

Individuals at or above the skill level of individual i again receive a change in the lump-sum transfer and face a change in wages, but they also pay higher taxes; since the tax t_i applies to income from $(y_{i-1}, y_i]$, the reduction in after-tax income resulting from a one-unit increase in t_i is $y_i - y_{i-1}$, which I will denote as Δy_i. Therefore, the effect of a change in t_i is:

$$\frac{dU_q}{dt_i} = U_{Cq}\left[\frac{db}{dt_i} - \Delta y_i + (1-t_q)L_q\frac{dw_q}{dt_i}\right].$$

Total Effect of t_i on Welfare:

If I denote welfare as $W \equiv \sum_{q=1}^{Q} f(z_Q)U_q$, then the total impact of t_i on welfare is:

$$\frac{dW}{dt_i} = \sum_{q=1}^{Q} U_{Cq}f(z_q)\left[\frac{db}{dt_i} + (1-t_q)L_q\frac{dw_q}{dt_i}\right] - \sum_{q=i}^{Q} U_{Cq}f(z_q)\Delta y_i. \tag{15}$$

Finally, I need to solve for $\frac{db}{dt_i}$. Given a government budget constraint that requires that net tax revenues must be equal to some exogenous quantity of spending G (where the size of G does

[41] Since the income distribution is more compact in the uniform case, I only use two 13th-order polynomials in that case, one below z^* and one above.

not directly affect the optimal tax formula, only affecting the marginal utilities), this is equal to the derivative of the total tax revenues collected by the government with respect to t_i, so:

$$\frac{db}{dt_i} = \sum_{q=1}^{Q} f(z_q) t_q \frac{dy_q}{dt_i} + \Delta y_i \sum_{q=i}^{Q} f(z_q) \tag{16}$$

where $\frac{dy_q}{dt_i}$ can be expressed as:

$$\frac{dy_q}{dt_i} = L_q \frac{dw_q}{dt_i} + w_q \frac{dL_q}{dt_i}. \tag{17}$$

To solve for $\frac{dL_q}{dt_i}$, I use the first-order condition for labour supply: $(1-t_q)w_q = \frac{-U_{Lq}}{U_{Cq}} \equiv s_q$, where s is the marginal rate of substitution. Then I differentiate to obtain:

$$(1-t_q)\frac{dw_q}{dt_i} - w_q \frac{dt_q}{dt_i} = s_{Cq}\frac{dC_q}{dt_i} + s_{Lq}\frac{dL_q}{dt_i}$$

where $\frac{dt_q}{dt_i} = 1$ for $q = i$ and is zero otherwise. Substituting into (17), I find:

$$\frac{dy_q}{dt_i} = \left[L_q + (1-t_q)\frac{w_q}{s_{Lq}}\right]\frac{dw_q}{dt_i} - \frac{w_q^2}{s_{Lq}}\frac{dt_q}{dt_i} - \frac{w_q s_{Cq}}{s_{Lq}}\frac{dC_q}{dt_i}.$$

Finally, substituting this into (16), I obtain the following equation for $\frac{db}{dt_i}$:

$$\frac{db}{dt_i} = \sum_{q=1}^{Q} f(z_q) t_q \left[\left(L_q + (1-t_q)\frac{w_q}{s_{Lq}}\right)\frac{dw_q}{dt_i} - \frac{w_q s_{Cq}}{s_{Lq}}\frac{dC_q}{dt_i}\right] + \Delta y_i \sum_{q=i}^{Q} f(z_q) - f(z_i) t_i \frac{w_i^2}{s_{Li}}. \tag{18}$$

If I define $\sum_{q=i}^{Q} f(z_q) \equiv Q_i$, (15) and (18) can be combined to give:

$$\begin{aligned}\frac{dW}{dt_i} &= \Delta y_i Q_i \left[E(U_{Cq}) - E(U_{Cq}|q \geq i)\right] - f(z_i) t_i \frac{w_i^2}{s_{Li}} E(U_{Cq}) \\ &+ \sum_{q=1}^{Q} f(z_q)\left[t_q\left(L_q + (1-t_q)\frac{w_q}{s_{Lq}}\right)E(U_{Cq}) + (1-t_q)L_q U_{Cq}\right]\frac{dw_q}{dt_i} \\ &- \sum_{q=1}^{Q} f(z_q) t_q \frac{w_q s_{Cq}}{s_{Lq}} E(U_{Cq}) \frac{dC_q}{dt_i}.\end{aligned}$$

This equation is very complicated, depending not only on marginal utilities and incomes, but also on wages, changes in consumption, and derivatives of the marginal rates of substitution. To simplify somewhat, let me make a common assumption from the optimal income tax literature and assume away income effects; I make the same assumption in the analysis of the model in section 3. In particular, let me assume that the utility function takes the form $U = U\left(C - \frac{1}{\gamma}L^{\gamma}\right)$. Then $s_{Cq} = 0$ and $s_{Lq} = (\gamma-1)L_q^{\gamma-2}$, so the equation simplifies to:

$$\begin{aligned}\frac{dW}{dt_i} &= \Delta y_i Q_i \left[E(U_{Cq}) - E(U_{Cq}|q \geq i)\right] - f(z_i)\frac{y_i}{(\gamma-1)}\frac{t_i}{1-t_i}E(U_{Cq}) \\ &+ \sum_{q=1}^{Q} f(z_q) L_q \left[\frac{\gamma}{\gamma-1} t_q E(U_{Cq}) + (1-t_q)U_{Cq}\right]\frac{dw_q}{dt_i}.\end{aligned}$$

This equation is the result stated in Proposition 2.

This expression can also easily be generalized to a continuous distribution: the only variables whose meaning depends on the number and spacing of mass points in the distribution are $f(z)$, Δy_i, and $\frac{dw_q}{dt_i}$, where the latter depends on the size of the interval affected by the tax increase t_i. Therefore, if I divide the equation above by $z_i - z_{i-1}$ and let Q go to infinity so that the gap between mass points goes to zero, the $f(z)$ terms become $\lim_{z_{i-1}\to z_i} \frac{F(z_i)-F(z_{i-1})}{z_i - z_{i-1}}$, which turns into a density $f(z)$ rather than a mass, and Δy_i is replaced by $\lim_{z_{i-1}\to z_i} \frac{y(z_i)-y(z_{i-1})}{z_i-z_{i-1}} = \frac{dy}{dz_i}$. Finally, $\lim_{z_{i-1}\to z_i} \frac{\frac{dw_q}{dt_i}}{z_i-z_{i-1}}$ can be denoted as $\lim \frac{dw_q}{dt_i}$, and the expression for $\frac{dW}{dt_i}$ can be written in continuous terms:

$$\frac{dW}{dt_i} = \frac{dy}{dz_i} Q_i \left[E(U_{Cq}) - E(U_{Cq}|q \geq i)\right] - f(z_i)\frac{y(z_i)}{(\gamma-1)}\frac{t_i}{1-t_i}E(U_{Cq}) + \int_0^1 f(z_q)L_q\left[\frac{\gamma}{\gamma-1}t_q E(U_{Cq}) + (1-t_q)U_{Cq}\right]\lim\frac{dw_q}{dt_i}dq$$

where $q, i \in [0,1]$ and $Q_i = \int_i^1 f(z_q)dq$.

F Additional Numerical Results

I now present a series of additional numerical and graphical results. Subsections F.1 and F.2 present sensitivity analyses, showing that the main results are not sensitive to the use of a flat skill distribution or a larger elasticity of taxable income, while section F.3 shows that the effect of wage bargaining on optimal taxes is similar when a Pareto tail is added above my distribution. Finally, section F.4 presents optimal bracketed taxes, which are qualitatively similar to the full non-linear taxes.

F.1 Sensitivity Analysis with Uniform Distribution

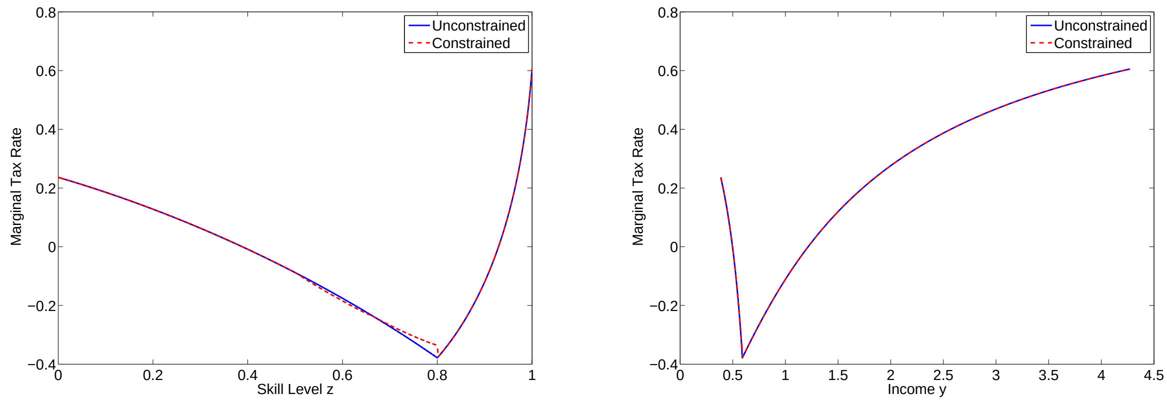
In this appendix, I present the efficient and optimal taxes with a uniform skill distribution and $h = 0.5$. This numerical exercise should not be taken literally, as there is no reason to believe that these assumptions are appropriate, but it provides a simple demonstration of the robustness of my general findings to an alternative calibration

I solve for the efficient taxes - both the first-best and the second-best within the class defined in section 4 - as before, and the results are displayed in Figure 11. Once again, the efficient taxes deviate from zero by a large amount, but with a slightly different pattern this time, featuring smaller positive taxes at the bottom of the distribution, and larger tax rates at the top that reach as high as 60%. With a uniform skill distribution and flat h, the cutoff between workers and managers is higher than in the calibrated results, and so the top managers can obtain larger rents from their workers, who are both more highly skilled (and thus easy to supervise) and significantly underpaid, necessitating larger taxes at the top. The second-best tax schedule is nearly identical to the first-best skill tax; in fact, it seems to coincide almost perfectly over the vast majority of the distribution, with a very small difference between $z = 0.5$ and $z = 0.8$.

I also use the same procedure as in section 5 to solve for the optimal taxes in both the competitive and wage bargaining frameworks, for $G = 0.12$, which is a bit less than 20% of average income prior to taxes in both the competitive and bargaining frameworks, and thus meant as a rough approximation to total tax revenues as a percent of GDP in the U.S. The results can be found

Figure 11: Efficient Taxes with Wage Bargaining and Uniform Distribution & Constant h

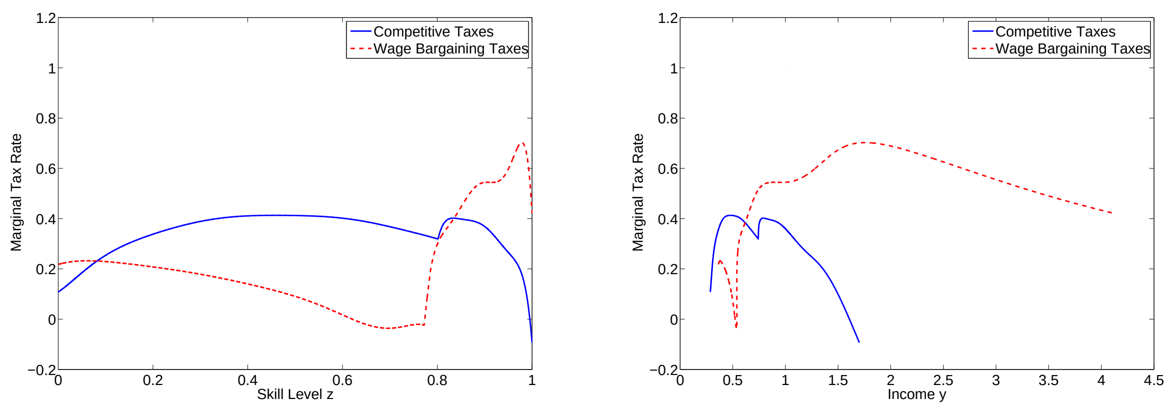
(a) As Function of Skill (b) As Function of Income



in Figure 12, where we can observe that, as in the main analysis, the optimal marginal tax rates roughly follow an inverted-U shape with competitive wages, but V-shaped at the bottom with bargained wages, then rising to about 70% before declining to about 40% at the top. Moving from zero marginal taxes to the optimal taxes generates welfare gains that are equivalent to a 6.37% and 7.37% increase in consumption respectively in the two cases.

Figure 12: Optimal Tax Schedules with Uniform $F(z)$, Constant h & Log Utility

(a) As Function of Skill (b) As Function of Income



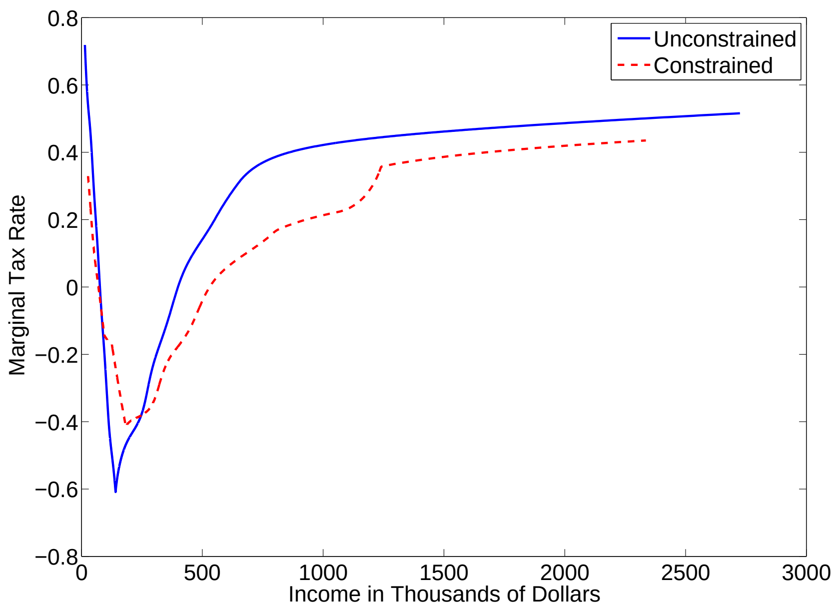
The results in this appendix demonstrate that the findings in sections 4 and 5 with the calibrated skill distribution are robust to the assumption of a uniform distribution; the efficient tax schedule remains strongly V-shaped, and the effect of wage bargaining on optimal taxes with log utility is also V-shaped.

F.2 Sensitivity Analysis on Elasticity of Taxable Income

In this appendix I redo all of the numerical analysis of sections 4 and 5 using a higher value of the elasticity of taxable income (ETI). The correct value of this elasticity has been the subject of a significant controversy, with many of the earlier estimates being considerably larger than the 0.12-0.4 range stated as plausible by Saez, Slemrod, and Giertz (2012).[42] A dramatic example is Feldstein (1995), who finds that the ETI is at least one; therefore, in this appendix, I assume an elasticity of one, or a value of $\gamma = 2$.

I calibrate h and the skill distribution to the U.S. income distribution using the same procedure as before. The efficient taxes with wage bargaining can be found in Figure 13, and a comparison of this figure with Figure 5 demonstrates that the conclusion about efficient taxes with wage bargaining is unaffected by a higher value of ETI: the overall shape is the familiar V, and the efficient top tax is even higher than before, at over 40%. The welfare gain from implementing the efficient tax schedule is a very large 15.20% of mean consumption. The optimal unconstrained tax both drops and rises faster than is incentive-compatible, and the adjustment for the constraint means that individuals around a spike in the skill distribution near $z = 0.1$ are taxed too little and work too hard; therefore, more managers are required due to the constraint, z^* declines, and the top managers are matching with less skilled workers, which is why the top income drops considerably.

Figure 13: Efficient Taxes with ETI = 1, Wage Bargaining and Calibrated Distribution



The optimal taxes with log utility are in Figures 14 and 15. The solution procedure was more complicated in the bargaining case, with the concave and convex baseline tax schedules both multiplied by 0.5, and the final tax schedule is smoothed using a triangular moving average. The

[42] Furthermore, some research has indicated that the ETI may not be a constant, with Gruber and Saez (2002) finding larger values at higher income levels, and Keane and Rogerson (2012) highlight the difference between the standard micro-level estimates and a long-run macro labour supply elasticity. Additionally, in a recent paper, Weber (2014) argues that most recent estimators of ETI are inconsistent; she finds an elasticity of 0.858 with a new estimator.

level of the optimal tax schedule drops considerably with competitive wages, but the drop at the top is much smaller in the wage bargaining case, compared to the ETI = 0.25 case. The welfare gains are 1.04% with competitive wages and 5.66% with wage bargaining in Figure 14, and 0.71% and 6.44% respectively in the Pareto tail analysis in Figure 15.

Figure 14: Optimal Tax Schedules with ETI = 1 and Log Utility

(a) As Function of Skill (b) As Function of Income

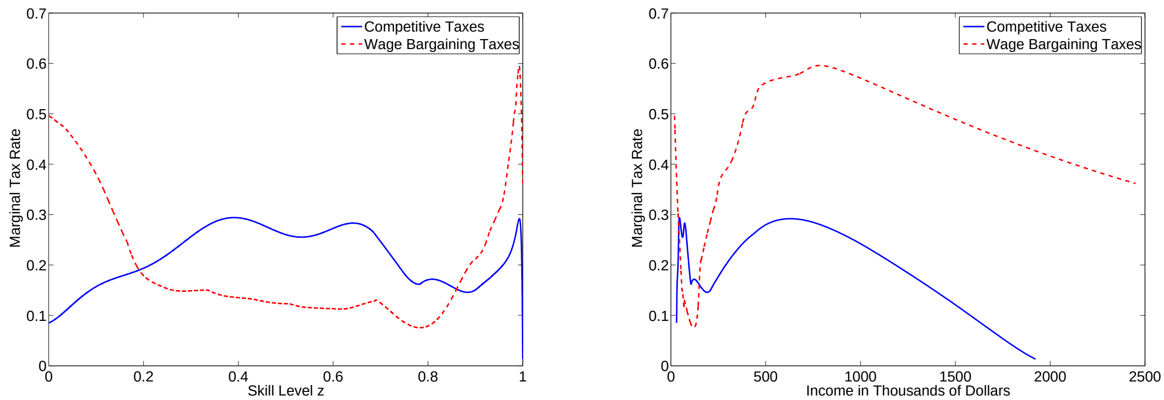
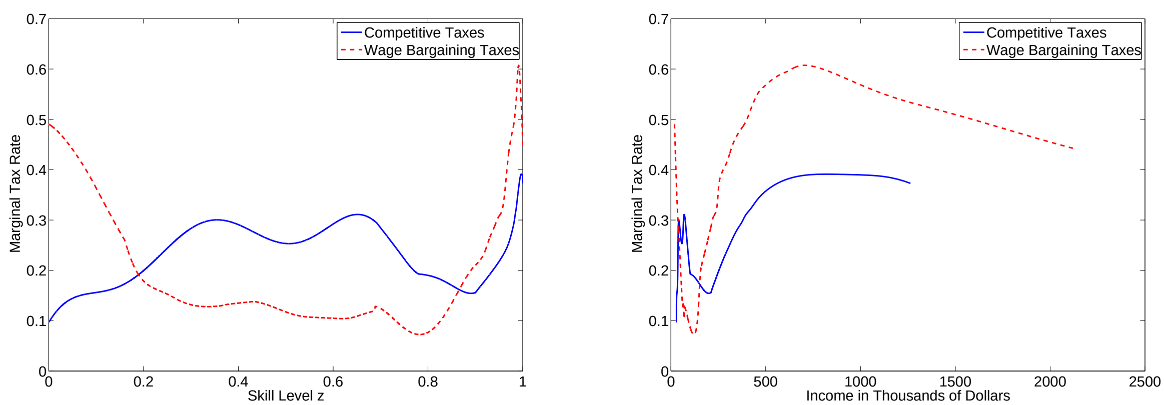


Figure 15: Optimal Tax Schedules with ETI = 1, Log Utility and Pareto Tail

(a) As Function of Skill (b) As Function of Income



This appendix has demonstrated that the case for an efficient V-shaped tax is not sensitive to the elasticity of taxable income; meanwhile, the optimal tax schedule with wage bargaining is only moderately altered, even while optimal taxes with competitive wages drop significantly. Furthermore, the welfare gains from optimal taxes might even increase with a higher value of ETI. The reason for these results is quite simply that the efficiency role of marginal taxes with wage bargaining does not depend directly on the responsiveness of individual behaviour to taxes; what matters is how far the wage deviates from the efficient level, and this deviation remains large with a higher ETI.

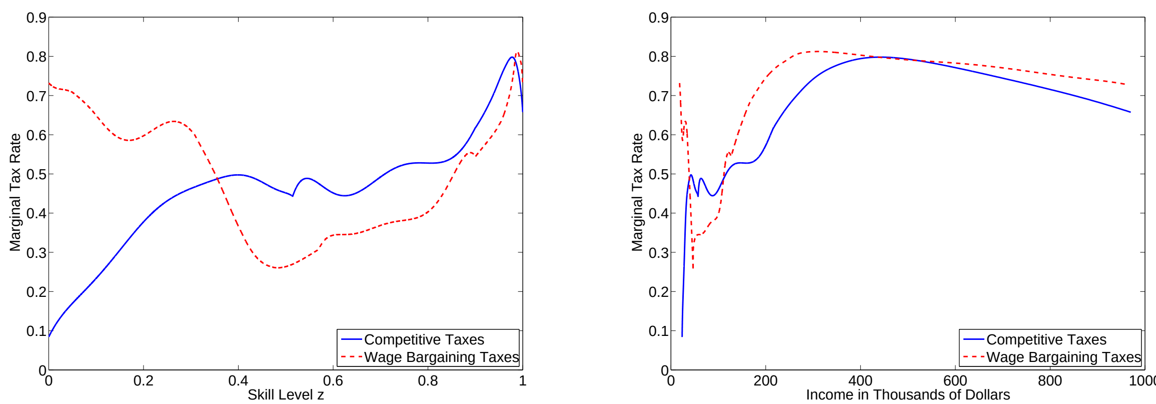
F.3 Optimal Taxes with Pareto Tail

The results for optimal income taxes presented in section 5.2 feature a strong decline above about \$400 thousand, due primarily due to the fact that there is a finite top to the income distribution.[43] A finite top to the income distribution eliminates the redistributional role of marginal taxes at the top of the distribution; however, a Pareto tail to the income distribution generates a positive optimal asymptotic top rate, as seen in Diamond and Saez (2011). I cannot solve for an asymptotic top tax rate in my model, because computationally, the model must have a top, a highest income for the highest-skill individual.[44]

However, an approximation can be provided if we assume that there is some positive mass of very-high-income individuals outside our model. Assume that their incomes are so large that the planner acts as if their marginal utilities of income are zero, and thus they are not considered as part of the measure of welfare; they are only a source of revenue when considering taxes on individuals within our model. Then, if f_{Q+1} is the size of this population, we simply need to add $\Delta y_i f_{Q+1} E(U_{Cq})$ to equation (13), to capture the value of the additional revenues from the super-rich.[45] Diamond and Saez (2011) find that the first percentile of the U.S. income distribution was at about \$400 thousand in 2007, whereas the top of my model's distribution is at about \$1.2 million; assuming a Pareto tail with shape parameter 1.5, as in Diamond and Saez (2011), this implies that about 20% of the top percentile had incomes above the top of my model, and therefore $f_{Q+1} = \frac{0.002}{0.998}$. The resulting optimal tax schedules are presented in Figure 16.

Figure 16: Optimal Tax Schedules with Log Utility and Pareto Tail

(a) As Function of Skill (b) As Function of Income



The results for optimal taxes are very similar to Figure 10 across most of the distribution, as is the pattern of effects of wage bargaining relative to competitive wages. However, the decline in tax

[43] Additionally, I am using a general equilibrium model, and thus taxes generate general equilibrium effects somewhat like those in Stiglitz (1982), whereby taxing managers at the top reduces their labour supply and thus forces workers to be supervised by less talented managers, with redistributional costs from lower wages further down the distribution.

[44] Note that having $z > 1$ is meaningless given that problems can only be of difficulty 1 or less. Since skill z only has meaning relative to its place in the distribution of problems d, the latter can always be normalized to be uniform between zero and one, so that $z \in [0, 1]$ represents the fraction of problems that can be solved.

[45] To prevent discontinuous results, I set $\Delta y_1 = \Delta y_2$, assuming that the bottom tax rate only raises revenue on a small amount of income rather than all income from zero to $y(z_1)$.

rates at the top is much less pronounced, with optimal top rates of about 70%. The top tax rate is again higher with wage bargaining, though by less than in Figure 8, because the redistribution term dominates and taxes cannot be raised as much from a high level. The welfare gains from implementing the optimal tax now increase to 5.56% in the competitive case and 6.19% with wage bargaining.

F.4 Optimal Bracketed Taxes

To complement the full non-linear optimal taxes calculated earlier, I now present the optimal tax schedule when marginal taxes are restricted to be constant within brackets. I have fixed the brackets at values close to the existing tax brackets facing single taxpayers in the U.S. (counting the personal exemption and standard deduction): less than $50k, $50-100k, $100-200k, $200-400k, and $400k+. In the current federal tax code, the marginal tax rates facing individuals over most of these ranges are, respectively, 15%, 25%, 28%, 33%, and 39.6%.

I search numerically for the set of marginal tax rates that maximize welfare, smoothing the tax rates near the thresholds to prevent discontinuous jumps (this does not prevent large numbers of people from clustering near the threshold). The results are presented below in Tables 1 and 2, for each of 4 cases: competitive and wage bargaining in both the baseline calibrated setting and the calibration with an ETI of one. Table 1 presents the results from the standard approach, while Table 2 adds the Pareto tail approximation. The results are less dramatic than in the fully non-linear case because individuals at the top and bottom of the distribution, where most of the variation in optimal tax rates occur, are collected together into relatively large brackets, but they continue to show the usual results: optimal taxes with competitive wages are generally in the shape of an inverted U, whereas they are more V-shaped with wage bargaining, higher at the top and bottom and lower in the middle. The final column in the table presents the welfare gains as a percentage of consumption as before; in every case, the gains are at least 50% as large as those from the full non-linear tax.

Table 1: Optimal Bracketed Taxes

	$0-50k	$50-100k	$100-200k	$200-400k	$400k+	Welfare Gain
Competitive (Baseline)	41.09%	45.34%	51.51%	68.38%	60.02%	3.67%
Bargaining (Baseline)	50.43%	37.00%	48.81%	70.95%	65.55%	4.02%
Competitive (ETI = 1)	22.62%	23.82%	17.90%	23.46%	22.25%	0.60%
Bargaining (ETI = 1)	41.92%	13.81%	12.42%	37.00%	47.04%	4.50%

Table 2: Optimal Bracketed Taxes with Pareto Tail

	$0-50k	$50-100k	$100-200k	$200-400k	$400k+	Welfare Gain
Competitive (Baseline)	41.01%	45.40%	53.73%	74.14%	64.97%	5.26%
Bargaining (Baseline)	51.93%	37.80%	49.55%	74.89%	66.54%	5.69%
Competitive (ETI = 1)	23.26%	23.29%	19.66%	31.74%	31.22%	0.55%
Bargaining (ETI = 1)	41.26%	14.10%	13.51%	35.21%	47.60%	5.05%

G Optimal Taxation in the Mirrleesian Method: A Simple 2-Type Model

The optimal income tax analysis in section 5 uses a perturbation method, analogous to that proposed by Saez (2001). The use of a Mirrleesian method, in which the optimal allocation of consumption and labour supply (or income) is solved subject to incentive compatibility constraints, is computationally infeasible for the reasons described at the beginning of section 5. However, to illustrate the method, I now present a simple 2-type model of matching in which the complications preventing the use of the Mirrleesian method are no longer present.

The model is related to the one presented in section 2.3.1, but with occupation as a choice rather than exogenously fixed. The economy consists of individuals of two skill types, $z_2 > z_1 \geq 1$, with a mass of measure 1 of each. Individuals decide whether to become workers or managers, and each manager receives a team size equal to $n(L_i) = \frac{L_i}{E(L_m)}$, where $E(L_m)$ is the average labour supply among managers. Within a team of one manager and n workers, team output is given by $Y = z_m(z_m L_m + n(L_m)E(z_p L_p))$, where $E(z_p L_p)$ is the average among the workers in the team.

I assume that individuals working alone receive only $z_i L_i$, so each individual is always weakly better off working in a team. Production efficiency then requires that type-2 individuals become managers and type-1 become workers.[46] Alternative matching patterns can easily be shown to lead to lower output; thus, switching to the efficient allocation, holding labour supplies and consumption fixed, would lead to additional revenues to the government, which could be used to increase consumption for one or both types. Therefore, I will focus on the equilibrium with type-2 managers and type-1 workers, and I will need to consider the incentive-compatibility constraints that ensure that such an equilibrium holds with taxation.

As in the models throughout the paper, the manager is assumed to be residual claimant, and pays a wage to the worker. Clearly, each type-1 worker's marginal product per unit of labour supply is equal to $z_2 z_1$, and this is true regardless of the tax system, so any competitive equilibrium will feature a wage fixed at $w = z_2 z_1$. I will also consider inefficient bargaining allocations in which the wage is exogenously fixed at a different value.

Individuals receive utility from consumption and disutility from labour supply, $U = U(C, L)$, where the marginal utilities are $U_C > 0$ and $U_L < 0$. The planner is a utilitarian with equal weights on all individuals, and thus wants to maximize:

$$V = U(C_1, L_1) + U(C_2, L_2)$$

subject to a set of incentive-compatibility constraints, as in the standard Mirrleesian problem; I will also need to verify that income is higher for type 2, as in the identification constraint in my model. However, in this case, I assume that the planner chooses C and L, rather than C and income (which I will denote as y), because the manager's income depends on the wage and on both types' labour supply. The tax system that implements the optimal allocation can then be backed out, with $T(y_1) = y_1 - C_1$ and $T(y_2) = y_2 - C_2$, where $y_1 = wL_1$ and $y_2 = z_2^2 L_2 + (z_2 z_1 - w)L_1$ in equilibrium (as $n = 1$ for all managers in equilibrium). The marginal taxes are not well defined in the usual sense, since with only two types and thus two income levels, I only need to define the tax schedule at two points. However, I define marginal tax rates in an implicit sense as in Stiglitz (1982): if the

[46] Output is unchanged if only some fraction of type-2 individuals become managers, while the rest join the type-1 individuals as workers. However, I assume that this is inefficient for reasons outside the model. For example, suppose each team produces a slightly different variety, and the number of produced varieties enters positively into utility up to a maximum of one unit of varieties; then if fewer than one unit of teams are formed, utility decreases.

tax schedule $T(y)$ was differentiable, any individual with wage w_i would set $\frac{U_{L_i}}{U_{C_i}} = -w_i(1-T'(y_i))$. Therefore, I can define the marginal tax rate faced by type i as $t_i = 1 + \frac{U_{L_i}}{w_i U_{C_i}}$ at the optimal allocation.

Finally, I need to define the incentive-compatibility constraints. The planner chooses consumption C_1 and C_2, and labour supply L_1 and L_2, but what they can actually observe is income y_1 and y_2, and they assign consumption C_i to individuals producing income y_i. Therefore, each type faces a choice of whether to produce an income of $y_1 = wL_1$ or $y_2 = z_2^2 L_2 + (z_2 z_1 - w)L_1$, and whether to do so by becoming a worker or a manager. The welfare-maximizing choices of C and L must therefore be consistent with type-1 individuals becoming workers and providing income y_1, and type-2 individuals becoming managers and providing income y_2. Thus, each type faces three incentive-compatibility constraints.

Consider first the type-1 individuals. They must prefer being a worker at income y_1 over the three alternatives, which imply the following constraints:

(I) to reach income y_1 as a manager, a type-1 individual would require a labour supply equal to $L_I = \frac{y_1}{z_1^2 + \frac{L_1}{L_2}(z_1^2 - w)}$; holding consumption C_1 fixed, I need that $L_I \geq L_1$, or $w \geq z_1^2 + \frac{L_1}{L_2}(z_1^2 - w)$;

(II) to reach income y_2 as a worker, a type-1 individual would require a labour supply equal to $L_{II} = \frac{y_2}{w}$; therefore, I need that $U(C_1, L_1) \geq U\left(C_2, \frac{y_2}{w}\right)$;

(III) to reach income y_2 as a manager, a type-1 individual would require a labour supply equal to $L_{III} = \frac{y_2}{z_1^2 + \frac{L_1}{L_2}(z_1^2 - w)}$; therefore, I need that $U(C_1, L_1) \geq U\left(C_2, \frac{y_2}{z_1^2 + \frac{L_1}{L_2}(z_1^2 - w)}\right)$.

However, constraints (I) and (II) imply that constraint (III) is satisfied, since $w \geq z_1^2 + \frac{L_1}{L_2}(z_1^2 - w)$ implies that $L_{III} \geq L_{II}$; therefore, I only need constraints (I) and (II).

A similar set of constraints can be presented for the type-2 individuals:

(IV) to reach income y_2 as a worker, a type-2 individual would require a labour supply equal to $L_{IV} = \frac{y_2}{w}$; holding consumption C_2 fixed, I need that $L_{IV} \geq L_2$, or $w \leq z_2^2 + \frac{L_1}{L_2}(z_2 z_1 - w)$;

(V) to reach income y_1 as a worker, a type-2 individual would require a labour supply equal to $L_V = \frac{y_1}{w}$; therefore, I need that $U(C_2, L_2) \geq U\left(C_1, \frac{y_1}{w}\right)$;

(VI) to reach income y_1 as a manager, a type-2 individual would require a labour supply equal to $L_{VI} = \frac{y_1}{z_2^2 + \frac{L_1}{L_2}(z_2 z_1 - w)}$; therefore, I need that $U(C_2, L_2) \geq U\left(C_1, \frac{y_1}{z_2^2 + \frac{L_1}{L_2}(z_2 z_1 - w)}\right)$.

Once again, only two of these constraints are relevant: if $w \leq z_2^2 + \frac{L_1}{L_2}(z_2 z_1 - w)$, then $L_V \geq L_{VI}$, and I only need constraints (IV) and (VI).

Some of the constraints can be further simplified; (I) can be rearranged simply to give $w \geq z_1^2$, while (IV) similarly can be rewritten as $w \leq \frac{z_2(L_1 z_1 + L_2 z_2)}{L_1 + L_2}$, thus creating a range of feasible values of w. Finally, constraint (VI) can be simplified to give $U(C_2, L_2) \geq U\left(C_1, \frac{y_1 L_1}{y_2}\right)$. If G is the resource requirement of the planner, then the full planner's problem can then be written out as follows:

$$\begin{aligned}
\max_{C_1,L_1,C_2,L_2} \; & V = U(C_1,L_1) + U(C_2,L_2) \\
s.t. \; & w \geq z_1^2 \\
& U(C_1,L_1) \geq U\left(C_2, \frac{y_2}{w}\right) \\
& \frac{z_2(L_1 z_1 + L_2 z_2)}{L_1 + L_2} \geq w \\
& U(C_2,L_2) \geq U\left(C_1, \frac{y_1 L_1}{y_2}\right) \\
& (y_1 - C_1) + (y_2 - C_2) \geq G
\end{aligned}$$

where $y_1 = wL_1$ and $y_2 = z_2^2 L_2 + (z_2 z_1 - w)L_1$ as stated above, and where I also need to verify that $y_2 > y_1$.

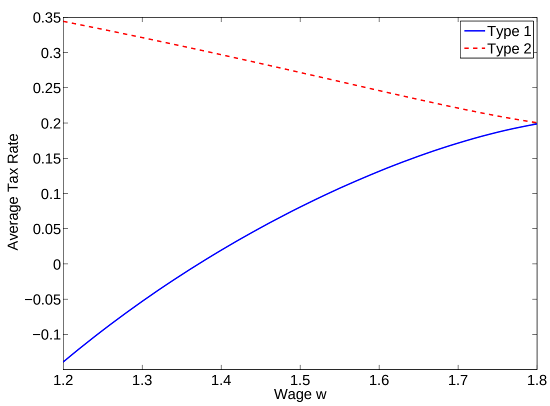
At this point, it would be possible to derive the first-order conditions, and characterize the solution depending on which constraints are binding. However, in this case which constraints are binding is likely to depend on the wage, and the first-order conditions are quite complicated because of the matching structure; thus, the insight available from an analytical solution is limited. For a numerical solution, I will instead solve the problem as explicitly presented above: I search for the vector $\{C_1, L_1, C_2, L_2\}$ that maximizes V while satisfying all of the constraints. I consider a case in which $U(C,L) = \log\left(C - \frac{1}{\gamma}L^{\gamma}\right)$, with $\gamma = 5$ as before; I also set $z_1 = 1$ and $z_2 = 1.5$, and the revenue requirement is set to $G = 0.88$, which is approximately 20% of total output at the optimum when the wage takes the competitive value of $w = 1.5$.

The wage is varied between 1.2 and 1.8, and the resulting optimal taxes are calculated, both the average tax paid $\frac{T(y_i)}{y_i}$ and the implicit marginal tax rate $t_i = 1 + \frac{U_{L_i}}{w_i U_{C_i}}$. Figure 17 presents the result, with the average taxes in panel (a) on the left and the marginal taxes in panel (b) on the right; in every case, $y_2 > y_1$ is satisfied along with all of the incentive-compatibility constraints. The results are exactly analogous to those from section 5: the optimal marginal tax rate on the type-1 workers increases with their wage, while the tax rate on the type-2 managers decreases with w. At the efficient wage of $w = 1.5$, the optimal marginal rate on the workers is positive, while that on the managers is essentially zero; however, if the wage is inefficiently low, the workers should face a lower tax, and in fact a negative tax rate if w is sufficiently low, whereas the managers should face an increased tax, to offset the distortions to both of their labour supplies. As usual, the reverse applies if the wage is inefficiently high.

This analysis demonstrates that the main result from section 5 - that optimal marginal income taxes should go up for individuals receiving inefficiently high wages, and down for those receiving too little - is unchanged in this analysis. However, this appendix also allows us to see even more clearly why the Mirrleesian method is infeasible with the full general equilibrium matching model of section 3: to be able to write down the incentive-compatibility constraints, I need to know the wages faced by agents at the optimum. Strong assumptions in the current model pin down these wages and ensure that I can write out the incentive-compatibility constraints, though even here they are relatively complicated due to matching. Once those strong assumptions about fixed wages are removed, the difficulty of the problem is greatly increased: to find out what wages would be at the optimum, I would need to know who is matching with whom, but solving the "outer problem" (in the terminology of Rothschild and Scheuer (2014)) to find the optimal matching function is simply not computationally feasible in the main model of section 3.

Figure 17: Optimal Taxes from 2-Type Model

(a) Average Taxes $\frac{T(y_i)}{y_i}$



(b) Marginal Taxes $t_i = 1 + \frac{U_{L_i}}{w_i U_{C_i}}$