Taxing the Job Creators: Efficient Progressive Taxation with Wage Bargaining

Nicholas Lawson*

Aix-Marseille University (Aix-Marseille School of Economics), CNRS & EHESS

nicholas.lawson@univ-amu.fr

September 12, 2014

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 it tends to increase equality. However, when wages deviate from marginal product, the laissez-faire equilibrium is inefficient, and there can be an independent efficiency rationale for income taxation. I study a setting of wage bargaining within hierarchical teams of workers and managers, and show that the efficiency case for taxing managers depends on a "job-creation" effect: if increased labour supply allows managers to supervise larger teams and thus collect larger rents, they will have an incentive to work too hard to create jobs at their firm. In other words, it is *because* of their job-creation activity that the "job creators" should be heavily taxed. Simulation of a calibrated model suggests an efficient tax schedule that is progressive over most of the income distribution with a top marginal rate of between 50% and 60%, and this result is not sensitive to the magnitude of the labour supply response to taxation. For a planner with redistributive motives, optimal marginal tax rates are also considerably higher at the top of the distribution in the presence of wage bargaining rather than a competitive labour market.

Keywords: optimal income taxation, progressive taxation, wage bargaining, team production

*I thank participants of seminars at Princeton University, the Aix-Marseille School of Economics, and the Delhi School of Economics, for their input. I also thank Robin Boadway, Andrew Clark, Cecilia García-Peñalosa, Jacob Goldin, Nikolaj Harmon, Marc Sangnier, Dean Spears and Alain Trannoy for their helpful comments and suggestions. Any errors or omissions are the responsibilty of the author.

1 Introduction

The standard economic view of the personal income tax is that it is a distortionary fiscal instrument which nonetheless has value because it is an equity-enhancing way of raising revenue.[1] This view has shaped the optimal income taxation literature that started with Mirrlees (1971) and is surveyed by both Mankiw, Weinzierl, and Yagan (2009) and Diamond and Saez (2011), a literature that focuses on the simple setting of a perfectly competitive labour market despite growing evidence that wages are not generally equal to marginal product.[2] However, if wages deviate from marginal product, the laissez-faire 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 the pre-existing bargaining distortion and return labour supply to the efficient level.

In this paper, I consider optimal taxation in a general equilibrium setting that accommodates the real-world divergence between employees' wages and their marginal product. I focus on production that occurs in hierarchical teams, in which lower-skill workers match with higher-skill managers. This setting represents two essential features of real-world labour markets: 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 level 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 present a general model, and show that if wage bargaining causes the allocation to deviate from efficiency, a tax or subsidy can be used to correct workers' incentives, but that 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 only one distortion and one tax

[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).

instrument needed to fix it. However, if team size is increasing in manager effort and wages are below marginal product, then 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 a lower level of the firm hierarchy that they can supervise and exploit for rents; therefore, the manager's "wage" per unit of labour supply is too high regardless of the level of worker effort. As a result, 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 should tax the "job creators" *because* they want to create too many jobs at their firm.[5]

I then provide a general characterization of the efficient tax schedule, and I also consider optimal taxes from the perspective of a planner who cares about distribution,[6] using a perturbation method to derive the optimal tax rate at any point in the distribution as a function of a direct redistribution effect, the distortion effect on the marginal individual, and a new component measuring how taxes shift the wage distribution. This general characterization provides insight into how the presence of a non-competitive labour market interacts with the other economic forces that shape the optimal income tax schedule. However, while useful for highlighting the nature of the solution, the results of the general model are dependent on a number of quantities for which there is no clear empirical counterpart; a specific parametric model is required to implement the solution in a particular application, and thereby provide numerical results.

I therefore focus on a specific case of the general 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-

[3] A management literature presents an alternative story in which 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. Research by Alan Manning, among others, suggests the opposite conclusion.

[5] This intuition holds regardless of whether the added jobs would be filled by workers poached from other firms or by unemployed individuals.

[6] Throughout the paper, I will use the standard definitions of the terms "efficient" and "optimal": the former corresponds to the perfect-competition laissez-faire outcome, while the latter is the outcome which maximizes equally-weighted utilitarian social welfare. These outcomes are identical if utility is linear in consumption.

skill managers to produce according to the team's ability to overcome problems encountered in production.[7] I examine equilibrium outcomes under competitive wage-setting and a simple form of wage bargaining between workers and managers, and show that given an underlying skill distribution, wage bargaining generates a far more right-skewed income distribution, in which rents extracted from middle-income workers are captured by the highest-skill 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 is progressive over most of the income distribution with a top marginal rate of 50-60%. 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, and this result is not sensitive to the magnitude of the labour supply response to taxation. Finally, I evaluate the optimal tax schedule with diminishing marginal utility of income, using the perturbation method presented earlier. 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 to an S-shaped tax with declining but positive and significant marginal taxes at the top.[8]

This paper contributes to a small but growing literature on optimal taxation in non-competitive labour markets. Since Mirrlees (1971), the majority of the optimal income taxation literature has focussed 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 literature looking at taxation in the context of search and matching models began to develop a few decades later, 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

[7]Rothschild and Scheuer (2013) and Boadway and Sato (2014) also study taxation with different types of jobs, but in a different setting in which jobs are not hierarchical, but rather correspond to an extensive-margin decision between different occupations.

[8]The result of a declining optimal rate near the top of the income distribution is due to the assumption of a finite top to the distribution; in the usual Mirrleesian analysis with competitive labour markets, the optimal top tax rate is zero.

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 that lead to a positive optimal tax rate at the top of the distribution of workers.[9] However, inefficiency in the laissez-faire equilibrium is assumed away, and 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 therefore cannot say anything about income taxation at the upper end of the income distribution.

Two additional recent papers, meanwhile, highlight an important role for taxation in settings in which wages do not capture the social return to labour supply.[10] 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 bargaining over compensation 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), on the other hand, focus on the possibility that a few skilled 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.

In both of these papers, the results are driven by the fact that the impact of taxes on labour supply is dwarfed by the impact on another margin, either bargaining or other workers' production. My results demonstrate an entirely different argument for efficiency-improving non-linear taxes: even if managers respond to increased taxes *entirely* on the labour supply dimension, high marginal taxes on managers can improve efficiency if that labour supply is aimed at increasing team size. Additionally, my general results for efficient taxation are not sensitive to the magnitude of the labour supply response, in contrast to all previous findings

[9] Lehmann, Parmentier, and van der Linden (2011) extend the model to consider endogenous participation. Related papers also 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. Another study, Hungerbühler and Lehmann (2009), focusses on the role for a minimum wage in a search and matching framework.

[10] 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 (2013) examine a labour market with a separate "rent-seeking" sector, while Stantcheva (2014) considers optimal taxation with adverse selection.

in the literature; what matters is the magnitude of the deviation of wages from the efficient level.

Furthermore, although the existing literature recognizes the general principle that income taxation can have efficiency benefits in non-competitive labour markets, prior work has only focused on special cases or portions of the income distribution; no study that I am aware of considers such an efficiency role for taxation in a model of the entire income distribution, as in the standard Mirrleesian analysis. In such a setting, it is important to recognize that 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.

The rest of the paper is organized as follows. Section 2 presents the general model of team production, and characterizes efficient and optimal income taxes. Section 3 specifies the parametric model and describes the equilibrium under both competitive wage-setting and wage bargaining. Section 4 presents efficient taxes with wage bargaining, and section 5 contains estimates of optimal taxes with diminishing marginal utility in the case of both competitive and bargained wages. 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.

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. I show that non-zero taxes on managers can only be justified from an efficiency perspective if team size is increasing in managerial effort, and I characterize the shape of the efficient tax schedule. I conclude the section with a presentation of a perturbation method for calculating the optimal tax schedule.

2.1 General Model and Efficiency

I assume that the population consists of a continuum of individuals with skill levels z from some distribution $F(z)$ who match into 2-layer hierarchical teams, formed according to some general (and unspecified) matching mechanism. These teams consist of a measure-one manager with skill level z_m at the top of each team and some measure n of workers at the bottom, all of skill level z_p;[11] throughout the paper, I will use subscripts p and m to refer to workers (p for production) and managers respectively.

Output Y of a team depends on the labour supplies and skill levels of the manager and workers: $Y = Y(L_m, nL_p; z_m, z_p)$, where L represents labour supply. I allow for the possibility that the number of workers n may be an increasing function of manager 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. The utility function is defined over consumption C and labour supply L: $U = U(C, L)$, where $U_C > 0, U_L < 0, U_{CC} \leq 0$, and $U_{LL} < 0$.

I define the efficient allocation as the laissez-faire (no-tax) outcome under perfect competition, or equivalently as the allocation in which the marginal product of labour equals the marginal rate of substitution for all individuals. Throughout this section, I focus on one representative team (i.e. representative conditional on skill levels z_m and z_p). Solving for a worker's labour supply choice given a wage w, I find the condition that:

$$U_C w + U_L = 0.$$

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

$$U_C[Y_m + L_p n'(L_m)(Y_p - w)] + U_L = 0$$

where $Y_m \equiv \frac{\partial Y}{\partial L_m} > 0$ is the marginal product of managerial labour, and $Y_p \equiv \frac{\partial Y}{\partial nL_p} > 0$ is the marginal product of worker labour. By working harder, the manager not only receives their

[11] In the specific 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.

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 $Y_p > w$.

In a setting of perfect competition, however, the wage is equal to the marginal product of a unit of worker labour supply, so $w = Y_p$. Therefore, the efficient allocation is defined by:

$$Y_m = \frac{-U_{Lm}}{U_{Cm}} \tag{1}$$

$$Y_p = \frac{-U_{Lp}}{U_{Cp}}. \tag{2}$$

In general, for efficiency to be satisfied, it is clearly important for the worker's wage to be equal to marginal product, as this ensures that both the workers and the manager have the right incentives. 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; can they do so using marginal taxes $t_p(z_p)$ and $t_m(z_m)$?[12]

To answer this question, I solve for individual labour supply choices given marginal taxes, allowing for the possibility that bargained wages $w(t_m, t_p; z_m, z_p)$ depend on taxes. I find:

$$(1 - t_m)[Y_m + L_p n'(L_m)(Y_p - w)] = \frac{-U_{Lm}}{U_{Cm}} \tag{3}$$

$$(1 - t_p)w = \frac{-U_{Lp}}{U_{Cp}}. \tag{4}$$

If it is possible to choose a t_p to make $(1 - t_p)w = Y_p$ (that is, unless w decreases quickly as $1 - t_p$ increases), then the worker's labour supply will be restored to the efficient level as defined in (2). Then t_m must be set to ensure efficiency of the manager's labour supply decision: combining (1) and (3), I require $(1 - t_m)[Y_m + L_p n'(L_m)(Y_p - w)] = Y_m$.

[12] These marginal taxes are allowed to vary across the skill distribution, 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.

If firms are of fixed size, then $n'(L_m) = 0$ and the condition for efficiency simplifies to $(1 - t_m)Y_m = Y_m$, and so $t_m^* = 0$. Since w does not appear in Y_m, the latter depends 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. There is only one distortion, and so only one tax instrument is required to fix it.

If on the other hand $n'(L_m) > 0$, then the manager's level of labour supply depends on the wedge between the worker's wage and marginal product. If $w < Y_p$, then the manager earns positive rents from their workers, and even if worker labour supply is corrected using a subsidy, manager 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. This result is summarized in the following proposition.

Proposition 1. *With one-manager/n-worker teams, where the manager is residual claimant:*
(i) if n is fixed, then conditional on the worker's effort choice, 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 optimal 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 tells us that wage bargaining between workers and managers alone does not provide an efficiency argument for positive tax rates at the top of the income distribution. If wages are set below marginal product, 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. Efficient progressive taxes - in this context referring to positive marginal taxes among managers, who we may assume would tend to receive higher incomes - are justified 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) job-creation activities. 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 receive rents from individuals working below them in their firm should be taxed.

This result is related to the "Hosios condition" in the context of search and matching models: Hosios (1990) identified the condition on worker bargaining power that would need to be satisfied for the equilibrium to be efficient, 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.[13] However, this idea has until now only been applied to a setting of matching between workers and abstract "vacancies," with only an extensive margin; the idea of efficiency-enhancing marginal taxation has never been explored in the current framework, with a continuum of individual types, labour supply decisions, and individual managers. This paper therefore represents the first analysis of efficiency-enhancing taxation in a model that captures the entire income distribution, as in the standard Mirrleesian setup.

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 make a few extra assumptions, I can make some additional statements about the shape of the efficient tax schedule. First

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

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.[14] In the model presented in section 3, the equilibrium will necessarily take this form.

I also make a few standard assumptions about the production function: I assume diminishing marginal returns to both forms of labour, $Y_{pp} < 0$ and $Y_{mm} < 0$, and I assume complementarity of manager and worker labour supply, or $Y_{mp} > 0$. Then, I can consider the tax 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. If wages rise slowly with skill, $\frac{dw}{dz_p}$ will be small, and 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 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:

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

[14] 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. However, 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). Andrews, Gill, Schank, and Upward (2012) also find evidence in favour of positive assortative matching using German social security records.

Differentiating with respect to z_m, I obtain:

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

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 is also more likely to be satisfied if wages of workers increase slowly with skill; in that case, $\frac{d\rho}{dz_m}$ 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, and indicate that 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. 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.

2.3 Optimal Taxation Using Perturbation Method

Before proceeding to a parametric model, I now consider that a policy-maker is likely to not only be interested in the tax schedule that restores efficiency to the labour market; if the utility function exhibits diminishing marginal utility from income, a desire for redistribution will also play a role. Therefore, I consider a utilitarian planner searching for the optimal tax schedule, or the tax schedule that maximizes a social welfare function that adds up all individual utilities.

To solve for the optimal tax system, I will use a perturbation method, in which 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 consider a population of Q individual mass points, denoted by $q = \{1, ..., Q\}$, with mass $f(z_q)$ at skill levels $z_q = \{z_1, z_2, ..., z_Q\}$; a perfectly continuous case is the limit as $Q \to \infty$. 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;[15] given a tax schedule over income, individual labour supply choices will ensure that income is increasing in skill. To simplify notation, managers' "wages" $C'(L_m)$ will be denoted w just like those of production workers; it will then follow 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 $\Delta y_i = y_i - y_{i-1}$. I combine these effects in appendix A, and additionally make the standard assumption that there are no income effects,[16] 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.

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

$$\frac{dW}{dt_i} = \Delta y_i Q_i \left[E(U_{Cq}) - E(U_{Cq}|q \geq i)\right] - f(z_i)\frac{y_i}{(\gamma-1)}\frac{t_i}{1-t_i}E(U_{Cq}) + \sum_{q=1}^{Q} f(z_q)L_q\left[\frac{\gamma}{\gamma-1}t_q E(U_{Cq}) + (1-t_q)U_{Cq}\right]\frac{dw_q}{dt_i} \tag{5}$$

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

Proof. See appendix A. □

This equation can easily be understood as the sum of three effects. The first term in (5) 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

[15] 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.

[16] 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).

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.[17]

It is also easy to see how (5) generalizes to a continuous distribution: the only variables whose meaning depends on the number and spacing of mass points in the distribution are $f(z)$ and Δy_i. Therefore, if I multiply (5) 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 rather than a mass, and Δy_i is replaced by $\lim_{z_{i-1}\to z_i} \frac{y(z_i)-y(z_{i-1})}{z_i - z_{i-1}} = \frac{dy}{dz}$.

If I write $R = \Delta y_i Q_i \left[E(U_{Cq}) - E(U_{Cq}|q \geq i)\right]$ for the redistribution term and $S = \sum_{q=1}^{Q} f(z_q) L_q \left[\frac{\gamma}{\gamma-1} t_q E(U_{Cq}) + (1-t_q) U_{Cq}\right] \frac{dw_q}{dt_i}$ for the wage-shifting effect, at the optimal tax rate t_i it must be true that:

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

and rearranging, this gives:

$$t_i = \frac{(\gamma - 1)(R + S)}{f(z_i) y_i E(U_{Cq}) + (\gamma - 1)(R + S)}. \tag{6}$$

Equations (5) and (6) 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 $S = 0$, $Q_i = f(z_i)$, $E(U_{Cq}|q \geq i) = 0$, and the Pareto tail assumption which implies that $y_i = \frac{\alpha y_{i-1}}{\alpha - 1}$, my expression simplifies exactly to theirs: $t_i = \frac{1}{1+e\alpha}$, where $e \equiv \frac{1}{\gamma - 1}$ is the elasticity of taxable income and α is the Pareto parameter. Additionally, it is true that (5) and (6) can be applied in a wide variety of situations, beyond the team production setting studied in this paper. However, their practical applicability is limited by the fact that they require us to be able to measure

[17] A simple thought experiment shows why taxes must shift wages if they are not equal to marginal product, even in the fixed-team-size version of the model considered earlier: 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.

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 wages would make redistributional lump-sum taxes feasible. However, (5) and (6) 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 (5) to obtain the effect of changing t_i on social welfare. Therefore, a specific model of the labour market is required in order to say more, and the next section presents such a model.

3 Parametric Model of Hierarchical Teams

To move beyond 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.

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 a measure-one manager and measure n of workers. 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 the equilibrium must feature perfect positive assortative matching;[18] the proof of this result is discussed in appendix B.

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

[18]It is important to note that each agent represents an infinitesimally small space on the skill distribution, so the matching function will be continuous.

communication cost $h \in (0,1)$[19]: the manager must spend hL_p units of time on each problem that is forwarded to them,[20] and can solve problems with $d \leq z_m$. If the problem is solved, L_p units of output are produced by that worker, whereas workers with unsolved problems produce nothing. The manager therefore spends $hL_p(1-z_p)$ units of time in expectation on each worker, and given that the manager faces a continuum of workers of measure n, the manager faces no uncertainty and has a managerial time constraint of $nhL_p(1-z_p) = L_m$, and the team's total output is $nL_p z_m$.

Individuals receive utility from consumption C and disutility from labour supply according to a utility function $U(C,L)$. As stated earlier, a utility function with no income effect is common in the optimal income tax literature; 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, I will therefore 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. 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 his labour supply L_m and the skill level of worker z_p he wishes to hire, which must be consistent with the equilibrium matching function; the manager receives total consumption of $C(L_m) = nL_p(z_m - w(z_p)) = L_m \frac{z_m - w(z_p)}{h(1-z_p)}$, and thus sets $L_m = \left(\frac{z_m - w(z_p)}{h(1-z_p)}\right)^{\frac{1}{\gamma-1}}$, so that $r(z_m; z_p) \equiv C'(L_m) = \frac{z_m - w(z_p)}{h(1-z_p)}$, 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 $w(z)$. 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 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 $n = \frac{L_m}{hL_p(1-z_p)}$, this can be rewritten as:

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

[19]The upper bound on h is needed only for the proof of positive assortative matching in appendix B; in my calibrations I will never need to use a value larger than 1.

[20]Workers who supply more labour are assumed to work on more aspects of production, and thus the problem they face takes longer for the manager to study.

and differentiating with respect to z_p:

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

Therefore, the matching function is defined by:

$$m'(z) = \left[\frac{(h(1-z))^{\gamma} w(z)}{m(z)-w(z)}\right]^{\frac{1}{\gamma-1}} \frac{f(z)}{f(m(z))}. \tag{7}$$

Although this differential equation has no simple analytical solution, there is a fairly simple intuition behind it. $m'(z)$ is increasing in both h and $w(z)$: higher wages mean higher worker labour supply, and both that and higher communication costs mean a larger measure 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 measure 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 two subsections present the two alternatives that I consider.

3.2 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 $C = L_m \frac{z_m - w(z_p)}{h(1-z_p)}$ 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 (7) 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

[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.

worker or a manager, and which simplifies to $w(z^*) = \frac{z^* - w(0)}{h}$. 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.

I solve these equations numerically, using a uniform skill distribution for now; here and for the remainder of the numerical analysis in the body of the paper, I assume a compensated elasticity of taxable income equal to 0.25,[22] 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. For illustrative purposes, I assume a value of $h = 0.5$, in the middle of the permissible values;[23] when I calibrate h (along with the skill distribution) in later analysis, I end up with $h = 0.3850$ in the competitive case and $h = 0.9826$ in the bargaining case, suggesting that $h = 0.5$ is a reasonable compromise for this initial analysis, as my first goal is to compare outcomes from competition and bargaining in the same parameterized model.

The resulting wage and matching functions are displayed in Figure 1. z^* takes a value of about 0.8, so that individuals with z above that value become managers and individuals below z^* become workers. The wage function in (a) displays workers' wages $w(z)$ up to z^*, and then displays 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. One important characteristic of the matching function is that it flattens 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, the wage function exhibits 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

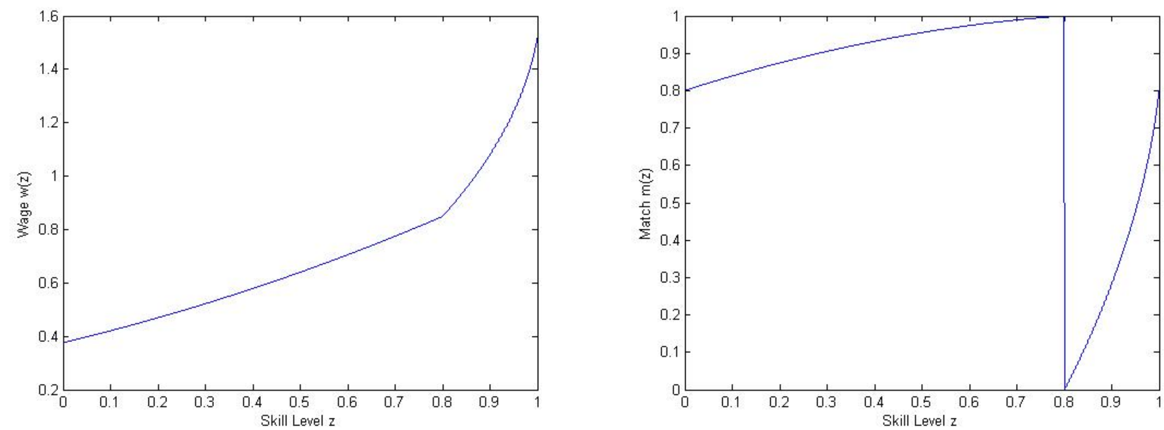
[22] An elasticity of 0.25 is suggested by Saez (2001), and Saez, Slemrod, and Giertz (2012) select it as the approximate midpoint of a range of plausible estimates from 0.12 to 0.4. I perform a sensitivity analysis in appendix E.2 using a value of 1 for the elasticity of taxable income, and show that the efficient taxes are nearly unaffected, while optimal taxes with diminishing marginal utility are still strongly raised at the top by wage bargaining.

[23] The maximum feasible h is about 0.916 with a uniform distribution and competitive wage-setting; above that values, communication is too costly to sustain a hierarchical equilbrium. With wage bargaining, the same constraint does not bind, but I never need to use a value above 1.

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 vice-versa for $z < z^*$.

Figure 1: Competitive Wage and Matching Functions with Uniform Distribution

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



3.3 Wage Bargaining

In this subsection, instead of perfectly competitive labour markets, I will consider a simple form of wage bargaining. Specifically, I assume fixed sharing of expected output, so $w = \beta z_m$, where β is set in equilibrium to clear the labour market. In the absence of taxes, this is also the outcome of a Nash bargain over the surplus; 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 to be extremely difficult.[24] However, an additional reason is that I 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 bargaining to taxation, 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.

[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 studied in section 4.1, 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.

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

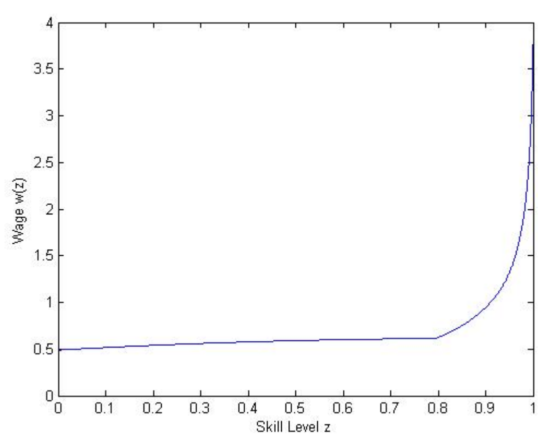
$$m'(z) = \left[\frac{\beta(h(1-z))^{\gamma}}{1-\beta}\right]^{\frac{1}{\gamma-1}} \frac{f(z)}{f(m(z))}. \tag{8}$$

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 $\beta = \frac{z^*}{h+z^*}$.

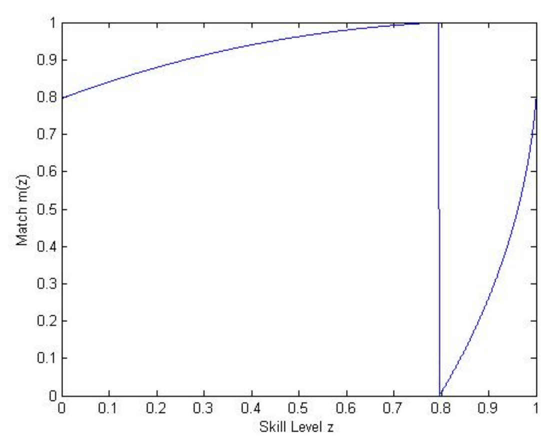
Once again, I can solve this equation numerically given a flat skill distribution and $h = 0.5$; the equilibrium value of β is 0.6142, and the resulting wage and matching functions are displayed in Figure 2. The matching function looks similar to the competitive case, but the wage function exhibits a much more dramatic rightward skew; because managers are able to extract rents from their workers, and higher-skill managers have larger teams working below them, the highest-skill managers can receive very large returns. Thus, a model with wage bargaining can more easily explain a long right tail to the income distribution than can a model with competitive wage-setting.

Figure 2: Wage and Matching Functions with Wage Bargaining and Uniform Distribution

(a) Wage Function $w(z)$



(b) Matching Function $m(z)$



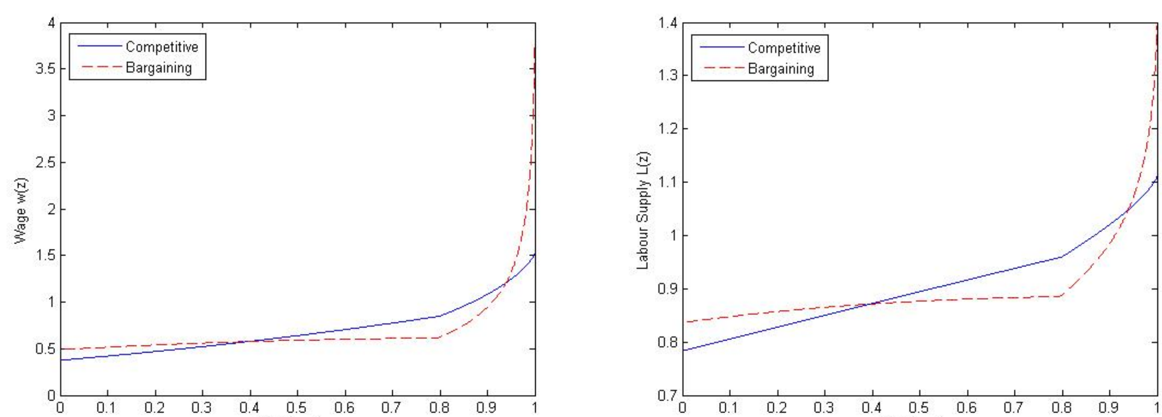
4 Efficiency and Taxation

Because the laissez-faire equilibrium in the competitive case is efficient, and wages in the bargaining case deviate from the competitive values, it is clear that the allocation with wage bargaining is not efficient. This happens because, with wages distorted from their laissez-faire competitive values, labour supplies are also distorted from the efficient values; some workers (those with wages that are too low) supply too little effort, and some supply too much, working beyond the point when their contribution to society equals the marginal utility cost from effort.

To illustrate this, panel (a) of Figure 3 overlays the equilibrium wages from the competitive and bargaining cases, while panel (b) does the same with labour supplies. Both pictures tell a similar story: middle-skill individuals are paid too little and thus work too little, while low-skill workers and especially the highest-skill managers receive excessive returns and work too hard. This pattern of inefficiency is not driven by the parameters chosen (recall that the output-sharing parameter β is endogenously determined in equilibrium), but rather by the following logic: within a team, if the manager is overpaid, the workers must necessarily be underpaid, so the pattern of inefficiency must necessarily be non-monotonic. Wages rise slowly with skill, and therefore the highest-skill workers near z^* are underpaid, and by indifference so are the lowest-skill managers. The partners of those groups - the lowest-skill workers and the highest-skill managers - are thus overpaid, especially the top managers, who are able to supervise large teams and thus accumulate very large rents from the individuals at the top of the set of workers.[25]

Figure 3: Wage and Matching Functions with Uniform Distribution

(a) Equilibrium Wages (b) Equilibrium Labour Supplies



This misallocation of labour can have significant efficiency consequences; in this case,

[25]While Nash bargaining leads to this outcome in all numerical cases that I have studied, alternative bargaining relationships could lead to different outcomes, including an inverted pattern of rents, or even an inverted matching pattern with the highest-skill managers matching with the lowest-skill workers. These outcomes seem less plausible, but they highlight the limitations on what can be said about efficiency in a context of wage bargaining in the absence of a fully-specified model.

if utility is 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. In such a setting, rather than being distortionary, marginal taxes could 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 will be attained. I consider this possibility in the following subsection.

4.1 Efficient Taxation with Wage Bargaining and Uniform Skill Distribution

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

$$m'(z) = \left[\frac{\beta(h(1-z))^{\gamma}}{1-\beta}\left(\frac{1-t(y(z))}{1-t(y(m(z)))}\right)\right]^{\frac{1}{\gamma-1}} \frac{f(z)}{f(m(z))}. \tag{9}$$

I assume, as is usual in the optimal taxation literature, that the government cannot observe skill levels of individuals; therefore, in practice, taxes can only be levied based on income. However, this 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 observations of 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.[26]

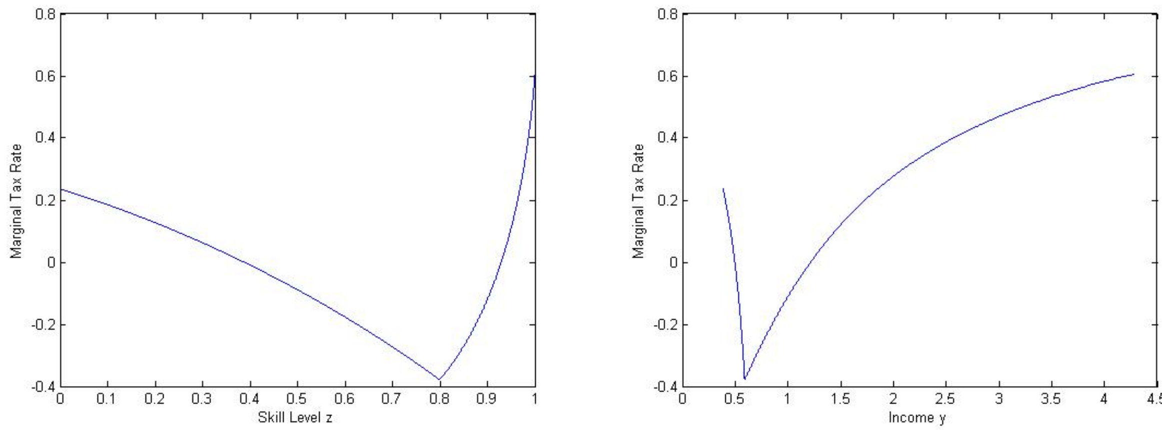
I therefore proceed to solve for the marginal taxes as a function of skill that restore efficiency to the labour market, and then check to see if they satisfy the necessary constraints. I use a simple procedure in which I iterate between choosing the marginal taxes that match labour supply to the competitive value at each point along the skill distribution, and re-solving for equilibrium at the new taxes. I continue to use a flat skill distribution, and

[26] 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).

calculate the optimal marginal tax rates as displayed in Figure 4, which do satisfy both the identification and incentive compatibility constraints.

Figure 4: Efficient Taxes with Wage Bargaining and Uniform Distribution

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



Panel (a) of Figure 4 presents optimal marginal taxes as a function of skill z, while panel (b) displays the optimal tax schedule as a function of income. The optimal taxes deviate from zero by a large amount, with small positive taxes at the bottom of the distribution, negative marginal rates in the middle, and rising tax rates at the top that reach as high as 60%. Since positive marginal taxes are used to offset excessive bargaining power, taxes are especially high at the top end of the income distribution, where high-skill managers extract considerable rents from their moderate-skill workers. This non-monotonicity was predicted in section 2.2, and follows from the pattern of distortions to wages and labour supply generated by wage bargaining that is illustrated in Figure 3: 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. At the cutoff z^*, individuals must be indifferent, so if the highest-skill managers are overpaid and need to be taxed, the workers just below z^* and, by indifference, the managers just above z^* should be subsidized. Finally, the latter implies that the lowest-skill workers are overpaid and should be taxed.

This demonstrates that non-zero marginal income taxes, and in particular taxes that are progressive over much of the range of the income distribution, can actually increase efficiency when wages are not set equal to marginal product; highly progressive taxes can be justified without any motive for redistribution. However, the calculations were carried out using an arbitrary parametrization of the model; next, I will demonstrate that similar results are found when h and the skill distribution are calibrated to the U.S. economy.

4.2 Efficient Taxes in Calibrated Model

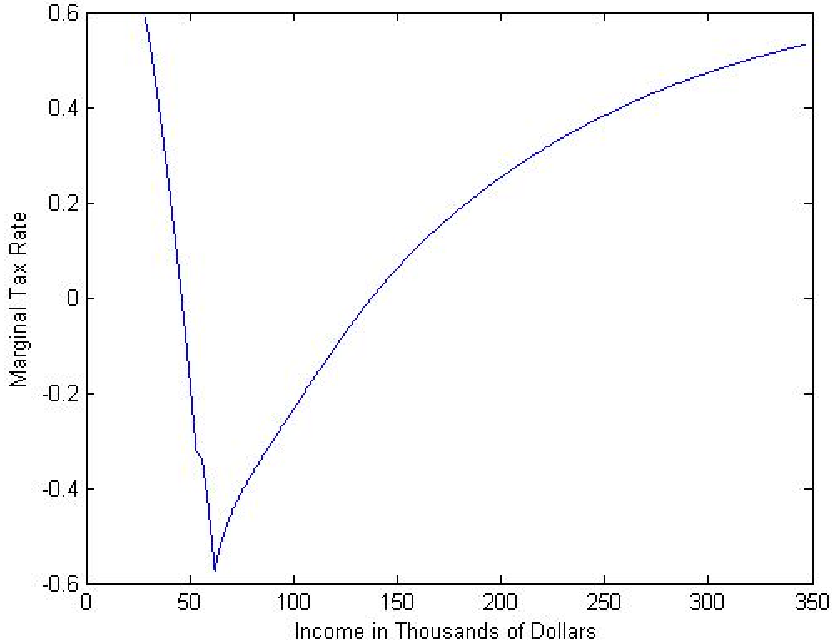
In this subsection, I present efficient taxes with a non-uniform skill distribution; specifically, I calibrate the wage bargaining case of my model to the income distribution measured by the 2012 March CPS, with both the CPS and model income distributions smoothed into a kernel density to allow for measurement error.[27] I assume that the baseline tax system is represented by an approximation to the U.S. income tax: specifically, I use the assumption first made in Jacquet, Lehmann, and van der Linden (2013) of 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 search over values of h and of the density at 11 nodes in the skill distribution, specifically $\{0, 0.1, ..., 1\}$, where the distribution is then defined as a cubic spline across those nodes; I find the set of these values that minimizes the sum of squares of a function defined as the difference between the resulting kernel income densities multiplied by income squared. The procedure of calibration is described in further detail in appendix C, where I also present the calibrated skill distribution; the resulting value of h is 0.9826.

I then solve for efficient taxes using the methods described in the previous subsection: I first solve for the competitive equilibrium with this skill distribution, and then find the tax rates that match the labour supply with wage bargaining to that obtained in the competitive laissez-faire equilibrium. However, the first-best is not incentive compatible in this setting, so the taxes are adjusted to ensure incentive compatibility according to a simple procedure described in appendix D.1. The result is displayed in Figure 5, as a function of real-world income in thousands of dollars. With the exception of a steeper drop in taxes at low incomes, from 0.6 to -0.6 rather than 0.2 to -0.4, the optimal tax schedule is quite similar to that presented in the flat-distribution case, and with quasi-linear utility the welfare gain from moving from a 27.9% flat tax to the optimum is a substantial 2.59% of mean consumption. Therefore, these results confirm the robustness of the results in the flat-distribution case: significant non-zero marginal taxes can improve efficiency, and the efficient tax schedule features high marginal tax rates at the top of the income distribution.

[27] I use a Gaussian kernel with a bandwidth of 0.3 times mean income.

Figure 5: Efficient Taxes with Wage Bargaining and Calibrated Distribution



5 Optimal Taxation with Diminishing Marginal Utility

I now focus on optimal taxation in a setting 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 skill distributions which have been calibrated to the U.S. economy; results with a uniform skill distribution can be found in appendix E.1.

I allow the government to choose a non-linear continuously differentiable tax schedule $T(y)$, for the purpose of financing lump-sum transfers and potentially a quantity of required government spending denoted by G. Therefore, the government's problem will be to choose the tax function to maximize average utility subject to their budget constraint:

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

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 (9) giving the matching function in the wage bargaining case, and the following equation for the competitive case:

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

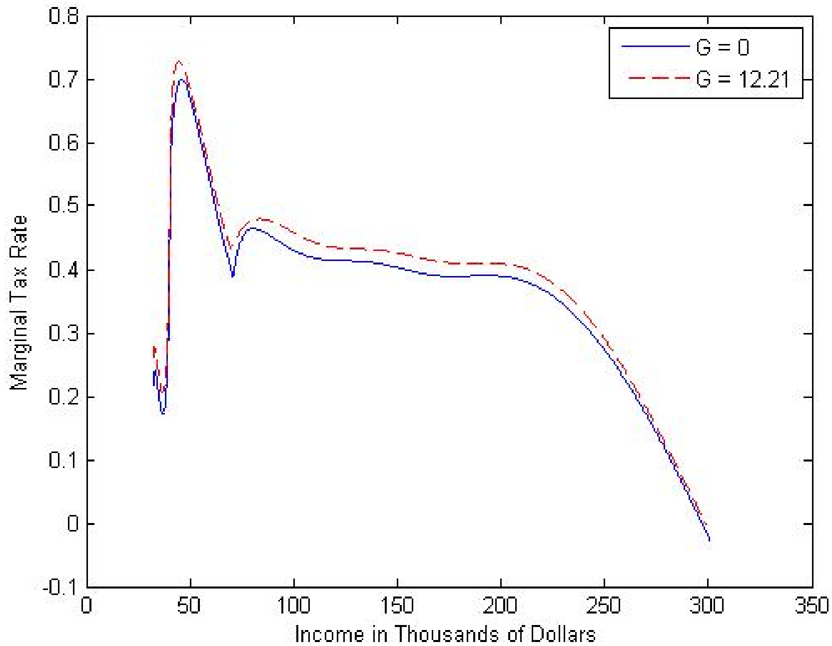
In the bargaining case, I use the same distribution and value of h as in section 4.2, and I use the same procedure for the competitive case to find $h = 0.3850$ and the skill distribution presented in Figure 11 in appendix C.

To solve for optimal taxes in this setting, I need to evaluate the welfare impact of changing taxes at each point in the distribution. However, there is a complication: not only do marginal tax rates depend on income, but in the current matching environment the income of any individual depends on marginal tax rates across the entire income distribution, since they all impact the matching equilibrium. Therefore, if a tax schedule is defined as a function of income, an iterative procedure between solutions for the income distribution and the marginal tax rates applied to each individual is required in order to solve for equilibrium and find out what tax rate each individual actually faces. In practice, this approach was computationally time-consuming, and I encountered severe difficulties in ensuring convergence. Therefore, I instead use the same basic approach as in section 4 to solve for optimal taxes $t(z)$ directly as a function of skill; by defining the tax according to skill, no iteration is required, as I know exactly what tax rate an individual of a given skill level faces, and I simply need to satisfy the constraints. A skill tax that satisfies the identification and incentive compatibility constraints can be implemented as an income tax, and since the set of income taxes is the implementable subset of skill taxes, the optimal implementable skill tax will also be the optimal income tax.

To look for the optimal skill tax in this case, I use the perturbation method introduced in section 2.3. I continue to consider a population of 10001 individual mass points at skill levels $\{0, 0.0001, ..., 1\}$. I use equations (5) and (6) to calculate the optimal tax schedule with my calibrated skill distributions: I find the optimal tax rate t_i for each individual subject to the identification and incentive-compatibility constraints. In practice, I use an iterative procedure and polynomial smoothing of the tax schedule, as described in further detail in appendix D.2.

The optimal taxes are presented for $G = 0$ and $G = 12.21$, the latter being the amount that balances the government budget given a baseline marginal tax of 27.9% and a minimum income of \$4024.90. The competitive results are presented in Figure 6; 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 1.07% and 1.99% of mean consumption, due to gains from redistribution.

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

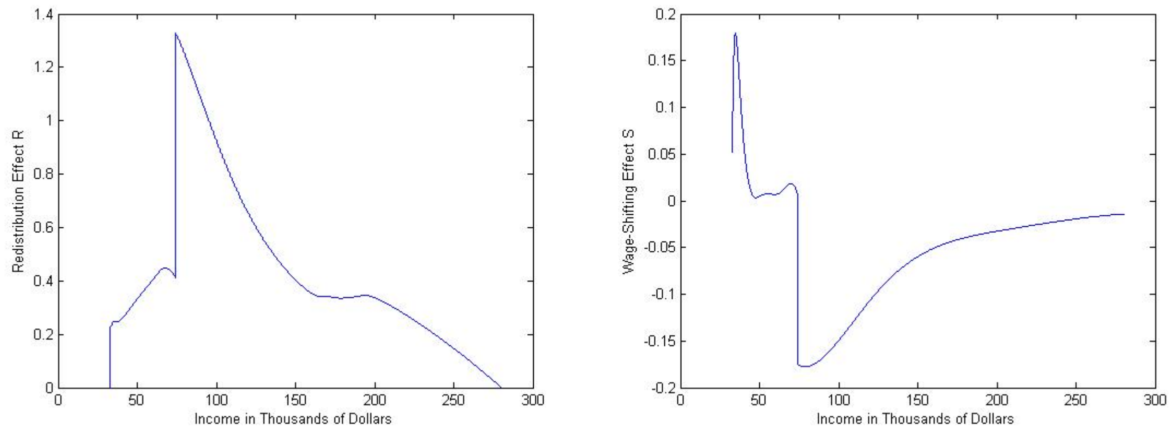


The reason for the inverted-U shape of the optimal tax schedule is simple: it is primarily driven by a redistribution effect that is generally in the shape of an inverted-U itself, as can be seen in Figure 7, which displays the values of R and S at baseline taxes in the $G = 12.21$ case (results are similar when $G = 0$). 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.

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

Figure 7: Values of R and S for Competitive Wages, Baseline Taxes and $G = 12.21$

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



incomes, optimal taxes are still generally V-shaped, but after rising to about 60% at a fairly high income, the optimal marginal rate declines to below 40% at the top. This decline at the top, however, may not be a robust finding, as it is dependent on the assumption of a finite top to the income distribution, which is a necessary component of my model; in typical competitive models with a finite top, a zero top tax rate will be optimal even with diminishing marginal utility of income, whereas a Pareto tail to the income distribution generates a positive optimal asymptotic top rate.[28] The resulting welfare gains from moving to the optimal tax are 0.74% and 1.28% of mean 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 below z^* (who occupy a very small space on the income distribution), but large and (aside from a spike just above z^*) 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.

The results in this section demonstrate that alternative models of wage-setting can lead to dramatically different results for optimal taxation. A competitive wage-setting environment leads to optimal taxes that are of a roughly inverted-U shape, whereas wage bargaining

[28]A lognormal income distribution also tends to lead to a zero optimal asymptotic top tax rate.

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

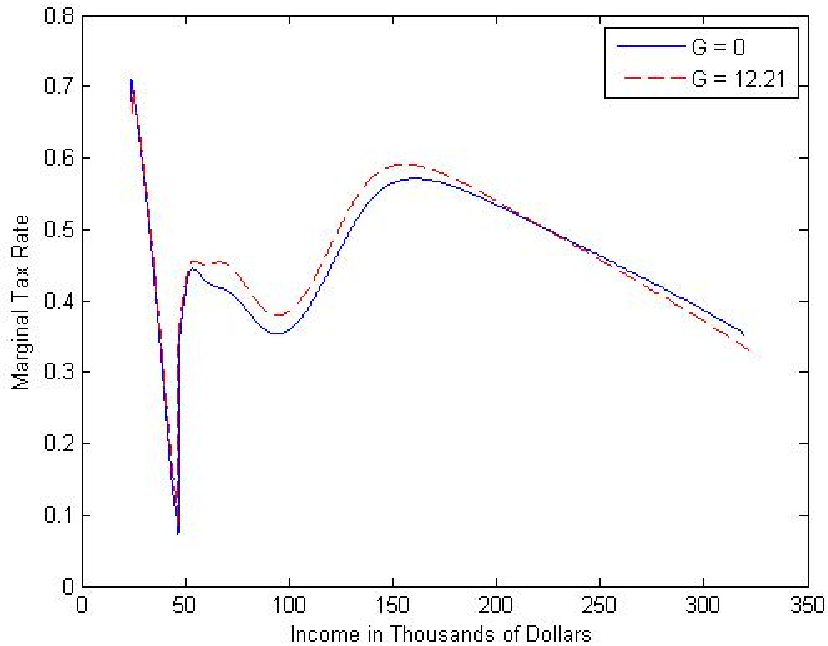
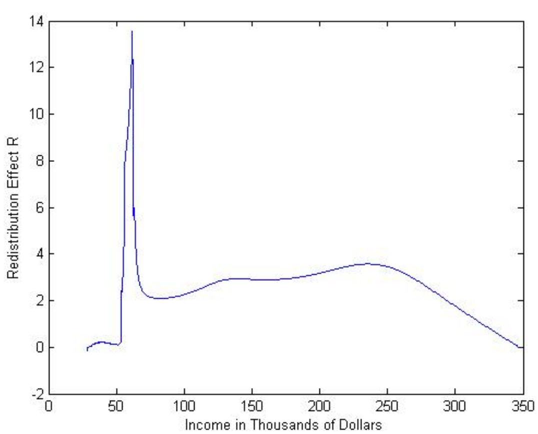
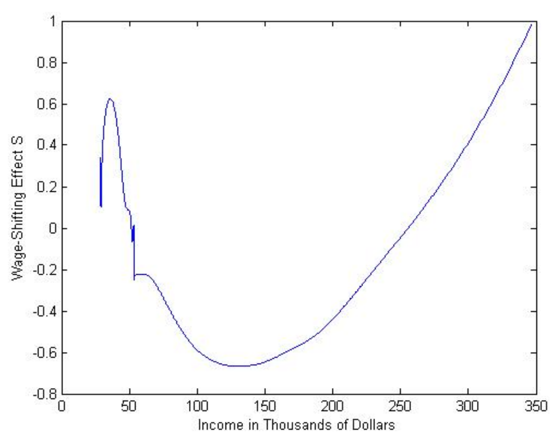


Figure 9: Values of R and S for Wage Bargaining, Baseline Taxes and $G = 12.21$

(a) Value of R



(b) Value of S



implies that taxes should be rising over much of the distribution and positive at the top, even though I assume a finite income distribution. In appendix E.1, I present results with a uniform skill distribution, and find very similar results, with even higher 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. Meanwhile, 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. Therefore, in appendix E.2, I redo the analysis with a considerably higher elasticity of taxable income of 1, and I find that my conclusions are largely unaltered: efficient taxes may actually be larger at the top end, and while the optimal taxes decline, the drop in the top tax rates is larger in the competitive case, leaving the qualitative conclusions about the impact of wage bargaining unaffected.

Finally, in appendix E.3, I present optimal bracketed taxes, with thresholds set at approximately the levels facing a single taxpayer in the US, 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 the tax rates rise over the top brackets in the wage bargaining case, and decline with competitive wages.

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 too much effort in trying to accumulate workers and the rents that come with them.

I also characterize efficient and optimal taxes using the general model, and then I 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, marginal taxes that deviate significantly from zero can play an important role in improving efficiency, and I show that the efficient taxes are progressive over most of the income distribution, reaching 50-60% at the top. Finally, I apply an optimal income tax analysis to the model, and show that wage bargaining dramatically alters the optimal tax schedule to feature high tax rates near the top 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 consider 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.

A 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 my general model from section 2; I will separately consider the impact on individuals $\{1, ..., i-1\}$ and on $\{i, ..., Q\}$.

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 also 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:

$$\frac{dW}{dt_i} = \sum_{q=1}^{Q} U_{Cq}f(z_q)\left[\frac{db}{dt_i} + (1-t_q)L_q\frac{dw_q}{dt_i}\right] - \sum_{q=i}^{Q} U_{Cq}f(z_q)\Delta y_i. \tag{10}$$

Finally, I need to solve for $\frac{db}{dt_i}$. If X is used to denote the total tax revenues collected by the government, $\frac{db}{dt_i} = \frac{\frac{dX}{dt_i}}{\sum_{q=1}^{Q} f(z_q)}$, and $\frac{dX}{dt_i}$ can be written as:

$$\frac{dX}{dt_i} = \sum_{q=1}^{Q} f(z_q)t_q\frac{dy_q}{dt_i} + \Delta y_i\sum_{q=i}^{Q} f(z_q) \tag{11}$$

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

$$\frac{dy_q}{dt_i} = L_q\frac{dw_q}{dt_i} + w_q\frac{dL_q}{dt_i}. \tag{12}$$

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 (12), 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 (11), I obtain the following equation for $\frac{dX}{dt_i}$:

$$\frac{dX}{dt_i} = \sum_{q=1}^{Q} f(z_q)t_q\left[\left(L_q + (1-t_q)\frac{w_q}{s_{Lq}}\right)\frac{dw_q}{dt_i} - \frac{w_q s_{Cq}}{s_{Lq}}\frac{dC_q}{dt_i}\right] + \Delta y_i\sum_{q=i}^{Q} f(z_q) - f(z_i)t_i\frac{w_i^2}{s_{Li}}. \tag{13}$$

If I define $\sum_{q=i}^{Q} f(z_q) \equiv Q_i$, (10) and (13) can be combined to give:

$$\frac{dW}{dt_i} = \Delta y_i Q_i\left[E(U_{Cq}) - E(U_{Cq}|q \geq i)\right] - f(z_i)t_i\frac{w_i^2}{s_{Li}}E(U_{Cq})$$

$$+\sum_{q=1}^{Q} f(z_q)\left[t_q\left(L_q + (1-t_q)\frac{w_q}{s_{Lq}}\right)E(U_{Cq}) + (1-t_q)L_qU_{Cq}\right]\frac{dw_q}{dt_i}$$

$$-\sum_{q=1}^{Q} f(z_q)t_q\frac{w_q s_{Cq}}{s_{Lq}}E(U_{Cq})\frac{dC_q}{dt_i}.$$

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:

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

This equation is the result stated in Proposition 2.

B Proof of Perfect Positive Assortative Matching

In this appendix, I will describe the proof that the equilibrium of my model from section 3 features 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 within a team monotonically increasing in worker skill. I will start with competitive wage-setting, and then present the proof with wage bargaining.

In the competitive labour market, much of the proof is identical to that in Antràs, Garicano, and Rossi-Hansberg (2005), so I will only discuss my deviations from that proof, and readers are directed to that paper for the details. 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. There is only one small modification to the proof in appendix B of Antràs, Garicano, and Rossi-Hansberg (2005); for a manager who hires I different types of workers, the maximization problem is described by:

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

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

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

The labour supply terms L_i do not appear in the original proof, but they cancel out, and then the first-order conditions are exactly identical to those in Antràs, Garicano, and Rossi-Hansberg (2005), and the proof proceeds as in that paper. Therefore, I can use a matching function $z_m = m(z_p)$ to describe the matching equilibrium.

Next, I must prove that $m(z)$ exhibits perfect positive assortative matching; that is, that $m'(z) > 0$ and the equilibrium is hierarchical with a single cutoff z^*. Again, with one minor modification, the proof of Theorem 1 in Antràs, Garicano, and Rossi-Hansberg (2005) applies: first, substitute the manager's "wage" $r(z_m; z_p)$ for $\Pi(z_m, z_p)$, and the proof of $m'(z) > 0$ is exactly the same. The rest of the proof of Theorem 1 in Antràs, Garicano, and Rossi-Hansberg (2005) also goes through unchanged (though the solution derived there for $m(z)$ is altered in my analysis).

Therefore, I can conclude that the matching function features perfect positive assortative matching in competitive equilibrium.

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; 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 $h(1-z_p)$ 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; in the interior of the set of workers, the matching function follows equation (8). 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; those workers would be strictly better off matching amongst themselves), and managers at the top (otherwise, the top individual could hire someone with $z = 1-\epsilon$ and receive arbitrarily large rents as ϵ goes to zero), so 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:

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

The first equation requires that individuals at z_1 are indifferent between being workers and managers, and can be written simply as $\beta z_2 = \frac{(1-\beta)z_1}{h}$.[29] The two inequalities ensure that individuals marginally above and below the relevant cutoffs become workers or managers as required, and these will be evaluated assuming that the equality is true.

First, we can observe that the first inequality is always satisfied:

$$\begin{aligned} \lim_{z\uparrow z_1} w'_{01}(z) &= \left[\frac{(\beta h(1-z_1))^{\gamma}}{1-\beta}\right]^{\frac{1}{\gamma-1}} \\ &= (1-\beta)\left[\frac{z_1(1-z_1)}{z_2}\right]^{\frac{\gamma}{\gamma-1}} \\ &< (1-\beta) < \frac{1-\beta}{h} = \lim_{z\downarrow z_1} r'_{12}(z). \end{aligned}$$

[29] A similar indifference condition must be satisfied at z_2, but this condition is not relevant to the current proof.

Therefore, the second inequality needs to be disproved. I proceed as follows:

$$\begin{aligned}\lim_{z \downarrow z_2} w'_{23}(z) &= \left[\frac{(\beta h(1-z_2))^{\gamma}}{1-\beta}\right]^{\frac{1}{\gamma-1}} \\ &= (1-\beta)\left[\frac{z_1(1-z_2)}{z_2}\right]^{\frac{\gamma}{\gamma-1}} \\ &< (1-\beta) < \frac{1-\beta}{h(1-z_1)} = \lim_{z \uparrow z_2} r'_{12}(z).\end{aligned}$$

This disproves the second inequality, and proves that 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.

C Calibration of Skill Distributions

To calibrate the skill distribution in both the wage bargaining and competitive cases, I start at $h = 0.5$ and a flat distribution, and then test small changes in h and 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 h and the spline nodes in the direction that reduces an objective function.[30] In choosing the objective function, there is a conflict between matching the weight given to the center of the distribution and generating the long right tail to the income distribution that is observed empirically. The simple model presented in section 3 is unable to generate a right tail as long as that in the data, even in the case of the wage bargaining framework; to more accurately match the real-world income distribution, a model with more than two team layers would likely be required. However, as a compromise, I consider the differences between the resulting kernel densities multiplied by income squared; I choose h and the values of the distribution nodes that minimize the sum of squares of this function.

In the wage bargaining case, the resulting value of h is 0.9826, and the skill distribution is displayed in Figure 10. Meanwhile, in the competitive case, the calibration produces $h = 0.3850$ and the skill distribution in Figure 11.

D Solution Methods

In this appendix, I describe the method used to solve for efficient taxes with the calibrated distribution, and the more complex method used to solve for optimal taxes in all settings.

D.1 Solution Method for Efficient Taxes with Calibrated Distribution

In the case with a calibrated skill distribution and wage bargaining 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

[30] To prevent very low densities, which could cause my matching algorithm to function poorly, I impose a minimum value of $f(z) = 0.2$ for the distribution.

Figure 10: Calibrated Skill Density $f(z)$ with Wage Bargaining

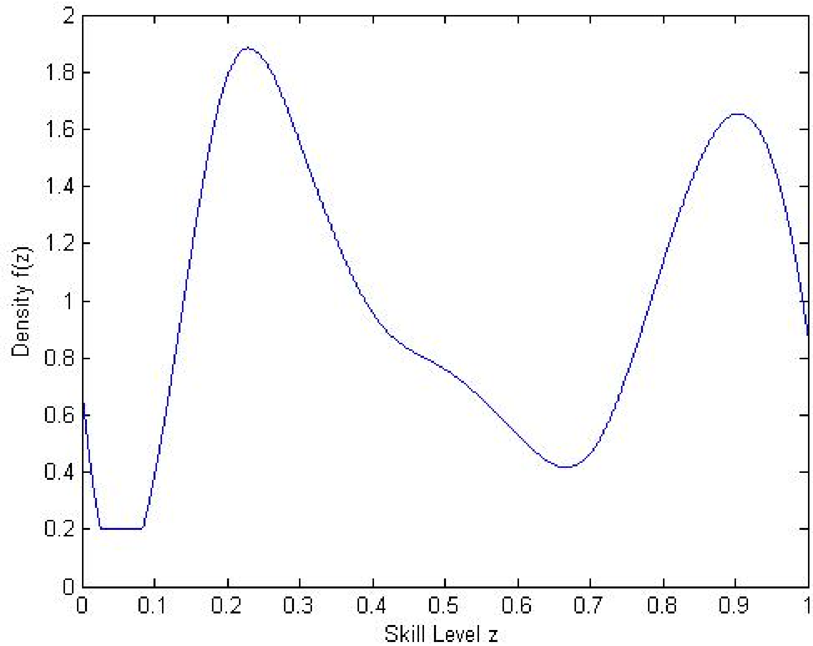
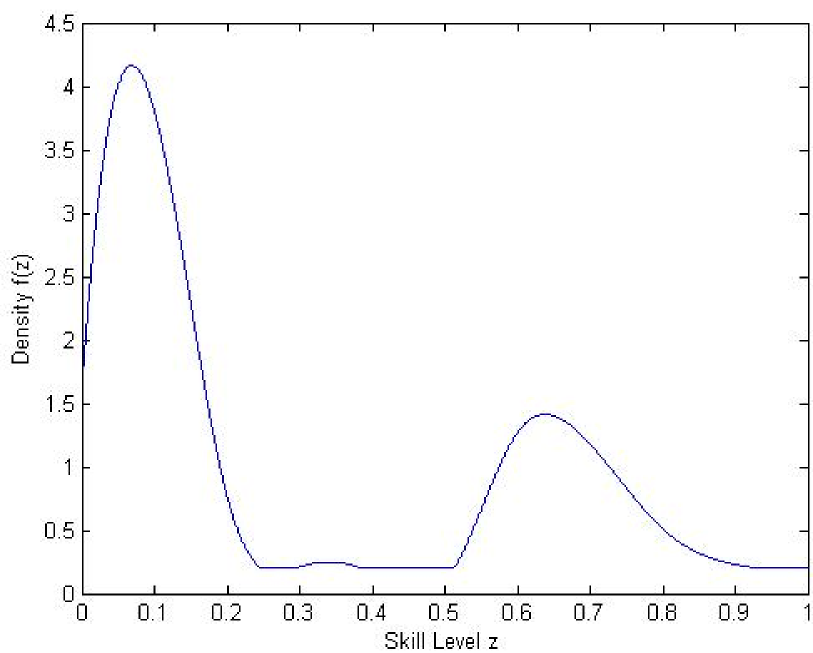


Figure 11: Calibrated Skill Density $f(z)$ with Competitive Wages



boundaries imposed by the constraints, starting at $0.625z^*$ and moving both right and left from there; the choice of $0.625z^*$ makes the integral of the adjustment to the tax schedule close to zero.

D.2 Solution Method for Optimal Taxes

To solve for the optimal tax schedule, I use equation (6) following an iterative procedure. For one round of iteration, 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; then I re-solve for the labour market equilibrium. 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.

I perform some number q of rounds at one time, and only re-solve the S term after all q rounds in order to save time; 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 (in the wage bargaining case, a cubic spline is used for the approximation above z^* as a polynomial does not adequately capture the dramatic increase in S at the top). At the end of each block of q rounds, I also smooth the tax schedule using a polynomial best fit, choosing 13th-order polynomials on each side of z^*, as otherwise S becomes unstable and poorly behaved; then I go through the tax rates and adjust them as needed to ensure that they fit inside the bounds imposed by the constraints.[31] I allow q to decline over time as the tax schedule converges, and the process concludes when q reaches 6 and the squared sum of shifts in the tax schedule drops below 0.005 (which corresponds to a shift of about 0.0007 per individual).

E Additional Numerical Results

I now present a series of additional numerical results. Subsections E.1 and E.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, and section E.3 presents optimal bracketed taxes.

E.1 Optimal Taxes with Uniform Distribution

In this appendix, I present the optimal taxes with a uniform skill distribution and $h = 0.5$. I use the same procedure as in section 5 to solve for the optimal taxes in both the competitive and wage bargaining frameworks, for each of $G = 0$ and $G = 0.12$; the latter is a bit less than 20% of average income prior to taxes in both the competitive and bargaining frameworks, which is meant as a rough approximation to total tax revenues as a percent of GDP in the U.S. The results for the competitive case can be found in Figure 12, where we can observe that, as in the main analysis, the optimal marginal tax rates roughly follow an inverted-U shape. As in the calibrated analysis, this is largely due to a redistribution effect in the shape of an inverted-U, as shown in Figure 13, which displays the values of R and S at zero taxes in the $G = 0$ case (results are very similar when $G = 0.12$). Moving from zero marginal taxes to the optimal taxes generates welfare gains that are equivalent to 2.95% and 5.50% of mean consumption respectively in the two cases, due to gains from redistribution.

The optimal marginal taxes for the wage bargaining case are presented in Figure 14. The results are generally similar to those in Figure 8: V-shaped at the bottom, then rising above 60%

[31] In the competitive analysis, I start the adjustment at z^* and then move left and then right from there. In the main wage bargaining analysis, I start at $0.625z^*$, whereas I start at $0.55z^*$ with ETI = 1.

Figure 12: Optimal Tax Schedule with Uniform $F(z)$, Log Utility and Competitive Wages

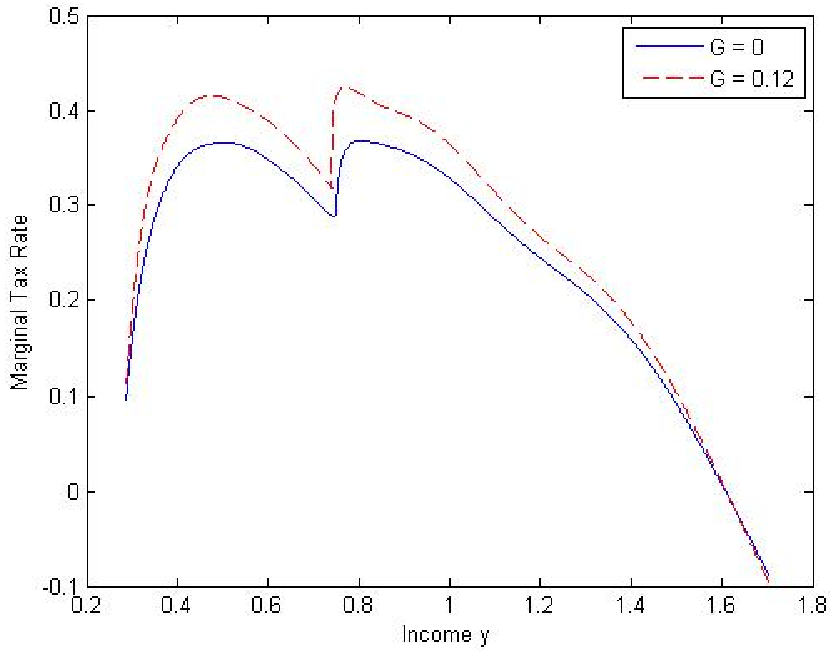
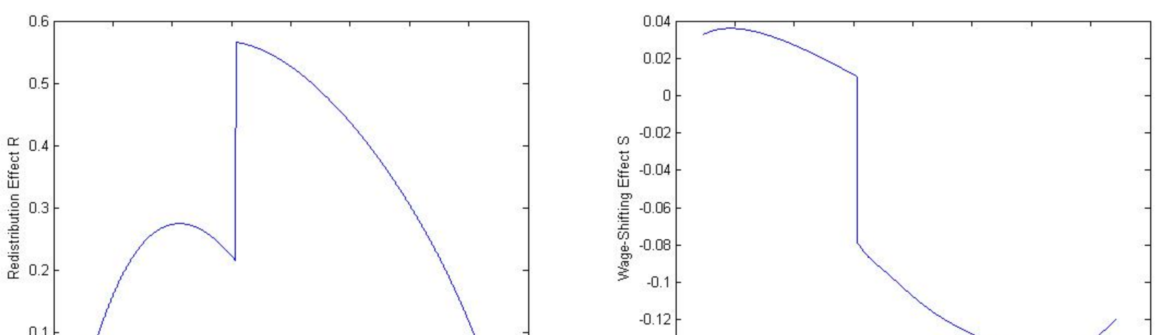


Figure 13: Values of R and S for Uniform $F(z)$, Competitive Wages, Zero Taxes and $G = 0$

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



before declining to about 40% at the top. The values of R and S are displayed in Figure 15, and tell a broadly similar story to those in Figure 9 earlier. The welfare gains from shifting from zero marginal taxes to the optimum are 4.44% and 6.30% of mean consumption in the two cases.

The results in this appendix demonstrate that the findings in section 5 with the calibrated skill distribution are robust to the assumption of a uniform distribution.

Figure 14: Optimal Tax Schedule with Uniform $F(z)$, Log Utility and Wage Bargaining

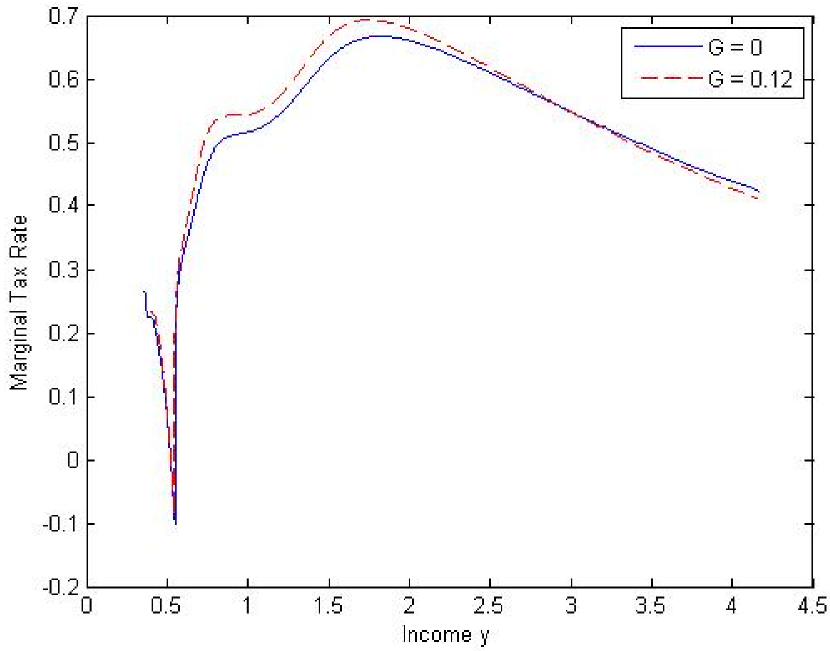
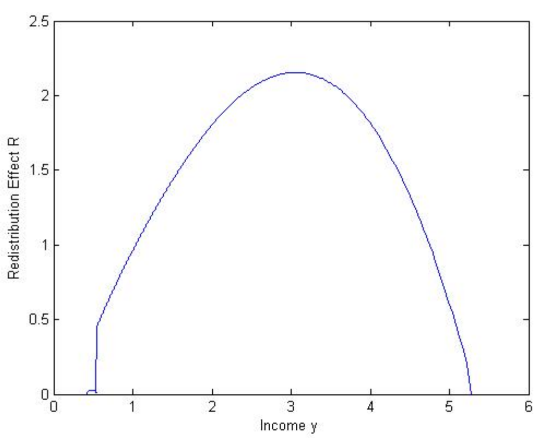
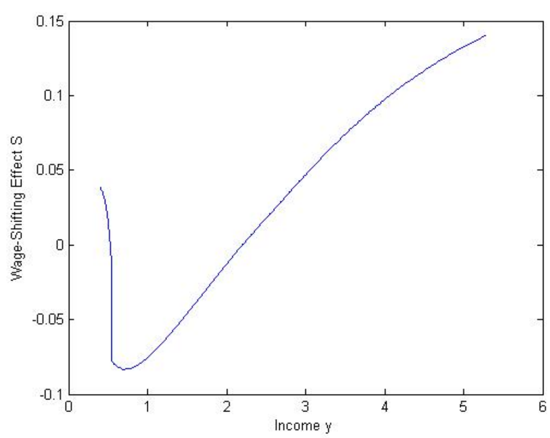


Figure 15: Values of R and S for Uniform $F(z)$, Wage Bargaining, Zero Taxes and $G = 0$

(a) Value of R



(b) Value of S



E.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).[32] A dramatic example is

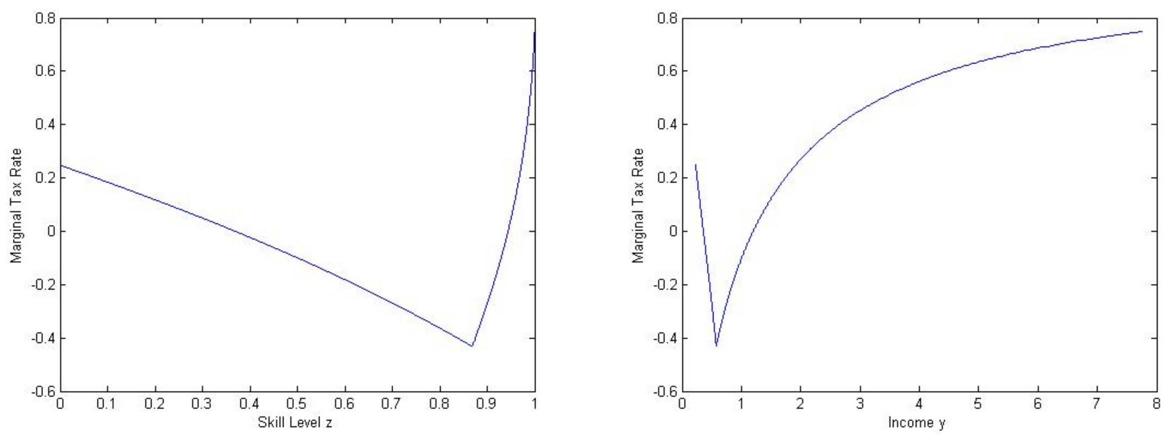
[32]Furthermore, some research has indicated that the ETI may not be a constant, with Gruber and Saez (2002) finding larger values at higher income levels, and Keane and Rogerson (2012) highlight the difference

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 begin with the flat-skill-distribution analysis. Figure 16 presents the efficient taxes with quasi-linear utility and wage bargaining, and a comparison of this figure with Figure 4 demonstrates that the conclusion about efficient taxes with wage bargaining is unaffected by a higher value of ETI; in fact, the top tax rate now increases to above 70%. The welfare gains in this case increase to 3.54% of mean consumption, as compared to 0.68% with ETI = 0.25.

Figure 16: Efficient Taxes with ETI = 1, Wage Bargaining and Uniform Distribution

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



Next, I calibrate h and the skill distribution to the U.S. income distribution using the same procedure as before. With a lower value of γ, it is easier for both models to generate a long right tail to the income distribution. The efficient taxes with wage bargaining can be found in Figure 17. As before, the overall shape of the tax schedule is similar to the ETI = 0.25 case, and the optimal top tax again is higher than before; the welfare gain from implementing the efficient tax schedule is a very large 8.36% of mean consumption.

Finally, the optimal taxes with log utility are in Figures 18 and 19. The level of the optimal tax with wage bargaining is somewhat lower than with ETI = 0.25, but the drop is much smaller than with competitive wages, and the optimal tax at the top remains above 30%, with a peak of over 40%. The welfare gains are 1.57% and 1.60% with competitive wages, and 2.05% and 1.98% with wage bargaining, which are larger than when ETI = 0.25 in every case but one.

This appendix has demonstrated that the case for efficient progressive taxes at medium-to-high incomes 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 increase substantially 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.

between the standard micro-level estimates and a long-run macro labour supply elasticity.

Figure 17: Efficient Taxes with ETI = 1, Wage Bargaining and Calibrated Distribution

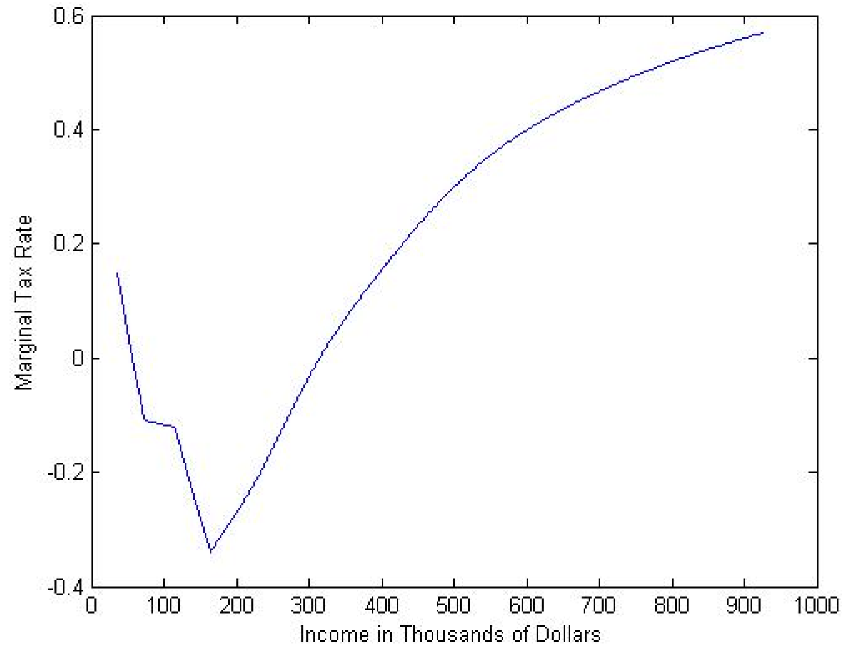
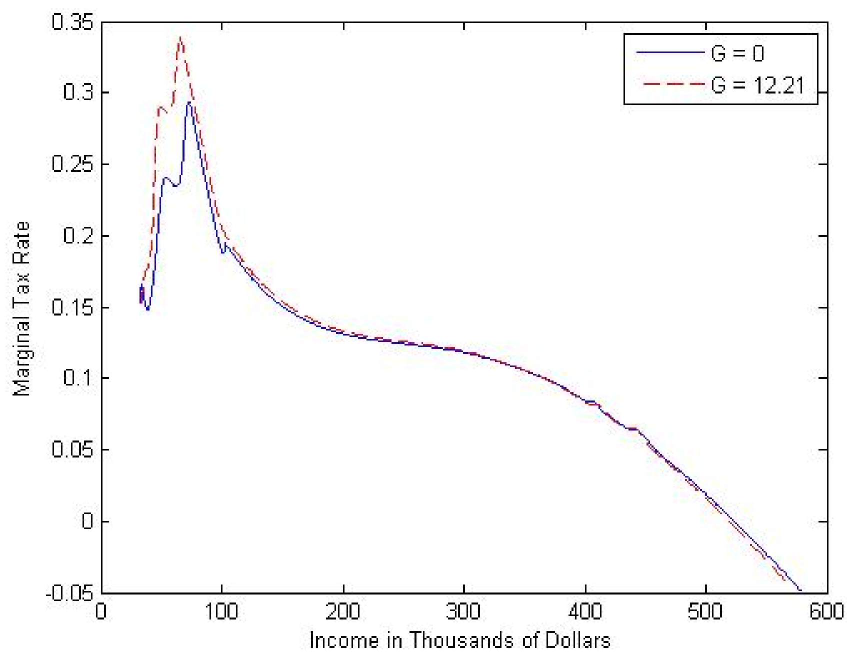


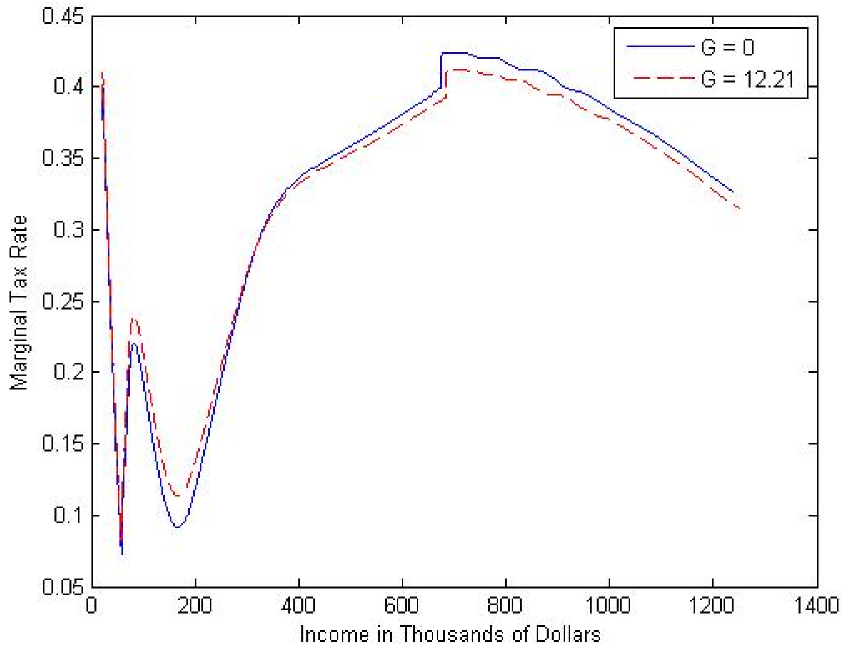
Figure 18: Optimal Tax Schedule with ETI = 1, Log Utility and Competitive Wages



E.3 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

Figure 19: Optimal Tax Schedule with ETI = 1, Log Utility and Wage Bargaining



at values close to the existing tax brackets facing single taxpayers in the US (counting the personal exemption and standard deduction): less than \$50k, \$50-100k, \$100-200k, \$200-400k, and (where necessary) \$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 Table 1, for $G = 12.21$ (though results are quite similar with $G = 0$), and for each of 4 cases: competitive and wage bargaining in both the baseline calibrated setting and the calibration with an ETI of one. 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 in the shape of an inverted U, whereas they are rising over most of the distribution with wage bargaining. The final column in the table presents the welfare gains as a percentage of mean consumption as before, and in brackets I also list the fraction of the welfare gain from the optimal non-linear tax that is attainable with a bracketed tax, a value that is at least 50% in each case.

Table 1: Optimal Bracketed Taxes

	\$0-50k	\$50-100k	\$100-200k	\$200-400k	\$400k+	Welfare Gain
Competitive (Baseline)	31.18%	47.51%	43.80%	37.53%	-	1.08% (54.4%)
Bargaining (Baseline)	38.08%	38.10%	47.01%	49.50%	-	0.84% (65.5%)
Competitive (ETI = 1)	17.24%	23.30%	16.27%	14.01%	4.72%	1.27% (79.1%)
Bargaining (ETI = 1)	28.78%	14.02%	9.35%	25.27%	42.36%	1.77% (89.0%)

References

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

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*, forthcoming.

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.

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., AND E. LEHMANN (2009): "On the Optimality of a Minimum Wage: New Insights from Optimal Tax Theory," *Journal of Public Economics*, 93(3-4), 464–481.

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.

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

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.

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*, forthcoming.

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