Taxation, Redistribution and Frictional Labor Supply[*]

Christopher Sleet[†]

Tepper School of Business, Carnegie Mellon University

Hakkı Yazıcı[‡]

Sabancı Üniversitesi

This Version: April 11, 2017

Abstract

We analyze the implications of ex ante dispersion in worker talents and a frictional labor market for the design of tax systems. Our model features on and off the job search, job ladders and equilibrium income and profit dispersion within talent markets. In a baseline setting with no talent dispersion, the optimal system consists of an unemployment benefit financed out of a simple lump sum tax on workers. The benefit is high enough to suppress worker income and firm profit dispersion, deter worker poaching and collapse job ladders. With talent dispersion, high benefit levels drive less talented workers out of the market and are prohibitively costly. Active talent markets are frictional. Taxes impact the dispersion of worker incomes and firm profits within these markets. These effects shape and modify conventional optimal tax formulas. Quantitative analysis implies that they create a motive for lower marginal tax rates. *(JEL D31, H21, H24, J60, J62, J65)*

[*]We thank seminar participants at Carnegie Mellon, Harvard and Rice Universities and at CERGE-EI. We especially thank Laurence Ales, David Childers, Nathan Hendren, Marek Kapička, Ali Shourideh, Ctirad Slavík and Stefanie Stantcheva for helpful comments and suggestions.

[†]csleet@andrew.cmu.edu. Tepper School of Business, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA.

[‡]yazici.hakki@gmail.com. Sabancı Üniversitesi, Faculty of Arts and Sciences, Orhanli, Tuzla, 34956 Istanbul, Turkey.

1 Introduction

The canonical optimal tax model assumes a frictionless labor market in which all worker income variation is attributable to variations in talent and intensive margin labor supply choices. This paper advances the optimal tax literature by deriving implications of off and on the job search frictions for tax design. The opportunity to search and (re)match with a firm that pays more and extracts less creates frictional wage dispersion and endogenous "job ladders" for equally talented workers. We show that tax policy reshapes such intra-talent market job ladders and, hence, impacts worker income and firm profit distributions. These effects supplement the effect of standard tax distortions on the intensive effort and extensive participation margins. In so doing they introduce new and readily interpretable terms into optimal tax equations. In addition, search frictions modify the mapping from the underlying talent distribution to the observed income distribution. Quantitative analysis that takes this into account implies that search frictions reduce optimal marginal tax rates relative to a frictionless benchmark.

Our model embeds Burdett and Mortensen (1998) type frictional labor markets into an otherwise standard public finance framework featuring heterogeneous worker talent, an intensive labor supply margin and taxes. In the model, workers search on and off the job. They experience "market luck" unrelated to their talent: job destruction shocks that consign them temporarily to unemployment and randomly arriving opportunities to climb job ladders and find less extractive employers.[1] The policymaker is assumed to observe (and tag) the current employment status (working or not working) and the income of the worker, but nothing else. She is, thus, constrained to select a tax function applied to the incomes of those who are and a benefit paid to those who are not in work. The policymaker observes firm profit and in our benchmark version (optimally) taxes it at 100%.[2] Within each talent market, tax and benefit policy determines a minimal utility that firms must deliver to workers and a corresponding maximal profit-per-worker level that they can extract. In choosing a profit-per-worker offer below this maximum, firms must trade off profit extracted from each worker against the number of workers attracted and retained. Some firms make high profit-per-worker offers, but attract and re-

[1] This combination of assumptions enriches the standard treatment of the labor market in public finance models and aligns with both recent microeconometric work that incorporates on-the-job search and macroeconomic work that stresses the role of on-the-job search in reconciling observed wage dispersion with the short duration of job search and unemployment.

[2] We assume that the social marginal value of public funds exceeds that of profits if left in private hands. Since the former can be redistributed as a lump sum transfer, it equals the average marginal utility of workers.

tain a small number of (lower earning) workers. Others make low profit-per-worker offers, but attract and retain many (higher earning) workers. In this way within talent market job ladders and frictional wage dispersion emerge.

If all workers have the same talent, the policymaker can collapse job ladders completely and, hence, purge the economy of frictional wage dispersion. This is achieved by driving up unemployment benefit to such a level that all firms make zero pre-tax profit on any job posting. Under this regime marginal income taxes are zero and there is no distortion to production: benefits are paid for out of a lump sum tax on identical workers. Thus, search frictions per se do not provide a rationale for distortionary redistributive taxation. The situation is very different when talent heterogeneity is introduced and combined with search frictions. Driving up unemployment benefit to collapse frictional wage dispersion in one talent market now means shutting down lower talent markets altogether. The cost of doing this ensures the survival of job ladders and wage dispersion in active talent markets at the optimum.

In the setting with search frictions and talent variation, marginal tax policy distorts both worker effort decisions in active talent markets and the threshold defining the marginally active talent. In addition to these intensive and extensive margin impacts, tax policy affects the equilibrium shape of job ladders and, hence, frictional wage dispersion in active markets. Specifically, to attract workers in the face of higher taxes, the most extractive employers at the bottom of job ladders must reduce the share of output captured as profit. The resulting diversion of profits to worker pay raises and partly offsets the tax induced impact of lower effort on bottom of the job ladder worker incomes. Competition between firms causes such (relative) pay rises to propagate up the job ladder. Income tax revenues are raised, but profit and profit tax revenues are squeezed.

The profit squeeze channel described above has novel implications for tax policy design in both affine and nonlinear tax settings. In an affine setting the policymaker applies the same marginal tax to all working agents and combines the proceeds with profit tax revenues to finance government spending, a lump sum transfer, and an unemployment benefit. As outlined, a rise in the tax rate diverts profits and profit tax revenues to worker incomes and income tax revenues. However, the costs and benefits of this diversion are not distributed uniformly across the population. On the one hand, *within* each talent market, workers lower on the job ladder benefit more from profit diversion. Since, under standard welfare criteria these poorer workers have higher social marginal welfare weights, this creates a motive for higher marginal taxation. On the other hand, workers in high talent markets work harder, earn more and pay more tax. A higher tax rate implies a larger mechanical after-tax income loss for those at the bottom of the job ladder in these markets. Compensating workers for this loss translates into a greater diversion of profits in high talent markets that is propagated up through the job ladder. Thus, relative to the frictionless benchmark economy, profit diversion redistributes across talent markets away from (on average poorer) low talents and towards (on average richer) high talents. Under standard welfare criteria this is a motive for lower marginal taxation. The impact of the profit squeeze channel on tax determination depends on the balance of these forces.

In the nonlinear income tax setting similar considerations play out. Now, a higher marginal tax *at* an income x diverts profits to the pre-tax incomes of workers earning *more than* x. The diversion of profits and the associated tax revenues to the incomes of high earners is socially costly and so a force for lower optimal marginal tax rates at high values of x is introduced. Optimal nonlinear tax formulas are extended to include readily interpretable terms that capture the marginal costs and benefits of profit diversion.

To assess the quantitative implications of labor frictions for optimal policy, we calibrate our model to US income data. This requires inverting the mapping from the underlying talent to the observed income distribution and, hence, disentangling that part of income variation attributable to talent and that part attributable to frictions. We then compute optimal affine and optimal nonlinear tax rates. To explore the implications of abstracting from labor market frictions, we raise the job finding-job destruction ratio (hence, dampening frictions), recalibrate the talent distribution and recompute optimal policy. We find that the profit squeeze channel works in the direction of modestly lowering marginal tax rates relative to the frictionless limit in the affine tax setting and of lowering them at higher incomes in the nonlinear setting.

The remainder of the paper proceeds as follows. After a brief literature review, Section 2 describes our baseline environment (with affine taxes). This section outlines how taxes impact competitive equilibria in our heterogeneous talent/frictional labor market setting. Sections 3 and 4 state and characterize the solution to the policymaker's (affine tax) problem in this setting. The latter also highlights two limiting cases: the non-frictional model and the homogenous talent model. Optimal nonlinear tax formulas are derived in Section 6 and quantitatively evaluated in Section 7. Section 8 concludes. Proofs and calibration details are provided in the appendices.

Literature The literature on optimal nonlinear income taxation originates with Mirrlees (1971). It was recast by Saez (2001) in terms of tax elasticities of income and attributes of the income distribution. Saez (2001) also supplies the first calibrated assessment of optimal nonlinear tax rates. These papers and most other contributions to the optimal nonlinear tax literature adopt a simple specification of the labor market in which there is a single market (and price) for effective labor and no market frictions. More recently, a new generation of normative public finance models has stressed fiscal spillovers and general equilibrium effects in richer labor market settings. Papers by Rothschild and Scheuer (2013) and Rothschild and Scheuer (2014) consider policymaking in a Roy model with multidimensional talent. Ales et al. (2015) analyze the implications of technical change for tax policy prescriptions in a related model. Ales and Sleet (2016) and Scheuer and Werning (2016b) use an assignment model to analyze the taxation of high income earners. Papers by Scheuer and Werning (2016a) and Sachs et al. (2016) provide unifying treatments across a variety of environments.

A smaller literature considers tax design in situations with search frictions. Boone and Bovenberg (2002) explore how taxes can be used to correct inefficiencies in settings with random (off-the-job) search, Nash bargaining between firms and workers and either free entry or a limited supply of firms. There is no on-the-job intensive effort margin, but unemployed workers exert costly effort in job search. Hungerbühler et al. (2006) augments the framework of Boone and Bovenberg (2002) with variation in talent (but omits search effort). In this setting, all workers with the same talent earn the same wage, but wages vary across talents. Optimal nonlinear taxation redistributes from more to less talented workers, but also deters workers from "bargaining aggressively". Thus, high marginal taxes distort the economy by encouraging too much job creation and output. Further variations on and extensions of these models are contained in Boone and Bovenberg (2006) and Lehmann et al. (2011). More recently Golosov et al. (2013) consider policy design in a model with directed search and no ex ante talent heterogeneity. Our paper advances this literature by analyzing optimal income taxation in an environment featuring both on and off the job search, ex ante talent heterogeneity and an intensive effort margin. Our focus on on-the-job search is partly motivated by the widespread adoption of the on-the-job search assumption in the structural job search literature.[3] It is also partly motivated by the observation of Hornstein et al. (2011) that large unem-

[3] Rich structural job search models featuring on-the-job search are analyzed by Bontemps et al. (2000), Postel-Vinay and Robin (2002), Lise et al. (2016) amongst many others. Although these papers contain elements not present in ours, they all share the Burdett-Mortensen model as the kernel of their more elaborate frameworks.

ployment to work transitions and the relatively short duration of unemployment is inconsistent with very much frictional wage dispersion in models that only feature off-the-job search, whether random or directed, and have plausible implications for the value of non-market time and for worker discount factors.

2 Baseline Environment

Worker preferences and constraints The economy is inhabited by a unit mass of agent-workers and a unit mass of firms. In each period, a fraction ρ of the worker population survives and a fraction $1-\rho$ dies and is replaced by an equal number of new borns. Workers have preferences over stochastic lifetime consumption and effort allocations $\{c_t, y_t\}_{t=0}^{\infty}$ of the form:

$$(1-\beta)\sum_{t=0}^{\infty}\beta^t E[U(c_t - h(y_t))],$$

where U and h are, respectively, concave and strictly convex functions and both are increasing and twice differentiable. In addition, $h(0)$ is normalized to 0. The discount factor $\beta \leq \rho$ incorporates the worker's survival probability and her patience.[4]

In our baseline environment we abstract from private insurance or savings. Hence, the consumption of an employed worker earning x and facing tax function T is just $c = x - T[x]$. We also initially restrict attention to affine tax functions of the form: $T[x] = -L + \tau x$, with L a lump sum transfer given to all employed workers and τ the marginal income tax rate. Later we generalize the analysis to allow for nonlinear income tax functions. Unemployed workers exert no effort and receive a benefit b.[5] Hence, for them $c = b$. It is convenient to call the argument $c - h(y)$ of U a worker's *current payoff* and to denote the current payoff of an employed worker earning income x and exerting effort y by:

$$\Psi^e(x,y) = x - T[x] - h(y).$$

[4]We also consider the limiting case $\beta = \rho = 1$. Worker payoffs are then identified with the aggregator $\lim_{T\to\infty}\frac{1}{T+1}\sum_{t=0}^{T}E[U(c_t - h(y_t))]$.

[5]We do not consider dynamic tax functions that condition on histories. In our model income is related to both talent and progress up the job ladder. Our restrictions on policy imply that a policymaker cannot distinguish between two workers earning the same income one of whom is an unlucky talented worker who has not progressed far up the job ladder and the other a lucky untalented worker who has. We abstract from policies that tie unemployment benefit to past incomes.

Similarly, the current payoff of an unemployed worker is:

$$\Psi^u(b, 0) = b.$$

Letting $i_t \in \{e, u\}$ denote the employment status of a worker at t and substituting the expressions for Ψ^i into the worker's preferences gives the following worker objective over employment-income-effort allocations $\{i_t, x_t, y_t\}_{t=0}^{\infty}$:

$$(1-\beta)\sum_{t=0}^{\infty}\beta^t E[U(\Psi^{i_t}(x_t, y_t))].$$

Worker employment, income and effort allocations Workers draw a permanent talent shock $\theta \in \Theta := [\underline{\theta}, \overline{\theta}]$ from a distribution K with density k at the beginning of their lives. Worker talent is observed by firms and, hence, workers are effectively partitioned into talent-specific labor markets. Firms sort themselves across these markets. A talent market is active only if firms can both attract workers from unemployment and earn non-negative profits. Each firm posts a profit-per-worker offer q in an active talent markets. A firm posting q commits to taking (no more than) q from each worker with which it matches. Let $F[\cdot|\theta]$ denote the distribution of firms across profit-per-worker offers in active market θ. A worker in talent market θ encounters a new firm drawn at random from $F[\cdot|\theta]$ with probability λ. If such a worker accepts a job from a firm offering q, she becomes residual claimant on the surplus after collection of q by the firm. She then selects an effort y, produces θy and takes the residual surplus $x = \theta y - q$ as pre-tax income. Equivalently, she chooses pre-tax income x to maximize:

$$\Psi^e\left(x, \frac{x+q}{\theta}\right) := L + (1-\tau)x - h\left(\frac{x+q}{\theta}\right).$$

Her optimal choice of income is $x(q,\theta) := \theta \frac{\partial h}{\partial y}^{-1}(\theta(1-\tau)) - q$ and her optimized current payoff is:

$$\Phi(q,\theta) := L - (1-\tau)q + \Gamma(\theta),$$

where $\Gamma(\theta) := (1-\tau)\{x(q,\theta) + q\} - h(\frac{x(q,\theta)+q}{\theta})$. Note that x, Φ and Γ all depend upon the policy parameters and, in particular, depend upon the marginal income tax rate τ, but for now we suppress this dependence in the notation.[6] Note also that this formulation is dual to assuming that firms post utilities, assign joint surplus

[6] Worker effort $x(q,\theta) + q = \theta \frac{\partial h}{\partial y}^{-1}(\theta(1-\tau))$ is independent of q and, hence, so too is $\Gamma(\theta)$.

maximizing efforts to accepting workers and pay incomes that attain the accepted utilities after reimbursement for effort.[7]

Since an employed worker's job finding probability is independent of her current job, her decision to accept a new job is a static one: she will accept if the offered q is below the one she currently has. A worker accepting such an offer will obtain a greater income and utility and, in this sense, climb the within-talent market job ladder. An unemployed worker obtains a current payoff b. Since an unemployed worker is no less (and no more) effective in finding jobs than an employed worker, her decision to accept a job is also a static one. She will do so if she encounters a firm making a profit-per-worker offer below $\overline{q}(\theta)$, where $\overline{q}(\theta)$ is the profit offer that sets a worker's current payoff from working equal to that from not working:

$$\Phi(\overline{q}(\theta), \theta) = b. \tag{1}$$

The value $\overline{q}(\theta)$ is the maximal profit offer that a firm can make (and have accepted with positive probability) in talent market θ. Inverting (1) gives an explicit expression for $\overline{q}(\theta)$:

$$\overline{q}(\theta) = -\frac{b-L}{1-\tau} + \frac{1}{1-\tau}\Gamma(\theta). \tag{2}$$

Let $\underline{q}(\theta)$ denote the minimal profit offer made in talent market θ.

The job of a surviving worker is destroyed with probability δ. Such a worker joins new born θ workers and other surviving workers who failed to find a job last period in the unemployment pool. Define the *survival-adjusted job finding* and *survival-adjusted job destruction* probabilities: $\tilde{\lambda} := \rho\lambda$ and $\tilde{\delta} := 1 - \rho + \rho\delta$. The ratio $\frac{\tilde{\lambda}}{\tilde{\delta}}$ is important: higher values of this ratio imply that ascent up the job ladder is rapid and descent by job destruction or death rare. As discussed below, under such values workers capture most of the surplus they create, income variation is largely attributable to talent variation and the economy more closely resembles a frictionless one. Conversely, lower values of $\frac{\tilde{\lambda}}{\tilde{\delta}}$ imply that job ladders play a greater role in explaining income dispersion amongst workers.

[7] Our profit posting formulation is simpler and aligns more closely with public finance in which workers select incomes. In Burdett and Mortensen (1998), workers do not supply effort and firms make simple income offers x. Their model could equally be formulated as one in which firms make profit-per-worker offers.

The firm problem Let $N(q|\theta)$ denote the number of workers attracted to a firm posting profit per worker offer q. Firm profits are given by:

$$\pi(q|\theta) = N(q|\theta)q$$

and firms select θ (the talent market in which they post) and q (the profit per worker offer they post) to maximize π. Profits in our model are pure rents to firm ownership. Provided that the social marginal value of a dollar of tax revenues distributed to all agents in the form of an addition to the lump sum transfer L and the unemployment benefit b exceeds the social marginal value of a dollar distributed to firm owners, it is optimal to tax this dollar at 100%. We assume that this is true for the first (and all subsequent) dollars potentially earned by firm owners. Consequently, we directly assume that the policymaker (optimally) taxes profits at 100%.[8]

Active and inactive talent markets As noted, firms allocate themselves across θ markets and profit-per-worker offers so as to maximize total profit. No firm enters a market in which the maximal profit-per-worker that a worker will accept is negative. Hence, firms only enter talent markets $\theta \in [\tilde{\theta}, \overline{\theta}]$, where:

$$\tilde{\theta} = \max\{\underline{\theta}, \hat{\theta}\} \quad \text{and} \quad \overline{q}(\hat{\theta}) = 0.$$

From (2), the threshold $\tilde{\theta}$ is given by:

$$\tilde{\theta} = \max\left\{\underline{\theta}, \Gamma^{-1}(\max\{b - L, 0\})\right\}. \tag{3}$$

We refer to talent markets $\theta \in (\tilde{\theta}, \overline{\theta}]$ as *active*, talent markets $\theta \in [\underline{\theta}, \tilde{\theta}]$ as *inactive* and $\tilde{\theta}$ as the *activity threshold*.

Steady-state equilibrium distributions We focus on steady-state competitive equilibria. The inflow of workers in active talent markets into unemployment con-

[8] In frictional settings income and profit taxation must be jointly considered. Our treatment of profit as a pure rent that is optimally taxed at 100% is standard in public finance. In principle, profit taxation may be constrained by disincentives to firm creation or by exogenous institutional considerations. Note, however, that we follow the Burdett-Mortensen framework and treat job finding probabilities as being independent of the number of firms. Under this assumption income and employment distributions are also independent of the number of firms (provided that number is non-zero). If more firms enter then profits-per-firm scale down, but total profits do not. In particular, if instead of a unit mass of firms, we assumed free entry subject to a fixed entry cost, profit tax rates close to 100% allow the government to capture nearly all aggregate profit (leaving a small amount over for the small number of entering firms to pay entry costs). Worker income and employment distributions are unchanged.

sists of new borns and survivors who lose their jobs. The outflow consists of unemployed workers who die and those who survive and find jobs. In steady state these must be identical. Hence, the unemployment rate in each active talent market satisfies:

$$1-\rho+\rho(1-u)\delta=(1-\rho)u+\rho\lambda u. \tag{4}$$

Rearranging (4) implies:

$$u=\frac{\tilde{\delta}}{\tilde{\lambda}+\tilde{\delta}}, \tag{5}$$

where recall $\tilde{\lambda}=\rho\lambda$ and $\tilde{\delta}=1-\rho+\rho\delta$ denote the survival adjusted job finding and job destruction probabilities. Workers with types $[\underline{\theta},\tilde{\theta}]$ are born into and never leave unemployment. They are effectively out of the labor force.

The total steady-state mass of unemployed (inclusive of those out-of the labor force) is:

$$K[\tilde{\theta}]+u(1-K[\tilde{\theta}]).$$

Recall that $F[\cdot|\theta]$ denotes the distribution of firms across profit-per-worker offers in active talent market θ and that a worker in this market meets a firms drawn at random from $F[\cdot|\theta]$ with probability λ. Let $G[\cdot|\theta]$ denote the distribution of workers across profit offers in talent market θ. Unemployed workers in market θ will accept any profit offer at or below $\overline{q}(\theta)$; employed workers will not accept any offer above their current one. Consequently, the inflow of workers into firms making profit-per-worker offers above q equals the mass of surviving unemployed θ-workers who meet such firms, i.e. the inflow equals $\tilde{\lambda}(1-F[q|\theta])u$. The outflow of workers from firms posting profit offers above q equals the mass of such workers who either die, exogenously separate or encounter a better (lower) profit-per-worker offer firm. This outflow equals $\{\tilde{\delta}+\tilde{\lambda}F[q|\theta]\}(1-G[q|\theta])(1-u)$. Equating these terms, substituting for u and rearranging gives:

$$G[q|\theta]=\left(\frac{\tilde{\delta}+\tilde{\lambda}}{\tilde{\delta}+\tilde{\lambda}F[q|\theta])}\right)F[q|\theta]. \tag{6}$$

The mass of θ workers recruited by a firm offering q consists of those who make contact with the firm and are either unemployed or are employed at firms offering $q'>q$. It is given by:

$$R(q|\theta)=\tilde{\lambda}[(1-u)(1-G[q|\theta])+u]\frac{k(\theta)}{m(\theta)},$$

where m is the density of firms across θ markets. Workers separate from firms when they die, exogenously loose their job or receive a better offer. The rate at which they separate from a firm offering q is, thus, $S(q|\theta) = \tilde{\delta} + \tilde{\lambda}F[q|\theta]$. Let $N(q|\theta)$ denote the number of θ workers per firm offering q. Equating steady-state recruitment and separations and using (5) and (6) implies:

$$N(q|\theta) = \frac{\tilde{\lambda}\tilde{\delta}}{\{\tilde{\delta} + \tilde{\lambda}F[q|\theta]\}^2}\frac{k(\theta)}{m(\theta)}.$$

Hence, the total profit of a firm entering market θ and offering q is:

$$\pi(q|\theta) = q\frac{\tilde{\lambda}\tilde{\delta}}{\{\tilde{\delta} + \tilde{\lambda}F[q|\theta]\}^2}\frac{k(\theta)}{m(\theta)}. \tag{7}$$

In particular, the profit made by a firm posting the maximum profit-per-worker offer $\overline{q}(\theta)$ is:

$$\pi(\overline{q}(\theta)|\theta) = \overline{q}(\theta)\frac{\tilde{\lambda}\tilde{\delta}}{\{\tilde{\delta} + \tilde{\lambda}\}^2}\frac{k(\theta)}{m(\theta)}. \tag{8}$$

Firms distribute themselves across θ markets and q demands so as to maximize profits. Hence, profits are equated at all θ and q in the support of F. This in turn determines the density of firms across markets:

$$m(\theta) = \frac{\overline{q}(\theta)k(\theta)}{\int_{\tilde{\theta}}^{\overline{\theta}} \overline{q}(\theta')k(\theta')d\theta'} \tag{9}$$

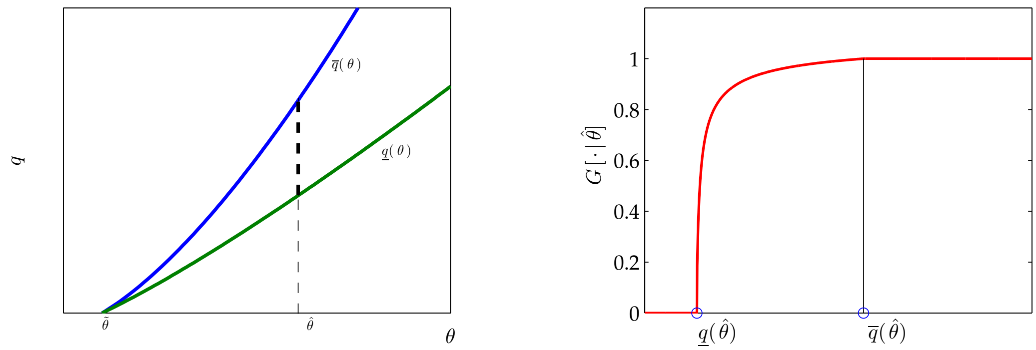
and the distribution of profit demands within each market, for $q \in [\underline{q}(\theta), \overline{q}(\theta)]$,

$$F[q|\theta] = -\frac{\tilde{\delta}}{\tilde{\lambda}} + \frac{\tilde{\delta} + \tilde{\lambda}}{\tilde{\lambda}}\sqrt{\frac{q}{\overline{q}(\theta)}}. \tag{10}$$

Since $F[\underline{q}(\theta)|\theta] = 0$, (10) implies that the minimal proft-per-worker offer made in the θ talent market is $\underline{q}(\theta) = \left(\frac{\tilde{\delta}}{\tilde{\delta}+\tilde{\lambda}}\right)^2 \overline{q}(\theta)$. The corresponding density on $[\underline{q}(\theta), \overline{q}(\theta)]$ is $f(q|\theta) = \frac{\tilde{\delta}+\tilde{\lambda}}{2\tilde{\lambda}}\frac{1}{q}\sqrt{\frac{q}{\overline{q}(\theta)}}$. Finally, combining (6) and (10) gives the distribution of workers across profit offers:

$$G[q|\theta] = \frac{\tilde{\delta} + \tilde{\lambda}}{\tilde{\lambda}} - \frac{\tilde{\delta}}{\tilde{\lambda}}\sqrt{\frac{\overline{q}(\theta)}{q}}. \tag{11}$$

The corresponding density is $g(q|\theta) = \frac{\tilde{\delta}}{2\tilde{\lambda}}\sqrt{\frac{\overline{q}(\theta)}{q}}\frac{1}{q}$. The following figure illustrates the situation. Figure 1(a) shows the activity threshold $\tilde{\theta}$ and the minimal $\underline{q}$ (green curve) and maximal $\overline{q}$ (blue curve) profit-per-worker in each active talent market (under a benchmark parameterization). The region between the $\underline{q}$ and $\overline{q}$ curves shows the



(a) Minimal and maximal profit offers.

(b) Distribution of workers across profit offers in market $\hat{\theta}$.

Figure 1: Talent and profit-per-worker in equilibrium.

support of the talent and profit-per-worker distributions in active talent markets. The interval of profit offers in a particular talent market $\hat{\theta}$ is indicated by the solid dashed line. Figure 1(b) shows the distribution of workers across profit-per-worker offers in this talent market. Note that $\underline{q}$ converges to zero and the concentration of workers at $\underline{q}$ becomes more pronounced as $\frac{\tilde{\lambda}}{\tilde{\delta}}$ increases. At large values of $\frac{\tilde{\lambda}}{\tilde{\delta}}$ workers cluster at the top of the job ladder and the economy more closely resembles a frictionless one.

Policy maker budget balance Equilibrium requires that government spending $\mathcal{G}$ and transfers L and b are funded through profit and income taxation and, hence, government budget balance achieved. We assume that the government cannot borrow or lend. Using the previously derived expressions for the fraction of workers unemployed, steady state budget balance implies:

$$\begin{aligned}\mathcal{G} + \left\{K[\tilde{\theta}] + (1 - K[\tilde{\theta}])\frac{\tilde{\delta}}{\tilde{\lambda} + \tilde{\delta}}\right\}b + \left\{(1 - K[\tilde{\theta}])\frac{\tilde{\lambda}}{\tilde{\lambda} + \tilde{\delta}}\right\}L \\ = \frac{\tilde{\lambda}}{\tilde{\lambda} + \tilde{\delta}}\int_{\tilde{\theta}}^{\overline{\theta}}\int_{\underline{q}(\theta)}^{\overline{q}(\theta)}\{q + \tau x(q, \theta)\}\, g(q|\theta)k(\theta)dqd\theta. \qquad (12)\end{aligned}$$

The impact of taxes Variations in the marginal tax rate τ impact the economy through three main channels. First, as in standard non-frictional public finance models, τ affects a (participating) worker's intensive effort margin and, so, the total output produced by that worker. Second, the income tax rate τ raises the activity threshold $\tilde{\theta}$ and, hence, reduces the fraction of active talent markets. Consequently, it impacts the fraction of agents who participate in the labor market.

Third, τ affects the distribution of profits-per-worker. Specifically, a higher tax rate reduces a worker's current payoff holding the profit extracted from that worker q fixed. In the face of an unchanged $b - L$, firms extracting the maximal profit-per-worker $\overline{q}(\theta)$ are compelled to reduce their profit offers in order to attract workers. As they do so they limit the ability of other firms previously making slightly lower profit-per-worker offers to poach workers. This reduces the size of such firms, lowers their total profits and modifies their tradeoff between firm size and profit-per-worker. They too respond by reducing their profit-per-worker offers. In this way competition amongst firms for workers ensures that the impact of the tax rise propagates through the entire profit offer distribution in each talent market. In doing so it directly reduces profit tax revenues. Moreover, since a worker's income is the difference between her output and the profit extracted by firms, this effect offsets the impact of reduced effort and output on worker income. Hence, it affects income tax revenues and has direct implications for worker welfare. This *profit squeeze channel* is the primary difference between the operation of tax policy in our frictional model and a more standard setting.

To formalize (and extend) the preceding discussion, let $J[\cdot|\theta]$ denote the distribution of workers across incomes in talent market θ and let $\xi(s,\theta)$ denote the pre-tax income earned by a worker at the s-th quantile of $J[\cdot|\theta]$, i.e. $s = J[\xi(s,\theta)|\theta]$. Using the first order condition of a worker and the expression (11) for the distribution $G[\cdot|\theta]$, $\xi(s,\theta)$ is given by:

$$\xi(s,\theta) := z(\theta) - B(s)\overline{q}(\theta), \tag{13}$$

where $B(s) = \left(\frac{\tilde{\delta}}{\tilde{\delta}+\tilde{\lambda}s}\right)^2$ and $z(\theta) = \theta \frac{\partial h}{\partial y}^{-1}(\theta(1-\tau))$ is the surplus produced by the worker. Clearly, $\xi(s,\theta) \leq z(\theta)$ since some output is captured by firm profit. In addition, using the definition of $\overline{q}(\theta)$, a small increase in τ causes a change in $\xi(s,\theta)$ of:

$$\frac{\partial \xi}{\partial \tau}(s,\theta) = \frac{\partial z}{\partial \tau}(\theta) + B(s)\left\{\frac{\xi(0,\theta)}{1-\tau}\right\} > \frac{\partial z}{\partial \tau}(\theta). \tag{14}$$

The first term captures the intensive effort margin effect of taxation on surplus and, hence, worker income. The second term gives the offsetting profit squeeze effect. Note that the second term can exceed the first: an increase in τ can squeeze profits sufficiently that worker incomes rise despite a decline in surplus.[9] Note

[9]Precisely, using the worker's first order conditions: $\frac{\partial \xi}{\partial \tau}(s,\theta) = -\frac{\theta^2}{h''(\frac{z(\theta)}{\theta})} + B(s)\left\{\frac{z(\theta)-\overline{q}(\theta)}{1-\tau}\right\} = \frac{z(\theta)}{(1-\tau)}\left\{-\frac{h'(\frac{z(\theta)}{\theta})}{\frac{z(\theta)}{\theta}h''(\frac{z(\theta)}{\theta})} + B(s)\left\{1-\frac{\overline{q}(\theta)}{z(\theta)}\right\}\right\}$. The term $\frac{h'(\frac{z(\theta)}{\theta})}{\frac{z(\theta)}{\theta}h''(\frac{z(\theta)}{\theta})}$ is the worker's effort supply elasticity, while the term $B(s)\left\{1-\frac{\overline{q}(\theta)}{z(\theta)}\right\} \in [0,1]$. If the effort elasticity is small enough, then some workers will experience an increase in the pre-tax income in response to a tax rise.

also that the profit squeeze effect is decreasing in the income quantile s (because $B(s)$ is decreasing), whereas the intensive margin effect is independent of s. Thus, workers at the bottom of the job ladder receive the smallest reductions (or largest increases) in income in response to a tax rise. Intuitively, competition amongst firms for workers propagates the depressing impact of the tax rise on firm profits up the job ladder, but this effect dissipates at higher rungs of the ladder. Figure 2 illustrates this. It isolates the profit squeeze effect on $J[\cdot|\theta]$ holding the intensive margin effect fixed. A tax rate increase from low to high causes a large horizontal shift in the income distribution at low income quantiles, but a negligible one at high quantiles. In combination the intensive margin and profit squeeze channels

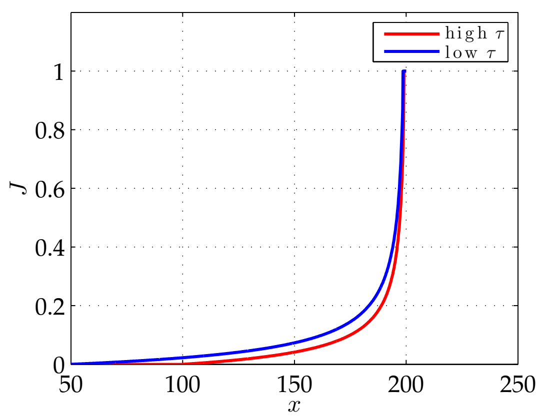


Figure 2: The figure shows the pure profit squeeze effect on a within-talent-market income distribution. Workers at low income quantiles experience larger falls in profit offers and greater rises in income from this channel.

imply that the income distribution *within* a given active talent market becomes more compressed as taxes rise and that those at the bottom of job ladders loose less (or benefit more) from a tax rate rise. Since the intensive margin and profit squeeze impacts depend upon talent θ, a tax rate rise also has implications for the distribution of incomes *across* talent markets. In particular, the profit channel component of (14), $B(s)\left\{\frac{\xi(0,\theta)}{1-\tau}\right\}$, is increasing in θ. Workers in high talent markets work harder, earn more and pay more tax. Consequently, a higher tax rate implies a larger mechanical after-tax income loss for those at the bottom of the job ladder in these markets. Compensating workers for this loss leads to a correspondingly greater profit squeeze in high talent markets. The profit squeeze channel thus works to redistribute from lower to higher talent markets where on average workers have higher current payoffs. This tension between the intra and inter talent market redistributions of income is important for our subsequent affine tax analysis.

The overall impact of a small increase in τ on income tax revenues combines the mechanical gain from a higher tax rate holding pre-tax incomes fixed with the effect on pre-tax incomes from the three channels described above. Further, there is a loss of profit tax revenues stemming from the profit squeeze channel. A higher tax rate also directly affects worker utility. This effect combines the mechanical loss in after-tax income (holding pre-tax income fixed) with the impact of taxes on pre-tax income and effort. It is given by: $-\xi(s,\theta) + (1-\tau)\frac{\partial x}{\partial \tau}(s,\theta) + \mathcal{E}(\theta)z(\theta)$, where $\mathcal{E}(\theta) = \frac{h'(y(\theta))}{h''(y(\theta))y(\theta)}$ is the conventional wage elasticity of effort. In the standard setting, an envelope theorem implies that the second and third terms in the preceding expression sum to zero and, hence, that the tax impact on a worker's current payoff is given entirely by the mechanical effect. However, in the current frictional model, the impact of taxes on incomes is dampened by the profit squeeze. Consequently, the second and third terms in the loss expression do not sum to zero and the overall utility loss is less than the mechanical loss.

3 The Affine Tax Policymaker Problem

Define $r := b - L$ to be the additional transfer paid to the unemployed over and above that paid to the employed and collect policy parameters into $P := (r, L, \tau)$. In the notation below we make dependence of variables on P explicit.

We assume that the policy maker maximizes the lifetime utility of each cohort of new born agents at the steady state. Hence, the policymaker's objective is:

$$\int_{\underline{\theta}}^{\overline{\theta}} (1-\beta) \sum_{t=0}^{\infty} \beta^t E[U(\Phi_t(P))|\theta; P] k(\theta) d\theta, \tag{15}$$

where the function $\Phi_t(P)$ takes the value $\Phi(q_t, \theta; P)$ for an employed θ talent agent with profit offer q_t at t (under policy parameter P) and b for an unemployed agent at t. The expectation operator in (15) is that induced by F, $\tilde{\theta}$, $\underline{q}$ and $\overline{q}$ (all of which depend on P) and the parameters λ and δ.

Let $\{Q_t[\cdot|\theta, P]\}$ denote the induced probability distributions over a θ talent worker's employment outcomes in each period of her life given policy P.[10] Define the discount-adjusted probability measure over lifetime employment outcomes for a talent θ

[10]Specifically, a θ worker is unemployed with probability 1 in period 0 of her life. If she survives into the next period, then with probability $\lambda F[q|\theta, P]$ she transitions to employment with a profit-per-worker offer of less than q and with probability $1-\lambda$ she remains unemployed and so forth.

worker by:

$$\tilde{Q}[\cdot|\theta,P] := (1-\beta)\sum_{t=0}^{\infty}\beta^t Q_t[\cdot|\theta,P].$$

Let $\tilde{E}[\cdot|\theta,P]$ denote the expectation operator implied by this probability measure and let $\tilde{Q}[\cdot|P]$ and $\tilde{E}[\cdot|P]$ denote the corresponding unconditional probability measure and expectation after integration over the talent distribution. If $\beta = \rho$ and survival risk is the worker's sole motive for discounting, then $\tilde{Q}[\cdot|\theta,P]$ equals the cross sectional distribution of θ talent workers across employment outcomes:

$$Q[\cdot|\theta,P] := (1-\rho)\sum_{t=0}^{\infty}\rho^t Q_t[\cdot|\theta,P],$$

with corresponding expectation $E[\cdot|\theta,P]$. More generally, if $\beta < \rho$, then $\tilde{Q}[\cdot|\theta,P]$ places less weight on outcomes attained (more frequently) later in life than the cross sectional distribution. The policymaker's problem may be conveniently stated as:

$$\sup_{P}\ \tilde{E}[U(\Phi(P))|P] \tag{16}$$

subject to the budget constraint:

$$\mathcal{G} + \left\{K[\tilde{\theta}(P)] + (1-K[\tilde{\theta}(P)])\frac{\tilde{\delta}}{\tilde{\lambda}+\tilde{\delta}}\right\}r$$
$$= -L + \frac{\tilde{\lambda}}{\tilde{\lambda}+\tilde{\delta}}\int_{\tilde{\theta}(P)}^{\overline{\theta}}\int_{\underline{q}(\theta;P)}^{\overline{q}(\theta;P)}\{q+\tau x(q,\theta)\}\,g(q|\theta;P)k(\theta)dqd\theta. \tag{17}$$

4 Analysis of the Policymaker's Problem

This section describes optimal policy in the affine tax environment. Optimal values are denoted by * superscripts, e.g. P^* denotes the optimal policy triple, $E^*[\cdot] = E[\cdot|P^*]$ an induced expectation under this optimal policy and so forth.

Choosing L Let μ^* denote the optimal multiplier on the government budget constraint. The policymaker's first order condition for L is simply $\mu^* = \tilde{E}^*[U'(\Phi^*)]$. Thus, the social shadow value of government funds equals the discount-adjusted expected worker marginal utility.

Choosing τ Proposition 1 derives an expression for τ^* in a form that is comparable to those found elsewhere in the literature. In particular, it incorporates additional terms that capture the role of the profit squeeze channel in shaping policy. Public finance models that treat the labor market as being non-frictional omit this channel and the corresponding terms. Let $\tilde{\text{Cov}}^*$ denote a covariance under the discount-adjusted probability measure at the optimal policy and let $\mathcal{E}_z^*$ denote the elasticity of a variable z with respect to $1-\tau$ at the optimal policy. We will refer to the latter as the "tax elasticity of z" for brevity.

Proposition 1. *The optimal marginal tax rate satisfies the condition:*

$$\begin{aligned} 0 = &- \tilde{Cov}^*(U'(\Phi), x) - \tilde{E}^*[U'(\Phi)]\tau^* \frac{E^*[x]\mathcal{E}^*_{E[x]}}{1-\tau^*} \\ &+ \tilde{E}^*[U'(\Phi)]r^* \frac{\tilde{\theta}^* \mathcal{E}^*_{\tilde{\theta}}}{1-\tau^*} \frac{\tilde{\lambda} k(\tilde{\theta}^*)}{\tilde{\lambda}+\tilde{\delta}} \\ &+ \tilde{Cov}^*(U'(\Phi), \mathcal{E}_{\bar{q}} q) - \tilde{E}^*[U'(\Phi)]\tau^* \frac{\pi^* \mathcal{E}^*_\pi}{1-\tau^*} \\ &+ \tilde{E}^*[U'(\Phi)]\{E^*[x] - \tilde{E}^*[x]\} - \tilde{E}^*[U'(\Phi)]\{E^*[\mathcal{E}_{\bar{q}} q] - \tilde{E}^*[\mathcal{E}_{\bar{q}} q]\}. \end{aligned} \tag{18}$$

Proof. See Appendix A. □

Proposition 1 has the following immediate corollary.

Corollary 1. *If $\beta = \rho$, the optimal marginal tax rate satisfies the condition:*

$$\begin{aligned} 0 = &- Cov^*(U'(\Phi), x) - E^*[U'(\Phi^*)]\tau^* \frac{E^*[x]\mathcal{E}^*_{E[x]}}{1-\tau^*} \\ &+ E^*[U'(\Phi)]r^* \frac{\tilde{\theta}^* \mathcal{E}^*_{\tilde{\theta}}}{1-\tau^*} \frac{\tilde{\lambda} k(\tilde{\theta}^*)}{\tilde{\lambda}+\tilde{\delta}} \\ &+ Cov^*(U'(\Phi), \mathcal{E}_{\bar{q}} q) - E^*[U'(\Phi)]\tau^* \frac{\pi^* \mathcal{E}^*_\pi}{1-\tau^*}. \end{aligned} \tag{19}$$

Interpreting Proposition 1. We first interpret the simpler equation (19) in Corollary 1 and then the more complicated expression (18) in Proposition 1. The first two terms in (19) are present in a standard (non-frictional) public finance setting with heterogeneous agents. Indeed in that setting (19) reduces to:

$$-\text{Cov}^*(U'(\Phi), x) - E^*[U'(\Phi)] \frac{\tau^*}{1-\tau^*} E^*[x]\mathcal{E}^*_{E[x]} = 0. \tag{20}$$

The first term in either (19) or (20) gives the marginal mechanical social benefit from collecting and redistributing income via a higher tax. It abstracts from behavioral and equilibrium responses. Precisely, a small increment in the tax rate of $\delta\tau$ levied on a (θ, q) worker mechanically depresses the worker's income and consumption.

This in turn depresses social welfare by: $-U'(\Phi^*)x^*(q,\theta)\delta\tau$.[11] Aggregating across workers gives a social cost of $-E^*[U'(\Phi)x]\delta\tau$. However, the increased tax rate also mechanically raises tax revenues by $E^*[x]\delta\tau$. The use of these revenues to fund an increased lump sum transfer has a social benefit of $E^*[U'(\Phi)]E^*[x]\delta\tau$.[12] Combining these terms gives an expression proportional to the first covariance term in (18), i.e. proportional to $-E^*[U'(\Phi)x] + E^*[U'(\Phi)]E^*[x]$. This covariance is negative and thus describes a redistributive motive for higher marginal taxes.

The second term in either (19) or (20) gives additional adverse behavioral consequences of a small increase in τ for income tax revenues. In the standard frictionless model this stems from an intensive margin response: working agents reduce their effort in response to higher taxes, so decreasing both their incomes and their income tax payments. This channel is present in the frictional model, but it is supplemented by two other effects. First, there is an extensive margin response: the activity threshold adjusts upwards, some additional talent markets become inactive and the corresponding workers no longer work or pay tax. This effect reinforces the usual intensive margin one. Second, some of the burden of the income tax falls on firm profits. Specifically, and consistent with our earlier description, the increase in τ reduces the maximal profit offer needed to make a worker indifferent between working and unemployment and, hence, causes the distribution of profit demands within each θ market to shift leftwards and place more mass on lower values. This raises and, hence, partially offsets the depressing effects of the intensive and extensive channels on aggregate pre-tax worker income. In our frictional model all three of these effects are folded into the variable $\mathcal{E}^*_{E[x]}$ in (19). It corresponds to the (aggregate) *elasticity of taxable income*.

The frictional model also introduces new terms into the standard expression (20). The first is:

$$E^*[U'(\Phi)]r^*\frac{\tilde{\theta}^*\mathcal{E}^*_{\tilde{\theta}}}{1-\tau^*}\frac{\tilde{\lambda}k(\tilde{\theta}^*)}{\tilde{\lambda}+\tilde{\delta}}. \tag{21}$$

Unemployed workers receive a transfer supplement $r^* = b^* - L^*$. The higher tax rate renders additional markets inactive and raises the number of workers claiming this transfer. The associated fiscal cost is captured in (21). It augments the social cost of lost income tax revenue at the extensive margin already folded into the second

[11]In the standard (non-frictional) setting, there are no further direct effects on the worker's welfare from the tax change: by an envelope theorem, the welfare loss from behavioral adjustments in income and consumption exactly match the benefit from reduced leisure. However, in our frictional setting there are additional behavioral and equilibrium effects which are accounted for by some of additional terms in (19) relative to (20).

[12]Where we use the fact that a unit of lump sum transfer has a social marginal benefit of $E^*[U'(\Phi)]$ when $\beta = \rho$ in our setting.

term of (19). The remaining term in (19) is:

$$\mathrm{Cov}^*(U'(\Phi), \mathcal{E}_{\bar{q}} q) - E^*[U'(\Phi)]\tau^* \frac{\pi^* \mathcal{E}^*_\pi}{1-\tau^*}. \tag{22}$$

This captures (further) implications of the profit squeeze channel. Recall that via this channel the higher tax rate diverts profits and profit tax revenues to pre-tax worker incomes. The rise in pre-tax worker incomes can be split into a rise in after-tax (and net of transfer) incomes and a rise in income tax revenues. The fall in profit tax revenues can be correspondingly split into parts that match these increases. By definition the profit squeeze-induced income tax revenue rise and the matching component of the profit tax revenue fall have a neutral overall impact on the policymaker's budget. However, the income tax revenue rise has already been incorporated into the second term in (19) and, hence, the matching profit tax revenue fall must be subtracted. The second term in (22) does this. The profit squeeze-induced rise in after-tax worker incomes and the matching fall in profits, profit tax revenues and the lump sum transfer implies a redistribution of consumption across workers. The marginal social benefit of this redistribution is captured by the covariance term $\mathrm{Cov}^*(U'(\Phi), \mathcal{E}_{\bar{q}} q)$ in (22) and in the final line of (19).[13] The sign of this covariance is ambiguous.[14] On the one hand, within each talent market, workers with higher values of $\mathcal{E}_{\bar{q}} q$ are lower on the job ladder and, hence, have lower current payoffs, Φ. If this force predominated, the covariance would be positive. A socially desirable redistribution would then be achieved via the suppression of profits and the profit squeeze channel would be a force for a higher marginal tax rate. On the other hand, as discussed previously, the profit squeeze is stronger in higher talent markets and tends to redistribute towards the (on average higher utility) workers in these markets. If this effect is dominant, then the covariance is negative, the redistribution associated with the profit squeeze is socially undesirable and the profit squeeze works in favor of a lower marginal tax rate.

Equation (18) in Proposition 1 generalizes (19) to the case in which $\beta < \rho$. When $\beta < \rho$, the discount-adjusted probability measure of workers across employment states no longer coincides with the actual distribution of workers across these states. The former weights employment states attained earlier in life more heavily. These include unemployment and higher profit-per-worker employment states

[13]This term is analogous to the first (covariance) term in (19), $-\mathrm{Cov}^*(U'(\Phi), x)$. Recall that this latter covariance describes the marginal social benefit of the (mechanical) income tax redistribution. In contrast, the new redistributive term enters (19) (and (22)) positively since the redistribution is away from rather than towards the lump sum transfer. In addition, its sign is more ambiguous.

[14]Sufficient conditions for signing it are discussed in Appendix A.

lower on the job ladder that are more likely to be attained earlier in a career. The relevant covariance terms and the shadow value of resources in (18) are computed using the discount adjusted probability measure. However, total labor income $E^*[x]$ and the impact of the tax rate on total profit per worker offers $E^*[\mathcal{E}_{\bar{q}} q]$ are computed using the actual rather than the discount adjusted probability measures requiring the adjustments to the covariance terms contained in the second line of (18).

Connection to Saez Saez (2001) analyzes the optimal linear tax on incomes above a threshold. His focus is on the taxation of top incomes, but the analysis applies to affine taxation of all incomes. He organizes the first order condition for the optimal marginal tax rate as: $(1-\bar{g})M + \tau B = 0$, where M is the mechanical effect on income tax revenues of a small income tax rate change, B is the behavioral effect and $\bar{g}$ is the marginal social value of dollars taken from people earning incomes greater than the threshold above which the higher tax rate is applied. The term $(1-\bar{g})M$ is analogous to the first term in (18). It captures the social value of income tax revenues mechanically collected via the tax rise net of the cost to tax payers. Saez models all agents as being on the intensive margin. Hence, if the higher tax rate is applied to all incomes, B reduces to the average uncompensated elasticity of income with respect to the retention rate $1-\tau^*$ scaled by the ratio of income to the retention rate. Thus, B corresponds to the second term in (18).

4.1 Special Cases and an Extension

We briefly describe two limiting cases. In the first, there is a frictional labor market, but no talent dispersion. In the second, there is talent dispersion, but $\tilde{\lambda}/\tilde{\delta}$ is arbitrarily large implying that there is no frictional labor market: in steady state all workers are "at the top of the job ladder" and collect all of the rents. We also describe an extension in which agents can save. To simplify, we assume $\beta = \rho$.

Frictional labor market without talent dispersion In the absence of talent dispersion (with the common θ normalized to 1), a first best allocation is attainable in which those in work supply the first best efficient level of effort and agents are fully insured against job loss. The optimal tax system consists of a lump tax paid by those in work, a zero marginal tax rate and a (fully insuring) unemployment benefit. This tax system eliminates dispersion in job offers and the associated ex post luck.

More precisely, suppose that the unemployment benefit b is set to $\frac{\tilde{\lambda}}{\tilde{\lambda}+\tilde{\delta}}\{y^* - h(y^*)\}$, where y^* satisfies $1 - h'(y^*) = 0$. Marginal tax rates are zero and the lump sum tax is set to $L = -\frac{\tilde{\delta}}{\tilde{\lambda}+\tilde{\delta}}\{y^* - h(y^*)\}$. Thus a worker matched with a firm offering q selects effort equal to the first best level and earns a current payoff of $\frac{\tilde{\lambda}}{\tilde{\lambda}+\tilde{\delta}}\{y^* - h(y^*)\} - q$. It follows that firms can only attract workers if they make $q = 0$ offers. Dispersion in such offers (and in firm profits) is endogenously eliminated by the tax system. Consequently, job ladders and frictional income dispersion cannot exist in the face of an optimally designed tax system absent private variation in talent. When talent variation *is* present, driving the unemployment benefit high enough to set all profit offers to zero in a given talent market involves closing markets for less talented workers. This is costly and is only done up to the threshold $\tilde{\theta}^*$. Dispersion in profit offers and, hence, job ladders in active talent markets above the threshold survive.

Dispersion in talent without a frictional labor market Suppose that the support of the productivity distribution h is not a singleton, but an interval $[\underline{\theta}, \overline{\theta}]$ and that $\tilde{\lambda}/\tilde{\delta}$ is arbitrarily large. In this case, there is no steady state unemployment and all θ workers receive the best possible profit offers from firms: $q = 0$. Studying either the original problem (16) or its first order condition with respect to τ as $\tilde{\lambda}/\tilde{\delta} \to \infty$ reveals that in the limit the first order condition for τ reduces to (20).

Private savings Our benchmark environment omits private savings. Thus it does not permit older workers to self-insure against job loss through the accumulation of private claims. However, our tax formula (18) in Proposition 1 survives in more general environments that do incorporate private savings. For example, if workers are able to purchase annuities at an exogenously given price and income from annuities is untaxed, then as described in Appendix A, tax formula (18) continues to hold. However, the values of the covariances (and other terms) that enter the formula are modified by this extension.[15]

[15] Our previous theoretical analysis implies that via the profit squeeze channel higher income tax rates create intra- and inter-talent market redistributions with the former a force for higher optimal rates and the latter a force for lower optimal rates. In our quantitative section below we find that the latter force is the dominating one and that the profit squeeze channel works to depress optimal marginal taxes. Savings permit agents to better insure themselves against frictional wage dispersion. We conjecture that this would dilute the benefits of intra-talent market redistribution and depress the covariance $\mathrm{Cov}^*(U'(\Phi), \mathcal{E}_{\tilde{q}} q)$. This would reinforce our results and strengthen the case for lower optimal taxes in frictional settings.

5 Quantitative Analysis of Optimal Affine Taxes

This section explores quantitative implications of our theory for optimal affine tax determination.

Calibration Worker preferences are assumed to be of the form:

$$U(c - h(y)) = \frac{1}{1-\sigma}\left(c - \frac{1}{1+\gamma}y^{1+\gamma}\right)^{1-\sigma}.$$

The primitives of our model are then the preference parameters σ, γ and β, the survival probability ρ, the exogenous job separation and job finding rates δ and λ and the talent distribution. We set $\sigma = 2$ and consider $\gamma = 1$ and 2. The model's time period is set to be one month and the discount factor β and survival probability ρ are set to a common and limiting value of 1.[16] We choose δ and λ to be in line with the empirical job destruction and job finding rates reported in the literature. Specifically, Shimer (2012) computes monthly job destruction rates in the neighborhood of 0.03 for the US in the post 1985 period. Thus, we set $\delta = 0.03$. Shimer (2012) reports monthly job finding rates for unemployed workers that fluctuate around 0.4 in the same period. Shimer (2012) does not report job finding rates for employed workers. Hornstein et al. (2011) calibrate a model with job-to-job transitions to US data and recover values for the job finding rates of employed workers of between 0.08 and 0.12. Our model does not incorporate distinct job finding rates for the employed and the unemployed. Consequently, we set λ equal to a weighted average of the value for the unemployed given in Shimer (2012) and the midpoint value for the employed in Hornstein et al. (2011): $\lambda = 0.118 = 0.4 \times 0.06 + 0.1 \times 0.94$.[17] Thus, our calibrated value for λ/δ is close to 4. As λ/δ increases, holding the other parameters fixed, labor market frictions are reduced and the model converges to a standard frictionless one. Hence, higher values of λ/δ reveal the implications of abstracting from labor market frictions and attributing all empirical income variation to talent dispersion (as is typically done in public finance). We consider such alternative values in our quantitative analysis. We set government spending $\mathcal{G}$ to equal 25% of equilibrium output.

[16]Hence, the actual and survival adjusted δ and $\tilde{\delta}$ and λ and $\tilde{\lambda}$ coincide.

[17]Hornstein et al. (2011) report an average unemployment rate over their sample of around 6%. The weight 0.06 used to compute λ corresponds to this. Allowing job finding rates to differ across employment and unemployment complicates the analysis by making the decision to accept a job dynamic: a worker must trade off a higher wage against the greater option value of unemployment. We leave this extension to future work.

Calibration of k The standard labor market model used in public finance assumes that differently talented workers supply labor in a frictionless market. The first order condition of these workers implies a one-to-one mapping between talent and income. This mapping can be inverted to allow the recovery of the underlying talent distribution from the empirically observed income distribution. In contrast our model does not imply a one-to-one mapping between talent and income. Rather the earnings density j is related to the talent density k via a mapping $j = \mathcal{T}(k; P)$, where:

$$j(x) = \mathcal{T}(k;P)(x) := \int_{\underline{\theta}(x;P)}^{\overline{\theta}(x;P)} \left\{ \frac{\delta}{2\lambda} \left(\frac{\overline{q}(\theta;P)}{\tilde{q}(\theta,x;P)} \right)^{\frac{1}{2}} \frac{1}{\tilde{q}(\theta,x;P)} \right\} k(\theta) d\theta, \tag{23}$$

$\tilde{q}$ is the iso-profit curve $\tilde{q}(\theta, x; P) = \theta(h')^{-1}(\theta(1 - T'[x])) - x$ and $P = (T, b)$. Inversion of $\mathcal{T}(\cdot; P)$ gives an empirical counterpart to k on $[\tilde{\theta}(P), \overline{\theta}]$.

The data source for our calibration exercise is the March release of the 2016 Current Population Survey (CPS).[18] We first form an affine approximation to current US government tax policy, $T[x] = -L + \tau x$, by regressing total taxes paid on labor income. Our estimates (with standard deviations in parentheses) are:

$$T[x] = \underset{(2.526)}{-302.56} + \underset{(0.000361)}{0.336}\; x.$$

The smallest active talent $\tilde{\theta}(P)$ is then determined by: $(1 - \hat{\tau}) - \frac{\underline{x}^{\gamma}}{\tilde{\theta}(P)^{1+\gamma}} = 0$, where $\hat{\tau} = 0.336$, $\underline{x}$ is the smallest income in our sample and γ equals its calibrated value. In addition, once $\tilde{\theta}(P)$ is determined, b is pinned down by:

$$\hat{L} + (1 - \hat{\tau})\underline{x} - \frac{1}{1+\gamma} \left(\frac{\underline{x}}{\tilde{\theta}(P)} \right)^{1+\gamma} = b, \tag{24}$$

where $\hat{L}$ equals the estimated value 302.56. In preceding sections b has been interpreted both as the value of inactivity to a worker and the direct cost of such inactivity to the policy maker. To better align the model with data, it is useful to distinguish these concepts. We (continue to) use b to label the per period value of inactivity to the worker, while denoting the transfer from the policymaker to the worker by b^U.[19] Given the values for T, $\tilde{\theta}$ and $\underline{x}$, an empirical counterpart for b can be recovered from the data using (24). A value for b^U is not needed to calibrate the talent distribution. Given empirical values for the tax policy parameters and

[18] Detailed description of the data and of our sample selection is given in Appendix C.

[19] The difference $b - b^U$ could be positive because the worker derives additional utility benefit from inactivity, has additional time to engage in home production, or receives transfers from other agents. Alternatively, it could be negative because there is a stigma attached to not working.

for b, empirical counterparts for the functions $\overline{q}(\cdot;P)$, $\tilde{q}(\cdot;P)$, $\underline{\theta}(\cdot;P)$ and $\overline{\theta}(\cdot;P)$ may be constructed using the relevant formulas from preceding sections. In addition, $\underline{x}(\cdot;P)$ and $\overline{x}(\cdot;P)$, the inverses of $\underline{\theta}(\cdot;P)$ and $\overline{\theta}(\cdot;P)$, may be obtained.

Let $\{x_m\}_{m=0}^M$ and $\{\theta_n\}_{n=0}^N$ denote fine grids of labor income and talent values and let $\mathbf{j} = \{j_m\}_{m=1}^M = \{J(x_m) - J(x_{m-1})\}_{m=1}^M$ and $\mathbf{k} = \{k_n\}_{n=1}^N = \{K(\theta_n) - K(\theta_{n-1})\}_{n=1}^N$ denote quantiles of the income and talent distributions. We recover estimates of the quantiles of the earnings distribution, $\mathbf{j}$, from our data set. Equation (23) implies the following approximate map between such $\mathbf{j}$ and the quantiles $\mathbf{k}$ of the talent distribution:

$$j_m \approx \sum_{n=1}^N \left\{ \int_{\underline{x}_{m-1}(\theta_n;P)}^{\overline{x}_m(\theta_n;P)} \left\{ \frac{\delta}{2\lambda} \left(\frac{\overline{q}(\theta_n)}{\tilde{q}(\theta_n,x)} \right)^{\frac{1}{2}} \frac{1}{\tilde{q}(\theta_n,x)} \right\} dx \right\} k_n,$$

where $\overline{x}_m(\theta;P) = \min\{x_m, \overline{x}(\theta;P)\}$, and $\underline{x}_m(\theta;P) = \max\{x_{m-1}, \underline{x}(\theta;P)\}$. Thus, we focus on the linear equations:

$$\mathbf{j} = \mathbf{T}\mathbf{k},$$

where $\mathbf{T}$ is an $M \times N$ matrix with (m,n)-th element:

$$\mathbf{T}(m,n) = \int_{\underline{x}_{m-1}(\theta_n;P)}^{\overline{x}_m(\theta_n;P)} \left\{ \frac{\delta}{2\lambda} \left(\frac{\overline{q}(\theta_n)}{\tilde{q}(\theta_n,x)} \right)^{\frac{1}{2}} \frac{1}{\tilde{q}(\theta_n,x)} \right\} dx.$$

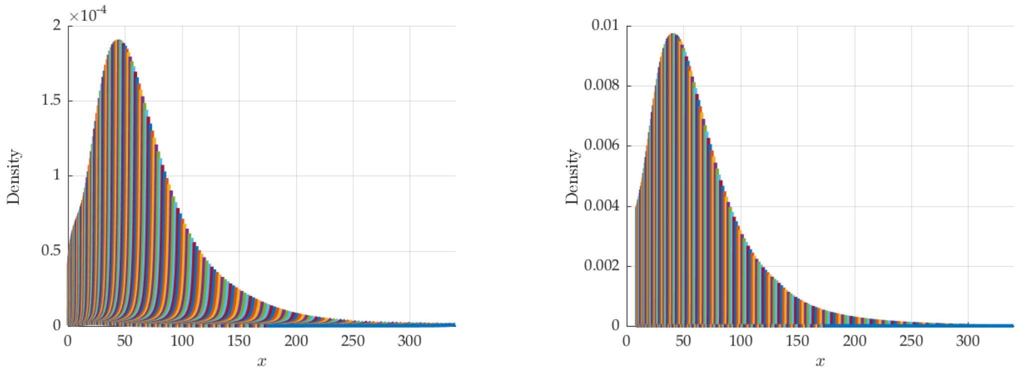
We calibrate $\mathbf{k}$ by selecting a value $\hat{k}$ that minimizes the sum of squares:

$$\min_{\mathbf{k}} \sum_{m=1}^M (\mathbf{j}(m) - \mathbf{T}(m,\cdot)\cdot\mathbf{k})^2.$$

This procedure provides values for $\{\mathbf{k}_n\}_{\tilde{n}+1}^N$, where $\tilde{n}+1$ denotes the least talented working quantile. In our model, $\theta_{\tilde{n}+1}$ is determined by the closest grid point in excess of $\tilde{\theta}(P)$. The left tail of the distribution is effectively censored: we observe the mass, but not the conditional talent distribution of workers who do not work. Thus, the only restriction on $\{\mathbf{k}_n\}_{n=1}^{\tilde{n}}$ is that $\sum_{n=1}^{\tilde{n}} k_n$ equals the mass of non-working agents. We apply a kernel density smoother to the part of the distribution we observe and fit a Pareto right tail. We extrapolate the density and fit a left tail.[20]

For our baseline calibration, Figure 3(a) shows the fitted income densities within each talent market as colored curves, i.e. it shows for different n the income densi-

[20]We have conducted sensitivity analysis with respect to the left tail of the talent distribution and considered various fitted left tails. This does not impact the choice of optimal affine tax rate very much. It does affect optimal nonlinear tax rates at low incomes.



(a) $\frac{\lambda}{\delta} = 0.118/0.03$

(b) $\frac{\lambda}{\delta} = 100$

Figure 3: Income densities for alternative values of the parameter describing labor frictions λ/δ and $\gamma = 1$.

ties $\frac{\delta}{2\lambda}\left(\frac{\bar{q}(\theta_n)}{\tilde{q}(\theta_n,x)}\right)^{\frac{1}{2}}\frac{k(\theta_n)}{\tilde{q}(\theta_n,x)}$, where $k(\theta_n)$ is the calibrated talent density value. Figure 3(b) repeats the picture for the case $\lambda/\delta = 100$. In this latter case job finding probabilities are high relative to job destruction probabilities and the economy is close to frictionless. Most workers are concentrated at the top of talent market job ladders and intra-talent market densities are relatively large and place most mass in a neighborhood of the highest incomes. They resemble a series of spikes whose heights closely track the shape of the empirical income histogram. In contrast in our baseline case, λ/δ is much lower and workers are more dispersed over incomes within taken markets. The intra-talent market densities have a J-like shape. Figure 4 shows the calibrated talent densities for these cases. The densities are

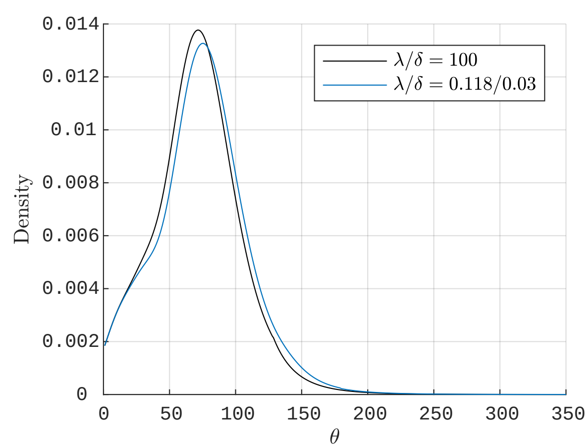


Figure 4: Calibrated talent densities for alternative values of the parameter describing labor frictions λ/δ and $\gamma = 1$.

similar across the cases, with some rightward shifting of mass in our benchmark frictional case relative to the near frictionless case.[21]

Quantitative results Table 1 reports results for a range of values for the frictions ratio λ/δ and for two values of γ.[22] Optimal marginal tax rates are unsurprisingly greater at higher values for γ (and, correspondingly, lower values of the effort supply elasticity). Of more interest is the monotonicity of optimal tax rates in λ/δ. At our benchmark value of the frictions parameter λ/δ optimal tax rates are lower than in the (near) frictionless limit. Table 2 provides perspective on this result. It reports the values of the various terms in (19) (at $\gamma = 1$ and with $\rho = 1$). These terms decompose the marginal benefits and costs of a small change in τ around the optimum. The terms in the last two rows of the table are associated with the profit squeeze channel. Recall that an increase in taxes diverts profits and profit revenues towards labor incomes and labor income tax revenues. The covariance term

Table 1: Optimal Affine Tax Policy

Variable	$\mathcal{G} = 0.25$ GDP, $\gamma = 1$			$\mathcal{G} = 0.25$ GDP, $\gamma = 2$		
	$\frac{\lambda}{\delta} \approx 4$	$\frac{\lambda}{\delta} = 50$	$\frac{\lambda}{\delta} = 100$	$\frac{\lambda}{\delta} \approx 4$	$\frac{\lambda}{\delta} = 50$	$\frac{\lambda}{\delta} = 100$
τ^*	30.4	32.8	35.0	44.8	47.3	48.4
L^*	162	259	334	714	812	827
b^*	749	727	699	1102	1054	1003
π^*	263	128	15	346	174	20

Notes: L^*, b^* and π^* are monthly 2015 US $ amounts. π^* is per capita monthly profit.

Table 2: Optimal Affine Tax Policy

Variable	$\mathcal{G} = 0.25$, $\gamma = 1$		
	$\frac{\lambda}{\delta} \approx 4$	$\frac{\lambda}{\delta} = 10$	$\frac{\lambda}{\delta} = 100$
$-\text{Cov}^*(U', x)$	2.3×10^{-3}	2.1×10^{-3}	1.9×10^{-3}
$-E^*[U'] \frac{\mathcal{E}^*_{E[x]}\tau^*}{1-\tau^*}$	-1.6×10^{-3}	-1.8×10^{-3}	-1.9×10^{-3}
$\text{Cov}^*(U', \mathcal{E}_{\bar{q}} q)$	-3.4×10^{-4}	-9.2×10^{-5}	-3.1×10^{-6}
$-E^*[U'] \frac{\mathcal{E}^*_{\pi}\tau^*\pi^*}{1-\tau^*}$	-2.1×10^{-4}	-9.7×10^{-5}	-1.1×10^{-5}

[21] Further analytical characterization of the relationship between income and talent distributions is given in Appendix F.

[22] For each alternative parameter value, the talent distribution is recalibrated.

in the third row captures the distributional benefit of diverting profit tax revenues towards after-tax labor incomes. The term in the fourth row nets out the cost of lost profit tax revenues that are diverted towards income tax revenues.[23] Both of the third and fourth row terms are negative indicating that they introduce an additional marginal cost of raising taxes and are force for lower taxes. However, as λ/δ rises these terms become small implying that as frictions become less important this force is diluted. The negative sign of the covariance between marginal utility and $\mathcal{E}_{\bar{q}}q$ in the third row of Table 2 implies that an increase in tax rates that suppresses profits and profit tax revenues is undesirable on redistributional grounds. Decomposition of this covariance term at the benchmark $\frac{\lambda}{\delta} \approx 4$ reveals:

$$\underbrace{\text{Cov}^*(U', \mathcal{E}_{\bar{q}}q)}_{-3.38\times 10^{-4}} = \underbrace{E^*[\text{Cov}^*(U', \mathcal{E}_{\bar{q}}q|\theta)]}_{7.69\times 10^{-5}} + \underbrace{\text{Cov}^*(E^*[U'|\theta], E^*[\mathcal{E}_{\bar{q}}q|\theta])}_{-4.15\times 10^{-4}}. \tag{25}$$

The first term captures the redistributional benefits of reducing profits within talent markets and the second the redistributional benefits of reducing profits across markets. The former is positive capturing the fact that the lowest paid in a talent market have more profit extracted from them. However, the second term is negative. Highly talented workers at the bottom of job ladders work harder and earn more than less talented ones. An increase in tax rates consequently leads to a greater mechanical reduction in the income of the former relative to the latter. Bottom-of-the-job-ladder firms must correspondingly reduce their profit offers further in high talent markets. Competition between firms propagates these larger reductions in profit offers up the job ladder in high talent markets. The overall effect is that profit suppression is stronger in high talent markets and redistribution from low to high talents occurs. This effect is captured by the second term in (25) and it is the dominating one.

6 Nonlinear Taxation

In this section the environment from the previous sections is modified to permit the policymaker to select a twice continuously differentiable non-linear tax function $T : \mathbb{R}_+ \to \mathbb{R}$ applied to the earnings of working agents. Preliminary concepts and notation are introduced first. To simplify the presentation $\beta = \rho = 1$ (so that $\tilde{\delta} = \delta$

[23]Recall that although these profit tax revenues are recaptured as higher income tax receipts, the benefits of the latter have already been incorporated into the second row terms. To avoid double counting the cost of this extra reduction profit tax revenues must be accounted for.

and $\tilde{\lambda} = \lambda$) and $\overline{\theta} = \infty$ is assumed.

Tax perturbations and elasticities A working agent's first order condition in the nonlinear tax setting is:

$$1 - T'[x(q,\theta;T)] = \frac{1}{\theta} h'\left(\frac{x(q,\theta;T)+q}{\theta}\right), \tag{26}$$

with $x(q,\theta;T)$ the worker's optimal income choice at tax function T. Consider a perturbation to this tax function $\tilde{T}[x] = T[x] + \kappa\Omega(x) = T[x] + \kappa\int_0^x \omega(x')dx'$, where ω is a smooth function. A worker confronted with this perturbed function has first order condition:

$$(1 - T'[x(q,\theta;T+\kappa\Omega)] - \kappa\omega(x(q,\theta;T+\kappa\Omega))) = \frac{1}{\theta} h'\left(\frac{x(q,\theta;T+\kappa\Omega)+q}{\theta}\right). \tag{27}$$

Assuming that $x(q,\theta;\cdot)$ is locally differentiable in κ around 0 (i.e. that the conditions of the implicit function theorem hold), differentiating (27) with respect to κ, rearranging and taking the limit as κ goes to 0 gives:

$$-\left.\frac{\partial x(q,\theta;T+\kappa\Omega)}{\partial\kappa}\right|_{\kappa=0} = \frac{x(q,\theta;T)\omega(x(q,\theta;T))}{1 - T'[x(q,\theta;T)]}\mathscr{E}(q,\theta;T), \tag{28}$$

where $\mathscr{E}(q,\theta;T)$ is the elasticity of income with respect to the tax perturbation:

$$\mathscr{E}(q,\theta;T) := \frac{1}{\frac{x(q,\theta;T)}{x(q,\theta;T)+q}\frac{\frac{x(q,\theta;T)+q}{\theta}h''\left(\frac{x(q,\theta;T)+q}{\theta}\right)}{h'\left(\frac{x(q,\theta;T)+q}{\theta}\right)} + \frac{T''[x(q,\theta;T)]x(q,\theta;T)}{1-T'[x(q,\theta;T)]}}. \tag{29}$$

Note that this differs from $\mathcal{E}(q,\theta;T) = \frac{\theta}{x(q,\theta;T)}\frac{h'(x(q,\theta;T)/\theta)}{h''(x(q,\theta;T)/\theta)}$, the conventional wage elasticity of effort (or tax elasticity of income), because workers do not receive all of the surplus they create and because taxes are non-linear.[24]

The income distribution The workers' income choice problems imply that all workers earning income x must share the same marginal rate of substitution between income and consumption:

$$\frac{1}{\theta} h'\left(\frac{x+\tilde{q}(\theta,x;T)}{\theta}\right) = 1 - T'[x],$$

[24] Non-linearity of taxes also ensures that worker effort choice depends on q as well as θ. See (26).

where $\tilde{q}(\theta, x; T)$ gives the profit extracted from a worker of type θ, earning x under tax system T. Thus, $\tilde{q}(\cdot, \cdot; T)$ defines a family of iso-income curves in (q, θ) space consistent with worker optimization and satisfying:

$$\tilde{q}(\theta, x; T) = \theta (h')^{-1}(\{1 - T'[x]\}\theta) - x. \tag{30}$$

The distribution of firms and workers across (q, θ) is determined analogously to the affine tax case using the steady state labor flow and firm profit maximization conditions. In particular, let $\underline{q}(\theta; T)$ and $\overline{q}(\theta; T)$ denote the minimal and maximal profit-per-worker offers made to type θ workers given tax function T. As in the affine tax case, $\overline{q}(\theta; T)$ is the maximal profit per worker such that θ-types are indifferent between accepting a job and remaining unemployed:

$$x(\overline{q}(\theta; T), \theta; T) - T[x(\overline{q}(\theta; T), \theta; T)] - h\left(\frac{x(\overline{q}(\theta; T), \theta; T) + \overline{q}(\theta; T)}{\theta}\right) = b. \tag{31}$$

Further, $\underline{q}(\theta; T) = (\frac{\delta}{\delta+\lambda})^2 \overline{q}(\theta; T)$. For $\tilde{q}(\theta, x; T)$ to be consistent with firm profit maximization, it must be that $\tilde{q}(\theta, x; T) \in [\underline{q}(\theta; T), \overline{q}(\theta; T)]$. Lemma 1 further characterizes $\tilde{q}(\cdot, x; T)$ (on a domain consistent with firm profit maximization).

Lemma 1. *$\tilde{q}(\cdot, x; T)$ is increasing in θ between two endpoints $\underline{\theta}(x; T)$ and $\overline{\theta}(x, T)$.*

Proof. See Appendix B. □

The upper end point $\overline{\theta}(x; T)$ in Lemma 1 is determined by the talent such that the iso-income curve $\tilde{q}(\cdot, x; T)$ reaches the upper bound on profits $\overline{q}(\cdot; T)$, i.e. $\tilde{q}(\overline{\theta}(x; T), x; T)$ $= \overline{q}(\overline{\theta}(x; T); T)$. Combining formulas (30) and (31) gives:

$$\overline{\theta}(x; T) = \frac{1}{1 - T'[x]} h'(h^{-1}\{x - T[x] - b\}).$$

The lower end point $\underline{\theta}(x; T)$ in Lemma 1 is such that the iso-income curve $\tilde{q}(\cdot, x; T)$ reaches the lower bound on profits $\underline{q}(\cdot; T)$, i.e. $\tilde{q}(\underline{\theta}(x; T), x; T) = \underline{q}(\underline{\theta}(x; T); T) = (\frac{\delta}{\delta+\lambda})^2 \overline{q}(\underline{\theta}(x, T); T)$. Let $\tilde{\theta}(T)$ denote the threshold talent such that all firms make zero profits per worker: $\overline{q}(\tilde{\theta}(T); T) = \underline{q}(\tilde{\theta}(T); T) = 0$. Figure 5 illustrates the situation. Talents below the threshold $\tilde{\theta}(T)$ do not work. A fraction $\frac{\lambda}{\lambda+\delta}$ of talents θ above this threshold do work and receive profit-per worker offers between $\underline{q}(\theta; T)$ and $\overline{q}(\theta; T)$. Workers earning income x have talents between $\underline{\theta}(x; T)$ and $\overline{\theta}(x; T)$ and profit-per-worker offers along the iso-income schedule $\tilde{q}(\cdot, x; T)$.

Let $J[\cdot; T]$ denote the income distribution given tax function T. The counter cumulative $1 - J[x; T]$ equals the mass of employed workers between the $\overline{q}(\theta; T)$ and

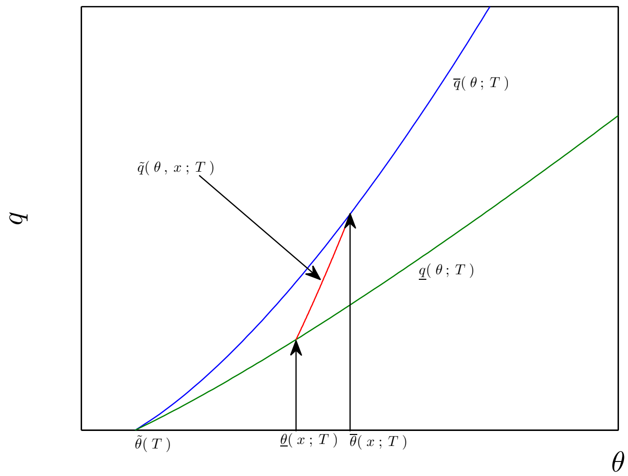


Figure 5: Talent, profit-per-worker and income combinations in equilibrium.

$\underline{q}(\theta;T)$ curves and to the right of the $\tilde{q}(x,\theta;T)$ curve. This is given by:

$$1 - J[x;T] := \int_{\underline{\theta}(x;T)}^{\overline{\theta}(x;T)} \int_{\underline{q}(\theta;T)}^{\tilde{q}(\theta,x;T)} g(q|\theta;T)dqk(\theta)d\theta + \int_{\overline{\theta}(x;T)}^{\infty} \int_{\underline{q}(\theta;T)}^{\overline{q}(\theta;T)} g(q|\theta;T)dqk(\theta)d\theta.$$

Let $j(\cdot;T)$ denote the corresponding income density.

Policymaker's problem The policymaker's nonlinear tax policy problem (with $\beta = \rho = 1$) is:

$$\sup_{T,b} \left(\frac{\delta}{\delta+\lambda} + \frac{\lambda}{\delta+\lambda} K[\tilde{\theta}(T)]\right) U(b) + \frac{\lambda}{\delta+\lambda} \int_{\tilde{\theta}(T)}^{\infty} \int_{\underline{q}(\theta;T)}^{\overline{q}(\theta;T)} U\left(\Phi(q,\theta;T)\right) g(q|\theta;T)dqk(\theta)d\theta. \tag{32}$$

subject to:

$$\mathcal{G} + \left(\frac{\delta}{\delta+\lambda} + \frac{\lambda}{\delta+\lambda} K[\tilde{\theta}(T)]\right) b = \frac{\lambda}{\lambda+\delta} \int_{\tilde{\theta}(T)}^{\infty} \int_{\underline{q}(\theta;T)}^{\overline{q}(\theta;T)} \left\{\tilde{q} + T[x(q,\theta;T)]\right\} g(q|\theta;T)dqk(\theta)d\theta,$$

where:

$$\Phi(q,\theta;T) = x(q,\theta;T) - T[x(q,\theta;T)] - h\left(\frac{x(q,\theta;T)+q}{\theta}\right).$$

We characterize the solution to (32) by analyzing perturbations of a tax function around an optimum. Exact definitions of the perturbations used and details of the analysis are given in Appendix B.[25] Theorem 1 summarizes the implications of a small increase in the marginal tax rate "at" an income x_0. We again use the * superscript to denote evaluations at an optimal tax function T^*, e.g. $\underline{\theta}^*(x_0) = \underline{\theta}(x_0, T^*)$, $\overline{\theta}^*(x_0) = \overline{\theta}(x_0, T^*)$ and so forth.

Theorem 1. *Let $\hat{q}^*(x_0, \theta) = \tilde{q}^*(x_0, \theta)$ if $\theta \in [\underline{\theta}^*(x_0), \overline{\theta}^*(x_0)]$ and $\hat{q}^*(x_0, \theta) = \overline{q}^*(x_0, \theta)$ if $\theta \in (\overline{\theta}(x_0, T^*), \infty)$ and let $\overline{\mathscr{E}}^*(x_0)$ denote the average value of the elasticity $\mathscr{E}^*(q, \theta)$ amongst those workers earning x_0. For x_0 such that $\underline{\theta}^*(x_0) > \tilde{\theta}^*$, the optimal nonlinear tax function satisfies the following equation:*

$$\begin{aligned}
&- \int_{\underline{\theta}^*(x_0)}^{\infty} \int_{\underline{q}^*(\theta)}^{\hat{q}^*(\theta, x_0)} U'\left(\Phi^*(q, \theta)\right) g^*(q|\theta) k(\theta) dq d\theta \\
&\quad + \int_{\overline{\theta}^*(x_0)}^{\infty} \frac{\mathcal{E}^*_{\overline{q}}(\theta)}{1 - T^{*\prime}[x_0]} \int_{\underline{q}^*(\theta)}^{\overline{q}^*(\theta)} U'\left(\Phi^*(q, \theta)\right) (1 - T^{*\prime}[x^*(q, \theta)]) q g^*(q|\theta) dq k(\theta) d\theta \\
&\qquad + E^*[U'(\Phi^*)] \Bigg\{ (1 - J^*[x_0]) - \frac{x_0 T^{*\prime}[x_0]}{1 - T^{*\prime}[x_0]} \overline{\mathscr{E}^*}(x_0) j^*(x_0) \\
&\qquad\quad + \int_{\overline{\theta}^*(x_0)}^{\infty} \frac{\mathcal{E}^*_{\overline{q}}(\theta)}{1 - T^{*\prime}[x_0]} \int_{\underline{q}^*(\theta)}^{\overline{q}^*(\theta)} T^{*\prime}[x^*(q, \theta)] \left\{ \frac{x^*(q, \theta) + q}{x^*(q, \theta)} \frac{\mathscr{E}^*(q, \theta)}{\mathcal{E}^*(q, \theta)} \right\} q g^*(q|\theta) dq k(\theta) d\theta \\
&\qquad\quad - \int_{\overline{\theta}^*(x_0)}^{\infty} \frac{\mathcal{E}^*_{\overline{q}}(\theta)}{1 - T^{*\prime}[x_0]} \int_{\underline{q}^*(\theta)}^{\overline{q}^*(\theta)} q g^*(q|\theta) dq k(\theta) d\theta \Bigg\} = 0. \qquad (33)
\end{aligned}$$

Proof. See Appendix B. □

Interpreting Theorem 1 In the standard Mirrlees setting, a local tax perturbation in the marginal tax rate around the optimum at an income x_0 has three effects. First, it leads to the mechanical collection of extra (socially valued) tax revenues from all workers earning more than x_0. Second, it directly reduces the utility of those who earn more than x_0 and, thus, pay more tax. Third, it induces a small intensive margin behavioral response from workers earning x_0. This lowers income tax revenues, but, since workers are optimizing before and after the tax change, this behavioral response has no additional direct impact on their welfare. Optimal policy trades off these effects. All of them are present in our model, but they are supplemented with effects stemming from the profit squeeze channel. Details are given below.

[25]From a mechanism design perspective, working agents have two characteristics θ and q that are not observed by the designer. In addition the distribution of one of these characteristics q is endogenous and shaped by the mechanism. In this setting direct tax perturbations provide a more straightforward characterization of the policymaker's problem than the mechanism design approach.

Impact on revenues The tax perturbation causes each worker earning more than x_0 to pay an extra dollar in tax. This leads to a mechanical increase in tax revenues equal to (after normalization by the fraction of agents working) $1 - J^*[x_0]$. In addition, θ-talent workers earning x_0 slightly reduce their effort and, hence, incomes by:

$$\frac{x_0}{1 - T^{*\prime}[x_0]} \mathscr{E}^*(\tilde{q}^*(x_0, \theta), \theta), \tag{34}$$

where $\mathscr{E}^* = \mathscr{E}(\cdot; T^*)$ is defined as in (29). This leads to a reduction in tax revenues collected from these workers of $\frac{x_0 T^{*\prime}[x_0]}{1 - T^{*\prime}[x_0]} \overline{\mathscr{E}}^*(x_0) j^*(x_0)$, where $\overline{\mathscr{E}}^*(x_0)$ is the average elasticity of income with respect to taxes amongst x_0 income earners.

The tax perturbation also has an impact on the profit offers received by workers earning x_0 and above. Specifically, the lowest paid workers with talents $\theta > \overline{\theta}^*(x_0)$ earn more than x_0 and, so, pay more tax (on the infra-marginal x_0-th dollar that they earn). Consequently, the maximal profit consistent with such a worker accepting a job, $\overline{q}^*(\theta)$, must be reduced. This is illustrated in Figure 6 by a downward pivot of $\overline{q}(\theta)$ at talents above $\overline{\theta}^*(x_0)$. Competition between firms propagates this reduced

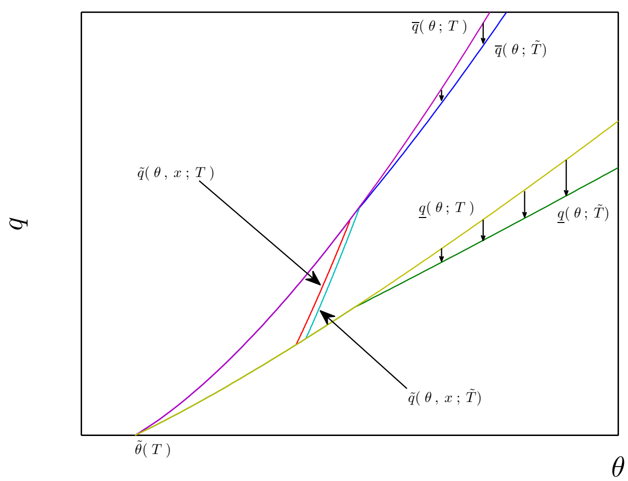


Figure 6: Impact of a Tax Perturbation at income x. The rise in taxes paid by workers earning more than x causes the maximal and minimal profit curves to pivot downwards.

profit offer up through the job ladder in these markets. In this way profits are diverted towards the pre-tax incomes of those earning more than x_0. The adverse effect of this on profit tax revenues is given by the term on the fifth line of (33), while the positive effect on income tax revenues is given by the term on the fourth line. The shadow value of tax revenues remains $E^*[U'(\Phi)]$ and together the terms on the third, fourth and fifth lines of (33) give the overall impact on revenues.

Direct impact on social welfare Workers earning more than x_0 pay additional taxes and this reduces their consumption and utility. The welfare cost of this is given by the first term in (33). In the standard Mirrlees model, this is the only direct welfare effect of taxation. However, in our setting there is an additional effect. As described previously, profit extracted from workers with talents greater than or equal to $\overline{\theta}^*(x_0)$ falls. The income and, hence, utility of these workers rises and this is captured by the term on the second line of (33).[26]

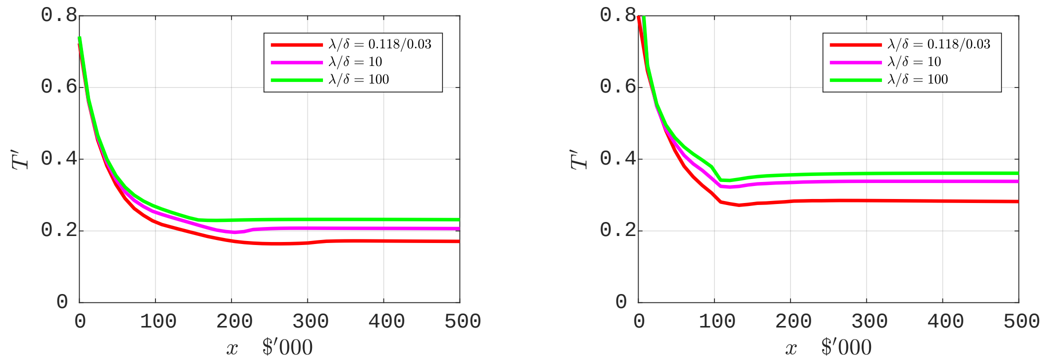
7 Quantitative Analysis of Optimal NonLinear Taxes

This section reports quantitative optimal nonlinear tax results. The benchmark parameter values and calibration procedure described in Section 5 is used. Figure 7 plots optimal tax rates at γ values of 1 and 2 for the baseline λ/δ value of $0.118/0.03 \approx 4$ and larger values of 10 and 100.[27] Optimal marginal tax rates are similar across λ/δ values for incomes below approximately $40,000, but rise moderately with λ/δ at incomes above this amount. Hence, an analysis that abstracted from labor market frictions and attributed all or most income variation to talent variation would prescribe greater than optimal marginal tax rates over a wide range of middle and higher incomes.

Evaluating tax wedges at the optimum To better understand the forces behind this result, it is convenient to use (33) to decompose marginal taxes into component

[26] Note that the presence of the profit-per-worker q variable in the integral indicates that this increase is more pronounced amongst workers from whom more profit is extracted, i.e those at the bottom of the job ladder. In this sense intra-talent market redistribution occurs in talent markets above $\overline{\theta}^*(x_0)$. Inter-talent market redistribution is present in the sense that only those with talent above $\overline{\theta}^*(x_0)$ experience increases in income and utility, those below this talent threshold do not.

[27] At each λ/δ optimal policy calculation, we recalibrate the talent distribution. Thus, Figure 7 gives the policy prescriptions implied by analyses that consistently assume frictions are at our benchmark value or are at smaller and, perhaps, much smaller values. An alternative exercise is to hold the talent distribution fixed and calculate the policy implications of changing λ/δ. This exercise gives similar results largely because the impact of λ/δ on the calibrated talent density is relatively modest. In Appendix F we further characterize the map $\mathcal{T}$ between the observed income and unobserved talent distribution. In particular, we show that if the talent distribution has a Paretian right tail, then the map between its Pareto coefficient and the limiting Pareto coefficient of the income distribution is not affected by λ/δ.



(a) $\gamma = 1$ (b) $\gamma = 2$

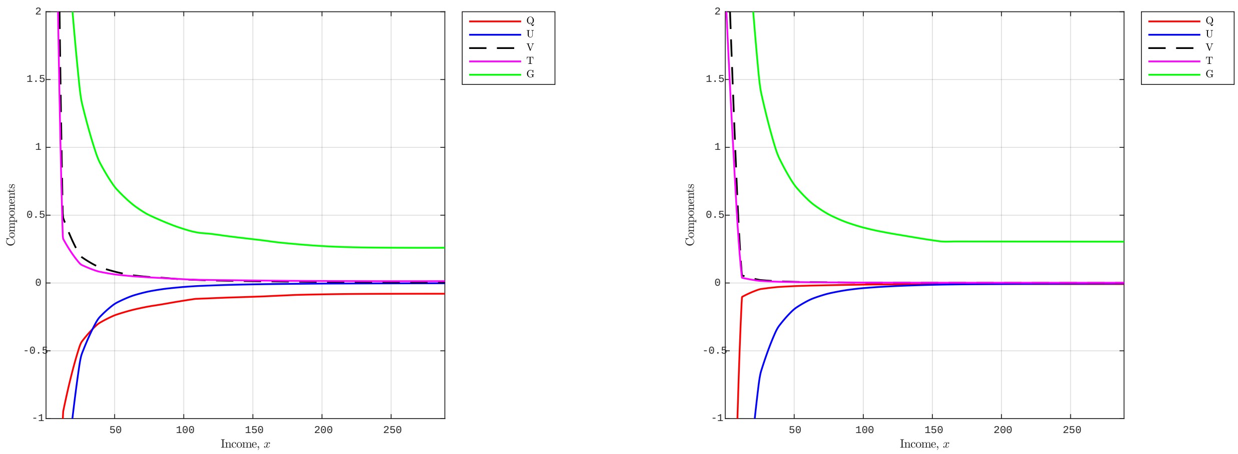
Figure 7: Marginal Tax Rates.

"wedges". As a first step, (33) may be rearranged as:

$$
\begin{aligned}
\frac{T^{*\prime}[x_0]}{1-T^{*\prime}[x_0]} = & \underbrace{\frac{1}{\overline{\mathscr{E}^*}(x_0)}\frac{1-J^*[x_0]}{x_0 j^*(x_0)}}_{G} \\
& \underbrace{-\frac{1}{\mathcal{L}}\int_{\underline{\theta}^*(x_0)}^{\infty}\int_{\underline{q}^*(\theta)}^{\hat{q}^*(\theta,x_0)} U'\left(\Phi^*(q,\theta)\right) g^*(q|\theta)k(\theta)dqd\theta}_{U} \\
& \underbrace{+\frac{1}{\mathcal{L}}\int_{\overline{\theta}^*(x_0)}^{\infty}\frac{\mathcal{E}^*_{\overline{q}}(\theta)}{1-T^{*\prime}[x_0]}\int_{\underline{q}^*(\theta)}^{\overline{q}^*(\theta)} U'\left(\Phi^*(q,\theta)\right)(1-T^{*\prime}[x^*(q,\theta)])qg^*(q|\theta)dqk(\theta)d\theta}_{T} \\
& \underbrace{+\frac{1}{\mathcal{M}}\int_{\overline{\theta}^*(x_0)}^{\infty}\frac{\mathcal{E}^*_{\overline{q}}(\theta)}{1-T^{*\prime}[x_0]}\int_{\underline{q}^*(\theta)}^{\overline{q}^*(\theta)} T^{*\prime}[x^*(q,\theta)]\mathscr{R}(q,\theta)qg^*(q|\theta)dqk(\theta)d\theta}_{V} \\
& \underbrace{-\frac{1}{\mathcal{M}}\int_{\overline{\theta}^*(x_0)}^{\infty}\frac{\mathcal{E}^*_{\overline{q}}(\theta)}{1-T^{*\prime}[x_0]}\int_{\underline{q}^*(\theta)}^{\overline{q}^*(\theta)} qg^*(q|\theta)dqk(\theta)d\theta.}_{Q} \qquad (35)
\end{aligned}
$$

where: $\mathcal{L} := \overline{\mathscr{E}^*}(x_0)x_0j^*(x_0)E^*[U'(\Phi^*)]$, $\mathcal{M} := \overline{\mathscr{E}^*}(x_0)x_0j^*(x_0)$ and $\mathscr{R} := \frac{x^*(q,\theta)+q}{x^*(q,\theta)}\frac{\mathscr{E}^*(q,\theta)}{\mathcal{E}^*(q,\theta)}$. The first left hand side term (labelled G) is a conventional term that appears in optimal tax equations derived under the assumption of frictionless labor markets (e.g. Saez (2001)). It equals (the reciprocal of) the product of the income Pareto coefficient and the elasticity of income with respect to a local marginal tax perturbation at x_0. The second term (labelled U) also appears in the frictionless setting and captures the social cost of redistributing away from those earning more than

x_0. The remaining terms are those arising in frictional economies. They decompose the social marginal cost of the tax induced redistribution from profits and towards incomes on, respectively, the utilities of those earning more than x_0 (term T), the income tax revenues collected from those earning more than x_0 (term V) and profit tax revenues (term Q).



(a) $\frac{\lambda}{\delta} = 0.118/0.03$

(b) $\frac{\lambda}{\delta} = 100$

Figure 8: Optimal Tax Wedges. The wedge terms Q, U, V, T and G in the figure correspond to the terms in equation (35).

Figure 8(a) shows the values of these terms in our baseline λ/δ case (and for $\gamma = 1$). The green curve is the conventional tax distortion term G. This inherits the shape of the (optimal) income Pareto coefficient reciprocal, which initially falls and then stabilizes. The blue curve is the conventional term U, which is initially large (in absolute terms) and negative. However, the term rapidly decays to zero, reflecting the low social marginal cost of redistributing from the average high income earner (under our assumed social criterion).[28] In a frictionless model only these terms would be present and at high incomes the former would shape the optimal marginal tax function. The additional terms introduced by the model with frictions correspond to the dashed black V, pink T and red Q curves. The figure indicates that these terms are large at incomes below \$30,000. In combination, however, they nearly net out and have a small overall impact on taxes at these lower incomes. As income rises the T and V terms decay to very small numbers: the social

[28]In our model there are some high earners at the bottom of job ladders who have low current payoffs and high marginal utilities. These workers have more extracted from them in the form of profit and must exert correspondingly higher effort. However, most high earners are not at the bottom of job ladders and average marginal utility conditional on income decreases with income.

benefit from diverting profit towards the income of high earners is very small (term T), while the corresponding income tax gain is also very small in part because the tax rate declines (term V). In contrast, the marginal cost of lost profit tax revenues captured by Q, while initially becoming smaller, asymptotes to a negative number. This implies that loss of profit revenues remains an inhibiting factor on tax rates at high incomes. Figure 8(b) shows that, when $\frac{\lambda}{\delta}$ is assumed equal to 100 and labor market frictions are correspondingly assumed small, the frictional terms V, T and Q are negligible at all but the very lowest incomes. In particular, the Q term now decays to zero: in near frictionless economies loss of profit revenues no longer restrains taxes on high income earners.

8 Conclusion

Most analysis in public finance is conducted in the context of a frictionless labor market in which all worker income variation is attributable to variations in talent and intensive margin labor supply choices. In contrast, large areas of macro and structural labor are cast in terms of search models. Increasingly these models emphasize on as well as off the job search. Our paper advances the policy design literature by deriving implications of off and on the job search frictions for tax design. In our setting a worker's pay depends upon her marketable talent and the extractiveness of her employer. Variation in the latter creates job ladders as workers search on the job for better firms offering higher wages. In this setting we highlight a novel "profit squeeze channel". Higher marginal income tax rates squeeze firm profits and, hence, raise the share of the surplus captured by workers. Through this channel they lower profits and profit tax revenues on the one hand, while raising worker incomes and income tax revenues on the other (relative to a standard public finance model that omits this channel). Such profit squeezing is socially desirable and a motive for higher income tax rates if it relatively benefits those at the bottom of job ladders from whom more profit is extracted (or, more generally, those with higher marginal social welfare weights). But it is undesirable if it primarily benefits highly talented workers earning higher incomes. Theory highlights the tradeoffs, but it is ambiguous about the overall implications of this channel for policy. Quantitative analysis suggests that it works in the direction of moderately lower marginal tax rates over a large range of incomes in the nonlinear setting and a lower overall tax rate in the affine setting.

We have explored the role of the profit squeeze channel in a particular (well known) frictional setting. But it is more general and would emerge in other frictional models in which higher income taxes (in combination with a given or policy determined outside option for workers) squeeze firm profits and different amounts of profits are collected from different workers. The latter might occur because workers are heterogeneous with respect to mobility or replaceability or because firms are heterogeneous with respect to their bargaining ability. We anticipate that our qualitative findings would be robust to these alternative settings, but leave exploration of this to future work.

References

Ales, L., M. Kurnaz, and C. Sleet (2015). Technical change, wage inequality and taxes. *American Economic Review 105*(10), 3061–3101.

Ales, L. and C. Sleet (2016). Taxing top CEO incomes. *American Economic Review 106*(11), 3331–3366.

Bontemps, C., J.-M. Robin, and G. J. Van den Berg (2000). Equilibrium search with continuous productivity dispersion: Theory and nonparametric estimation. *International Economic Review 41*(2), 305–358.

Boone, J. and L. Bovenberg (2002). Optimal labour taxation and search. *Journal of Public Economics 85*(1), 53–97.

Boone, J. and L. Bovenberg (2006). Optimal welfare and in-work benefits with search unemployment and observable abilities. *Journal of Economic Theory 126*(1), 165–193.

Burdett, K. and D. T. Mortensen (1998). Wage differentials, employer size, and unemployment. *International Economic Review 39*(2), 257–273.

Golosov, M., P. Maziero, and G. Menzio (2013). Taxation and redistribution of residual income inequality. *Journal of Political Economy 121*(6), 1160–1204.

Hornstein, A., P. Krusell, and G. L. Violante (2011). Frictional wage dispersion in search models: A quantitative assessment. *The American Economic Review 101*(7), 2873–2898.

Hungerbühler, M., E. Lehmann, A. Parmentier, and B. Van der Linden (2006). Optimal redistributive taxation in a search equilibrium model. *The Review of Economic Studies 73*(3), 743–767.

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), 1523–1537.

Lise, J., C. Meghir, and J.-M. Robin (2016). Matching, sorting and wages. *Review of Economic Dynamics 19*, 63–87.

Mirrlees, J. (1971). An exploration in the theory of optimum income taxation. *Review of Economic Studies 38*(2), 175–208.

Postel-Vinay, F. and J.-M. Robin (2002). Equilibrium wage dispersion with worker and employer heterogeneity. *Econometrica 70*(6), 2295–2350.

Rothschild, C. and F. Scheuer (2013). Redistributive taxation in the Roy model. *Quarterly Journal of Economics 128*(2), 623–662.

Rothschild, C. and F. Scheuer (2014). A theory of income taxation under multidimensional skill heterogeneity. Mimeo. Stanford University.

Sachs, D., Tysvinski, and Werquin (2016). Nonlinear tax incidence and optimal taxation in general equilibrium. NBER WP 22646.

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

Scheuer, F. and I. Werning (2016a). Mirrlees meets diamond-mirrlees: Simplifying nonlinear income taxation. NBER WP 22076.

Scheuer, F. and I. Werning (2016b). The taxation of superstars. NBER WP 21323.

Shimer, R. (2012). Reassessing the ins and outs of unemployment. *Review of Economic Dynamics 15*(2), 127–148.

A Proofs

Proof of Proposition 1

Proof. Let $W^*(\theta) = W(P^*;\theta)$ denote the payoff to a new born θ talent at the optimal policy P^*. If $\theta \in [\underline{\theta}, \tilde{\theta}^*)$, then τ has no impact on $W^*(\theta)$. For $\theta \in [\tilde{\theta}^*, \overline{\theta}]$ and recalling that new borns are initially unemployed:

$$W^*(\theta) = U(b) + \beta\lambda \int_{\underline{q}^*(\theta)}^{\overline{q}^*(\theta)} V^*(q|\theta) f^*(q|\theta) dq + \beta(1-\lambda) W^*(\theta),$$

where $V^*(q|\theta)$ denotes the continuation payoff to an employed θ talent at a firm making profit offer q and f^* is the profit offer density both evaluated at the optimal policy. Differentiating with respect to τ gives:

$$\begin{aligned}\frac{\partial W^*}{\partial \tau}(\theta) = {}& \beta\lambda \int_{\underline{q}^*(\theta)}^{\overline{q}^*(\theta)} \frac{\partial V^*}{\partial \tau}(q|\theta) f^*(q|\theta) dq + \beta\lambda \int_{\underline{q}^*(\theta)}^{\overline{q}^*(\theta)} V^*(q|\theta) \frac{\partial f^*}{\partial \tau}(q|\theta) dq \\ & + V^*(q^*|\theta) f^*(q^*|\theta) \frac{\partial \overline{q}^*}{\partial \tau}(\theta) - V^*(q|\theta) f^*(q|\theta) \frac{\partial \underline{q}^*}{\partial \tau}(\theta) + \beta(1-\lambda) \frac{\partial W^*}{\partial \tau}(\theta). \end{aligned} \tag{A.1}$$

Using the definition of f^*,

$$\int_{\underline{q}^*(\theta)}^{\overline{q}^*(\theta)} V^*(q|\theta) \frac{\partial f^*}{\partial \tau}(q|\theta) dq = \frac{1}{2} \frac{\mathcal{E}^*_{\overline{q}}(\theta)}{1-\tau^*} \int_{\underline{q}^*(\theta)}^{\overline{q}^*(\theta)} V^*(q|\theta) f^*(q|\theta) dq, \tag{A.2}$$

where $\mathcal{E}^*_{\overline{q}}$ is the elasticity of $\overline{q}^*(\theta)$ with respect to $1-\tau$ evaluated at the optimum. Integrating by parts:

$$\begin{aligned}\int_{\underline{q}^*(\theta)}^{\overline{q}^*(\theta)} V^*(q|\theta) f^*(q|\theta) dq &= -\int_{\underline{q}^*(\theta)}^{\overline{q}^*(\theta)} \frac{\partial V^*(q|\theta) f^*(q|\theta)}{\partial q} q dq \\ &\quad + V^*(\overline{q}^*(\theta)) f(\overline{q}^*(\theta)) \overline{q}^*(\theta) - V^*(\underline{q}^*(\theta)) f(\underline{q}^*(\theta)) \underline{q}^*(\theta) \\ &= -\int_{\underline{q}^*(\theta)}^{\overline{q}^*(\theta)} \frac{\partial V^*}{\partial q}(q|\theta) f^*(q|\theta) q dq - \int_{\underline{q}^*(\theta)}^{\overline{q}^*(\theta)} V^*(q|\theta) \frac{\partial f^*}{\partial q}(q|\theta) q dq \\ &\quad + V^*(\overline{q}^*(\theta)) f(\overline{q}^*(\theta)) \overline{q}^*(\theta) - V^*(\underline{q}^*(\theta)) f(\underline{q}^*(\theta)) \underline{q}^*(\theta).\end{aligned}$$

Using $\frac{\partial f(q|\theta)}{\partial q} = -\frac{1}{2q} f(q|\theta)$, we then obtain:

$$\begin{aligned}\frac{1}{2} \int_{\underline{q}^*(\theta)}^{\overline{q}^*(\theta)} V^*(q|\theta) f^*(q|\theta) dq = {}& -\int_{\underline{q}^*(\theta)}^{\overline{q}^*(\theta)} \frac{\partial V^*}{\partial q}(q|\theta) f^*(q|\theta) q dq \\ & + V^*(\overline{q}^*(\theta)) f(\overline{q}^*(\theta)) \overline{q}^*(\theta) - V^*(\underline{q}^*(\theta)) f(\underline{q}^*(\theta)) \underline{q}^*(\theta).\end{aligned} \tag{A.3}$$

Combining (A.1) to (A.3) gives:

$$\frac{\partial W^*}{\partial \tau}(\theta) = \beta\lambda \int_{\underline{q}^*(\theta)}^{\overline{q}^*(\theta)} \frac{\partial V^*}{\partial \tau}(q|\theta) f^*(q|\theta) dq + \beta\lambda \frac{\mathcal{E}^*_{\overline{q}}(\theta)}{1-\tau^*} \int_{\underline{q}^*(\theta)}^{\overline{q}^*(\theta)} \frac{\partial V^*}{\partial q}(q|\theta) q f^*(q|\theta) dq$$
$$+ \beta(1-\lambda)\frac{\partial W^*}{\partial \tau}(\theta). \tag{A.4}$$

Next note that:

$$V^*(q|\theta) = U(\Phi^*(q,\theta)) + \beta\lambda \int_{\underline{q}^*(\theta)}^{q} V^*(q'|\theta) f^*(q'|\theta) dq' + \beta(1 - \lambda F^*(q|\theta) - \delta) V^*(q|\theta) + \beta\delta W^*(\theta).$$

Differentiating $V^*(q|\theta)$ with respect to q gives:

$$\frac{\partial V^*}{\partial q}(q|\theta) = -(1-\tau^*)U'(\Phi^*(q,\theta)) + \beta(1 - \lambda F^*(q|\theta) - \delta)\frac{\partial V^*}{\partial q}(q|\theta). \tag{A.5}$$

Differentiating $V^*(q|\theta)$ with respect to τ (and proceeding along similar lines to the calculation of $\frac{\partial W^*}{\partial \tau}(\theta)$) gives:

$$\frac{\partial V^*}{\partial \tau}(q|\theta) = -x^*(q,\theta)U'(\Phi^*(q,\theta)) + \beta\lambda \int_{\underline{q}^*(\theta)}^{q} \frac{\partial V^*}{\partial \tau}(q'|\theta) f^*(q'|\theta) dq'$$
$$+ \beta(1 - \lambda F^*[q|\theta] - \delta)\frac{\partial V^*}{\partial \tau}(q|\theta) - \beta\lambda \frac{\mathcal{E}^*_{\overline{q}}(\theta)}{1-\tau^*} \int_{\underline{q}^*(\theta)}^{q} \frac{\partial V}{\partial q}(q'|\theta) q' f^*(q'|\theta) dq' + \beta\delta \frac{\partial W^*}{\partial \tau}(\theta). \tag{A.6}$$

In period 0 a new born worker is unemployed with probability 1. In period 1, the worker remains unemployed with probability $1-\lambda$ and transitions to employment with a profit-offer of less than q with probability $\lambda F^*[q|\theta]$. In period 2, the worker is unemployed with probability $(1-\lambda)^2 + \lambda\delta$ and so forth. Let $Q^*_t[\cdot|\theta]$ denote the probability distribution over employment outcomes of a θ talent worker at age t at the optimum and let:

$$\tilde{Q}^*[\cdot|\theta] = (1-\beta)\sum_{t=0}^{\infty} \beta^t Q^*_t[\cdot|\theta].$$

If $\beta = \rho$ and survival risk is the worker's sole motive for discounting, then $\tilde{Q}^*[\cdot]$ gives the cross sectional distribution of θ talent workers across employment outcomes. More generally, if $\beta < \rho$, then it places less weight on outcomes attained (more frequently) later in life than the cross sectional distribution. Let $\tilde{E}^*[\cdot|\theta]$ denote the expectation under $\tilde{Q}^*[\cdot|\theta]$. Combining (A.4) to (A.6) and iterating gives:

$$(1-\beta)\frac{\partial W^*}{\partial \tau}(\theta) = -\tilde{E}^*[xU'(\Phi)|\theta] + \mathcal{E}^*_{\overline{q}}(\theta)\tilde{E}^*[U'(\Phi)q|\theta]. \tag{A.7}$$

And taking the expectation over θ gives the direct marginal impact of a tax raise on the welfare of new born workers:

$$(1-\beta)\int_{\underline{\theta}}^{\overline{\theta}}\frac{\partial W^*}{\partial\tau}(\theta)k(\theta)d\theta = -\tilde{E}^*[xU'(\Phi)] + \tilde{E}^*[\mathcal{E}_{\overline{q}}U'(\Phi)q].$$

We turn next to the budget constraint. The shadow value of an extra unit of government funds is $(1-\beta)\tilde{E}^*[U'(\Phi)|\theta]$. The direct (mechanical) increase in revenue from a small marginal tax rate increase is:

$$E^*[x] = \frac{\rho\lambda}{1-\rho(1-\lambda-\delta)}\int_{\tilde{\theta}^*}^{\overline{\theta}}\int_{\underline{q}^*(\theta)}^{\overline{q}^*(\theta)} x^*(q,\theta)g^*(q|\theta)k(\theta)dqd\theta,$$

where $E^*[\cdot]$ denotes the cross sectional average labor income at the optimum. The intensive margin labor supply response is:

$$-\frac{\tau^*}{1-\tau^*}E[x^*]\mathcal{E}^*_{E[x]} = \tau^*\frac{\partial}{\partial\tau}\left\{\frac{\rho\lambda}{1-\rho(1-\lambda-\delta)}\int_{\tilde{\theta}}^{\overline{\theta}}\int_{\underline{q}(\theta)}^{\overline{q}(\theta)} x(q,\theta)g(q|\theta)k(\theta)dqd\theta\right\}\Bigg|_{\tau=\tau^*},$$

where $\mathcal{E}^*_{E[x]}$ denotes the elasticity of $E^*[x]$ with respect to $1-\tau$ at the optimum. In addition, the tax rate rise also impacts the extensive margin and since the optimal $r^* = b^* - L^*$ need not equal 0, this has the following marginal impact on revenues:

$$r^*\frac{\tilde{\theta}^*}{1-\tau}\mathcal{E}^*_{\tilde{\theta}}\frac{\rho\lambda}{1-\rho(1-\lambda-\delta)}k(\tilde{\theta}^*),$$

where $\mathcal{E}^*_{\tilde{\theta}}$ is the elasticity of $\tilde{\theta}$ with respect to $1-\tau$ at the optimum. Finally, the marginal tax rate affects profits and, hence, profit tax revenues:

$$\begin{aligned}\frac{\partial\pi^*}{\partial\tau} =& \tau^*\frac{\partial\pi^*}{\partial\tau} + (1-\tau^*)\frac{\partial}{\partial\tau}\int_{\tilde{\theta}^*}^{\overline{\theta}}\int_{\underline{q}^*(\theta)}^{\overline{q}^*(\theta)} qg^*(q|\theta)k(\theta)dqd\theta \\ =& -\frac{\tau^*}{1-\tau^*}\mathcal{E}^*_{\pi}\pi^* - \int_{\tilde{\theta}^*}^{\overline{\theta}}\mathcal{E}^*_{\overline{q}}(\theta)\int_{\underline{q}^*(\theta)}^{\overline{q}^*(\theta)} qg^*(q|\theta)dqk(\theta)d\theta = -\frac{\tau^*}{1-\tau^*}\pi^*\mathcal{E}^*_{\pi} - E^*[\mathcal{E}_{\overline{q}}q],\end{aligned}$$

where the derivation of the second equality is similar to the derivation for (A.4) (but with q replacing $V^*(q,\theta)$) and $\mathcal{E}^*_{\pi}$ is the elasticity of aggregate profit π with respect to $1-\tau$ at the optimum. Combining terms gives the overall first order condition for τ:

$$\begin{aligned}0 =& -\tilde{E}^*[xU'(\Phi)] + \tilde{E}^*[\mathcal{E}_{\overline{q}}U'(\Phi^*)q] \\ &+ \tilde{E}^*[U'(\Phi)]\left\{E^*[x] - \tau^*\frac{E^*[x]\mathcal{E}^*_{E[x]}}{1-\tau^*} + r^*\frac{\tilde{\theta}^*\mathcal{E}^*_{\tilde{\theta}}}{1-\tau^*}\frac{\rho\lambda k(\tilde{\theta}^*)}{1-\rho(1-\lambda-\delta)} - E^*[\mathcal{E}_{\overline{q}}q] - \tau^*\frac{\pi^*\mathcal{E}^*_{\pi}}{1-\tau^*}\right\}.\end{aligned} \quad \text{(A.8)}$$

Equivalently, using the definitions of $\tilde{\lambda}$ and $\tilde{\delta}$:

$$
\begin{aligned}
0 = & - \tilde{\text{Cov}}^*(U'(\Phi), x) + \tilde{\text{Cov}}^*(U'(\Phi), \mathcal{E}_{\bar{q}} q) \\
& + \tilde{E}[U'(\Phi^*)]\{E[x] - \tilde{E}[x]\} - \tilde{E}[U'(\Phi^*)]\{E[\mathcal{E}^*_{\bar{q}} q] - \tilde{E}[\mathcal{E}^*_{\bar{q}} q]\} \\
& + \tilde{E}^*[U'(\Phi)] \left\{ -\tau^* \frac{E^*[x]\mathcal{E}^*_{E[x]}}{1-\tau^*} + r^* \frac{\tilde{\theta}^* \mathcal{E}^*_{\tilde{\theta}}}{1-\tau^*} \frac{\tilde{\lambda} k(\tilde{\theta}^*)}{\tilde{\lambda} + \tilde{\delta}} - \tau^* \frac{\pi^* \mathcal{E}^*_{\pi}}{1-\tau^*} \right\}.
\end{aligned}
\tag{A.9}
$$

□

Extension to include private savings Formula (A.9) was derived under the assumption that workers cannot save. The argument may be extended to accommodate saving. Assume that workers can purchase annuities at an exogenously given price. Further assume that income from annuities is untaxed. By accumulating annuities older workers can self insure against job loss. Now, worker value functions W^* and V^* depend on wealth as well as employment status. However, by an envelope theorem, savings adds no additional terms to the expression for $\frac{\partial W^*}{\partial \tau}(\theta)$, which following the argument leading to (A.7) in the proof of Proposition 1 becomes:

$$
(1-\beta)\frac{\partial W^*}{\partial \tau}(\theta) = -\tilde{E}^*[xU'(D+\Phi)|\theta] + \mathcal{E}^*_{\bar{q}}(\theta)\tilde{E}^*[U'(D+\Phi)q|\theta], \tag{A.10}
$$

where D is consumption in excess of after-tax labor income consumed in each employment-wealth state and $\tilde{Q}^*[\cdot|\theta]$ is modified to be the (discount-adjusted) distribution over employment status *and* wealth and $\tilde{E}^*[\cdot|\theta]$ is the corresponding expectation. The other times entering the optimal tax first order condition are unaltered by the inclusion of savings. The optimality condition (A.9) becomes:

$$
\begin{aligned}
0 = & - \tilde{\text{Cov}}^*(U'(D+\Phi), x) + \tilde{\text{Cov}}^*(U'(D+\Phi), \mathcal{E}_{\bar{q}} q) \\
& + \tilde{E}^*[U'(D+\Phi)]\{E^*[x] - \tilde{E}^*[x]\} - \tilde{E}^*[U'(D+\Phi)]\{E^*[\mathcal{E}_{\bar{q}} q] - \tilde{E}^*[\mathcal{E}_{\bar{q}} q]\} \\
& + \tilde{E}^*[U'(D+\Phi)] \left\{ -\tau^* \frac{E^*[x]\mathcal{E}^*_{E[x]}}{1-\tau^*} + r^* \frac{\tilde{\theta}^* \mathcal{E}^*_{\tilde{\theta}}}{1-\tau^*} \frac{\rho \lambda k(\tilde{\theta}^*)}{1-\rho(1-\lambda-\delta)} - \tau^* \frac{\pi^* \mathcal{E}^*_{\pi}}{1-\tau^*} \right\}.
\end{aligned}
\tag{A.11}
$$

This expression is essentially identical to (18). The only modification is that marginal utilities of consumption depend on D in addition to Φ.

Signing the redistributive term $\tilde{\text{Cov}}^*(U'(\Phi), \mathcal{E}_{\bar{q}} q)$ Whether the redistributive term $\tilde{\text{Cov}}^*(U'(\Phi), \mathcal{E}_{\bar{q}} q)$ is positive or negative depends upon whether the reductions in profit-per-worker (and increases in pre-tax worker incomes) induced by a tax change fall mainly on poorer agents at the bottom of job ladders from whom more profit is extracted relative to others in their talent market or richer, more talented who generate more surplus. We explore this more formally below at a possibly non-optimal policy P.

Recall that: $x(q,\theta)+q=\theta\frac{\partial h}{\partial y}^{-1}(\theta(1-\tau))$ and

$$\overline{q}(\theta)=-\frac{b-L}{1-\tau}+\frac{1}{1-\tau}\Gamma(\theta)=-\frac{b-L}{1-\tau}+x(q,\theta)+q-\frac{1}{1-\tau}h(\frac{x(q,\theta)+q}{\theta}). \qquad \text{(A.12)}$$

Assume that $b\geq L$, then $\overline{q}(\theta)<x(q,\theta)+q$ and:

$$\mathcal{E}_{\overline{q}}(\theta)=\frac{1-\tau}{\overline{q}(\theta)}\frac{\partial\overline{q}}{\partial(1-\tau)}(\theta)=-1+\frac{x+q}{\overline{q}(\theta)}=-1+\frac{\theta\frac{\partial h}{\partial y}^{-1}(\theta(1-\tau))}{\overline{q}(\theta)}>0.$$

After some algebra, one can show that:

$$\frac{\partial\mathcal{E}_{\overline{q}}}{\partial\theta}(\theta)=\frac{x+q}{\theta\overline{q}(\theta)}\left[\mathcal{E}(\theta)+1-\frac{x+q}{\overline{q}(\theta)}\right].$$

Further with $b\geq L$, $\frac{x+q}{\overline{q}(\theta)}\geq\frac{1}{1-\frac{h(\frac{x+q}{\theta})}{\frac{x+q}{\theta}h'(\frac{x+q}{\theta})}}$. Hence, if $h(y)=\frac{1}{1+\gamma}y^{1+\gamma}$, $\gamma>1$, then $\frac{x+q}{\overline{q}(\theta)}>\mathcal{E}+1$ and we conclude that $\frac{\partial\mathcal{E}_{\overline{q}}}{\partial\theta}(\theta)\leq 0$. Thus, under plausible assumptions on preferences, the highest and, hence, the distribution of profit-per-worker offers is less sensitive to changes in tax rates in higher talent markets.

We have by the law of total covariance:

$$\text{Cov}(U',\mathcal{E}_{\overline{q}}q)=E[\text{Cov}(U',\mathcal{E}_{\overline{q}}q|\theta)]+\text{Cov}(E[U'|\theta],E[\mathcal{E}_{\overline{q}}q|\theta]).$$

Now,

$$\begin{aligned}\text{Cov}(U',\mathcal{E}_{\overline{q}}q|\theta)=&\mathcal{E}_{\overline{q}}(\theta)\int_{\underline{q}(\theta)}^{\overline{q}(\theta)}U'(\Phi(q,\theta;P))qg(q|\theta;P)dq\\&-\mathcal{E}_{\overline{q}}(\theta)\int_{\underline{q}(\theta)}^{\overline{q}(\theta)}U'(\Phi(q,\theta;P))g(q|\theta;P)dq\int_{\underline{q}(\theta)}^{\overline{q}(\theta)}qg(q|\theta;P)dq.\end{aligned}$$

Since $\Phi(q,\theta;P)$ is decreasing in q and U' is decreasing in Φ, we conclude that $\text{Cov}(U',\mathcal{E}_{\overline{q}}q|\theta)$ is positive, i.e. *within* each talent market those lowest on the job ladder with the highest marginal utility benefit the most from a tax induced reduction in profit per worker offer. Since each $\text{Cov}(U',\mathcal{E}_{\overline{q}}q|\theta)>0$, it follows that $E[\text{Cov}(U',\mathcal{E}_{\overline{q}}q|\theta)]>0$.

We now verify that $E[U'|\theta]$ is decreasing in θ. To see this let $v(q|\theta)=\Phi(q,\theta;P)$ and change variables to write $E[U'|\theta]$ as:

$$-\int_{b}^{\overline{v}(\theta)}U'(v)g\left(\frac{q(v|\theta)}{q(b|\theta)}\right)\frac{dq}{dv}dv.$$

We have that: $-\frac{dq}{dv}=h'\left(\frac{x(v|\theta)+q(v|\theta)}{\theta}\right)\frac{1}{\theta}=1-\tau$. Also, $\frac{\partial q}{\partial\theta}(v,\theta)=\frac{x(v|\theta)+q(v|\theta)}{\theta}$. Let $K(v,\theta):=\frac{q(b|\theta)^{\frac{1}{2}}}{q(v|\theta)^{\frac{3}{2}}}$. It follows that: $\frac{\theta}{K(v,\theta)}\frac{\partial K}{\partial\theta}(v,\theta)=-\frac{3}{2}\frac{x(v|\theta)+q(v|\theta)}{q(v|\theta)}+\frac{1}{2}\frac{x(b|\theta)+q(b|\theta)}{q(b|\theta)}=-\frac{x(v|\theta)}{q(v|\theta)}+$

$\frac{x(b|\theta)}{q(b|\theta)}$ -1. But by the worker's first order condition: $1-\tau = \frac{1}{\theta}h'\left(\frac{x+q}{\theta}\right)$. Hence, for a given θ, $x+q$ equals a fixed number and $\frac{x}{q}$ increases as v rises and q falls. It follows that $\frac{\theta}{K(v,\theta)}\frac{\partial K}{\partial \theta}(v,\theta) < 0$. Hence, at higher θ values, the distribution over v first order stochastically dominates the distribution over lower values. But U' is falling in v and so we conclude that $E[U'|\theta]$ is falling in θ.

Next consider $E[\mathcal{E}_{\bar{q}}q|\theta]$. Using the formula for $\mathcal{E}_{\bar{q}}$ (and recalling that this variable does not depend on q), the formula for $\overline{q}(\theta)$ and ignoring unimportant constants:

$$\mathcal{E}_{\bar{q}}(\theta)E[q|\theta] \propto \mathcal{E}_{\bar{q}}(\theta)\int_{\underline{q}(\theta)}^{\overline{q}(\theta)}\left(\frac{\overline{q}(\theta)}{q}\right)^{\frac{1}{2}}dq \propto \left(-1+\frac{x(q,\theta)+q}{\overline{q}(\theta)}\right)\overline{q}(\theta)$$
$$= \frac{b-L}{1-\tau}+\frac{1}{1-\tau}h\left(h'^{-1}(\theta(1-\tau))\right).$$

This term is falling in θ and so $\text{Cov}(E[U'|\theta], E[\mathcal{E}_{\bar{q}}q|\theta]) < 0$. This captures the fact that *even though* profits are proportionately less sensitive to the income tax rate in higher talent markets, since average profits collected per worker are greater in higher talent markets, workers in these markets benefit more from the suppression of profits and on average these workers earn higher incomes and have higher consumption and utility. These calculations highlight the basic tension: the profit channel benefits workers at the bottom of the job ladder (who tend to be poorer, suggesting that the profit channel enhances the motive for redistributive taxation) *and* more highly talented workers (who tend to be richer suggesting that the profit channel dampens the motive for redistributive taxation).

We briefly indicate properties that are needed for the covariance $\tilde{\text{Cov}}(U'(\Phi), \mathcal{E}_{\bar{q}}q)$ to be negative (and, hence, a motive for lower taxes). First, assume that at the optimum $b = L$ and that $h(y) = \frac{1}{2}y^2$. Under these conditions $\mathcal{E}_{\bar{q}} = 1$, the profit-per-worker distribution of higher talent workers is no less sensitive that that of lower ones and $\tilde{\text{Cov}}(U'(\Phi^*), \mathcal{E}_{\bar{q}}q) = \tilde{\text{Cov}}(U'(\Phi), q)$. To assess the sign of this term, first change the order of integration and compute:

$$E[U'q] = \int_{\underline{q}}^{\overline{q}}\int_{\underline{\theta}(q)}^{\overline{\theta}(q)} U'(\Phi(q,\theta;P))qg(q|\theta;P)dq,$$

where $\underline{\theta}(q)$ and $\overline{\theta}(q)$ are defined implicitly by $q = \overline{q}(\underline{\theta}(q))$ and $q = \underline{q}(\overline{\theta}(q))$. Next consider:

$$E[U'|q] = \int_{\underline{\theta}(q)}^{\overline{\theta}(q)} U'(\Phi(q,\theta;P))\frac{\frac{\delta}{2\lambda}\left(\frac{\overline{q}(\theta)}{q}\right)^{\frac{1}{2}}\frac{1}{q}k(\theta)}{\int_{\underline{\theta}(q)}^{\overline{\theta}(q)}\frac{\delta}{2\lambda}\left(\frac{\overline{q}(\theta')}{q}\right)^{\frac{1}{2}}\frac{1}{q}k(\theta')d\theta'}d\theta$$
$$= \int_{\underline{\theta}(q)}^{\overline{\theta}(q)} U'(\Phi(q,\theta;P))\frac{\overline{q}(\theta)^{\frac{1}{2}}k(\theta)}{\int_{\underline{\theta}(q)}^{\overline{\theta}(q)}\overline{q}(\theta')^{\frac{1}{2}}k(\theta')d\theta'}d\theta.$$

Next change variables in the preceding integration. Define $v = \Phi(q,\theta;P)$, so that (given $h(y) = \frac{1}{2}y^2$) $dv = \theta(1-\tau)^2 d\theta$ and $\theta(v;q) = \sqrt{\frac{2}{1-\tau}\left\{\frac{v-L}{1-\tau} + q\right\}}$, then:

$$E[U'|q] = \int_b^{\frac{\lambda}{\delta}q(1-\tau)+b} \psi'(v) \frac{\left[\frac{v-b}{1-\tau} + q\right]^{\frac{1}{2}} \frac{1}{(1-\tau)^2} \frac{k(\theta(v;q))}{\left[\frac{2}{1-\tau}\left\{\frac{v-L}{1-\tau}+q\right\}\right]^{\frac{1}{2}}}}{\int_b^{\frac{\lambda}{\delta}q(1-\tau)+b} \left[\frac{v'-b}{1-\tau} + q\right]^{\frac{1}{2}} \frac{1}{(1-\tau)^2} \frac{k(\theta(v';q))}{\left[\frac{2}{1-\tau}\left\{\frac{v'-L}{1-\tau}+q\right\}\right]^{\frac{1}{2}}} dv'} dv.$$

This can be simplified to give:

$$E[U'|q] = \int_b^{\frac{\lambda}{\delta}q(1-\tau)+b} U'(v) \frac{\left[\frac{v-b}{1-\tau} + q\right]^{\frac{1}{2}} \frac{k(\theta(v;q))}{\left[\frac{v-L}{1-\tau}+q\right]^{\frac{1}{2}}}}{\int_b^{\frac{\lambda}{\delta}q(1-\tau)+b} \left[\frac{v'-b}{1-\tau} + q\right]^{\frac{1}{2}} \frac{k(\theta(v';q))}{\left[\frac{v'-L}{1-\tau}+q\right]^{\frac{1}{2}}} dv'} dv.$$

Consider the term:

$$\left[\frac{v-b}{1-\tau} + q\right]^{\frac{1}{2}} \frac{k(\theta(v;q))}{\left[\frac{v-L}{1-\tau} + q\right]^{\frac{1}{2}}}. \tag{A.13}$$

Let $\gamma(q) = \left[\left\{\frac{2}{1-\tau}\right\}\frac{v-b}{1-\tau} + q\right]^{\frac{1}{2}}$. Using the definitions of $\theta(v;q)$ and $\gamma(v;q)$, the derivative of (A.13) with respect to q is:

$$\frac{\gamma(v;q)k(\theta(v;q))}{\theta(v;q)}\left[\frac{1}{\gamma(v;q)}\frac{\partial\gamma}{\partial q}(v;q) - \frac{1}{\theta(v;q)}\frac{\partial\theta}{\partial q}(v;q) + \left(\frac{\theta k'(\theta(v;q))}{k(\theta(v;q))}\right)\frac{1}{\theta(v;q)}\frac{\partial\theta}{\partial q}(v;q)\right].$$

If $b = L$, this reduces to:

$$k(\theta(v;q))\left[\left(\frac{\theta k'(\theta(v;q))}{k(\theta(v;q))}\right)\frac{1}{\theta(v;q)}\frac{\partial\theta}{\partial q}(v;q)\right],$$

which is negative if $k' < 0$ (e.g. if k is a Pareto distribution). Thus, in this case, higher values of q place more mass on higher values of v. But these higher v are associated with lower values for U'. Hence, U' and q are negatively correlated. We conclude that in this case $-E[U']E[q] + E[U'q] < 0$.

B Nonlinear Tax Analysis

We first prove Lemma 1 from the main text. For convenience we restate the lemma.

Lemma 1 : *$\tilde{q}(\cdot, x; T)$ is increasing in θ between two endpoints $\underline{\theta}(x;T)$ and $\overline{\theta}(x,T)$.*

Proof. Recall the definition of $\tilde{q}$: $\tilde{q}(\theta, x; T) = \theta(h')^{-1}(\{1 - T'[x]\}\theta) - x$. Hence:

$$\frac{\partial}{\partial\theta}\tilde{q}(\theta, x; T) = (h')^{-1}(\{1 - T'[x]\}\theta) + \frac{\theta\{1 - T'[x]\}}{h''(\{1 - T'[x]\}\theta)} > 0.$$

□

In the remainder of this appendix, we characterize the optimal nonlinear tax. For convenience, we restate the policymaker's problem:

$$\sup_{T,b} \left(\frac{\delta}{\delta+\lambda} + \frac{\lambda}{\delta+\lambda}K[\tilde{\theta}(T)]\right)U(b) + \frac{\lambda}{\delta+\lambda}\int_{\tilde{\theta}(T)}^{\infty}\int_{\underline{q}(\theta;T)}^{\overline{q}(\theta;T)} U\left(\Phi(q, \theta; T)\right) g(q|\theta; T)dqk(\theta)d\theta. \tag{B.1}$$

subject to:

$$\mathcal{G} + \left(\frac{\delta}{\delta+\lambda} + \frac{\lambda}{\delta+\lambda}K[\tilde{\theta}(T)]\right)b = \frac{\lambda}{\lambda+\delta}\int_{\tilde{\theta}(T)}^{\infty}\int_{\underline{q}(\theta;T)}^{\overline{q}(\theta;T)} \{\tilde{q} + T[x(q, \theta; T)]\}\, g(q|\theta; T)dqk(\theta)d\theta,$$

where:

$$\Phi(q, \theta; T) = x(q, \theta; T) - T[x(q, \theta; T)] - h\left(\frac{x(q, \theta; T) + q}{\theta}\right).$$

It is convenient to let $r := b + T[0]$, $\tilde{T}[x] = T[x] - T[0]$ and $P = (r, T[0], \tilde{T})$ and to re-express this problem as:

$$\begin{aligned}\sup_{P} & \left(\frac{\delta}{\delta+\lambda} + \frac{\lambda}{\delta+\lambda}K[\tilde{\theta}(P)]\right)U(r - T[0]) \\ & + \frac{\lambda}{\delta+\lambda}\int_{\tilde{\theta}(P)}^{\infty}\int_{\underline{q}(\theta;P)}^{\overline{q}(\theta;P)} U\left(\Phi(q, \theta; P) - T[0]\right) g(q|\theta; P)dqk(\theta)d\theta.\end{aligned} \tag{B.2}$$

subject to:

$$\begin{aligned}\mathcal{G} + & \left(\frac{\delta}{\delta+\lambda} + \frac{\lambda}{\delta+\lambda}K[\tilde{\theta}(P)]\right)(r - T[0]) \\ & = \frac{\lambda}{\lambda+\delta}\int_{\tilde{\theta}(P)}^{\infty}\int_{\underline{q}(\theta;P)}^{\overline{q}(\theta;P)} \left\{\tilde{q} + \tilde{T}[x(q, \theta; P)] + T[0]\right\} g(q|\theta; P)dqk(\theta)d\theta,\end{aligned}$$

where:

$$\Phi(q, \theta; P) = x(q, \theta; P) - \tilde{T}[x(q, \theta; T)] - h\left(\frac{x(q, \theta; P) + q}{\theta}\right).$$

First order condition for $T[0]$**.** Let μ^* denote the optimal multiplier on the policymaker's budget constraint. Differentiating the policymaker's Lagrangian with respect to $T[0]$ (holding r and $\tilde{T}$ fixed at their optimal values) gives the first order condition:

$$E^*[U'] = \mu^*,$$

where $E^*[U']$ denotes the optimal average marginal utility.

First order condition for r. We next derive the first order condition for r. Recall that r is the premium received by unemployed workers over the tax paid by workers who work, but earn nothing, i.e. $r = b + T[0]$. As precursors to the first order condition, we give formulas for the variables $\frac{\partial \overline{q}}{\partial r}(\theta; P)$ and $\frac{\partial \tilde{\theta}}{\partial r}(\theta; P)$ which appear in the condition. First, recall that the variable $\overline{q}$ satisfies:

$$x(\overline{q}(\theta;P),\theta;P) - \tilde{T}[x(\overline{q}(\theta;P),\theta;P)] - h\left(\frac{x(\overline{q}(\theta;P),\theta;P) + \overline{q}(\theta;P)}{\theta}\right) = r.$$

Hence,

$$\frac{\partial \overline{q}}{\partial r}(\theta;P) = -\frac{1}{\frac{1}{\theta}h'\left(\frac{x(\overline{q}(\theta;P),\theta;P)+\overline{q}(\theta;P)}{\theta}\right)}. \tag{B.3}$$

In addition, the variable $\tilde{\theta}(P)$ satisfies:

$$x(0,\tilde{\theta}(P);P) - \tilde{T}[x(0,\tilde{\theta}(P);P)] - h\left(\frac{x(0,\tilde{\theta}(P);P)}{\tilde{\theta}(P)}\right) = r.$$

And so,

$$\frac{\partial \tilde{\theta}}{\partial r}(\theta;P) = \frac{1}{\frac{x(0,\tilde{\theta}(P);P)}{\tilde{\theta}(P)^2}h'\left(\frac{x(0,\tilde{\theta}(P);P)}{\tilde{\theta}(P)}\right)}. \tag{B.4}$$

Denoting optimal quantities with * superscripts, the first order condition for r is:

$$\begin{aligned}
&\Big(\frac{\delta}{\delta+\lambda} + \frac{\lambda}{\delta+\lambda}K[\tilde{\theta}^*]\Big)U'(b^*) \\
&\quad + \frac{\lambda}{\lambda+\delta}\int_{\tilde{\theta}^*}^{\infty}\frac{1}{\overline{q}^*(\theta)}\frac{\partial \overline{q}^*}{\partial r}(\theta)\int_{\underline{q}^*(\theta)}^{\overline{q}^*(\theta)} U'\left(\Phi^*(q,\theta)\right)\left(1 - T^{*\prime}[x^*(q,\theta)]\right)qg^*(q|\theta)dqk(\theta)d\theta \\
&\quad\quad + \frac{\lambda}{\lambda+\delta}E^*[U'(\Phi^*)]\Bigg\{ -k[\tilde{\theta}^*]\Big(T^*[x^*(0,\tilde{\theta}^*)] + b^*\Big)\frac{\partial \tilde{\theta}^*}{\partial r} \\
&\quad\quad\quad + \int_{\tilde{\theta}^*}^{\infty}\frac{1}{\overline{q}^*(\theta)}\frac{\partial \overline{q}^*}{\partial r}(\theta)\int_{\underline{q}^*(\theta)}^{\overline{q}^*(\theta)} T^{*\prime}[x^*(q,\theta)]\left\{\frac{x^*(q,\theta)}{x^*(q,\theta)+q}\frac{\mathscr{E}^*(q,\theta)}{\mathcal{E}^*(q,\theta)}\right\}qg^*(q|\theta)dqk(\theta)d\theta \\
&\quad\quad\quad\quad - \int_{\tilde{\theta}^*}^{\infty}\frac{1}{\overline{q}^*(\theta)}\frac{\partial \overline{q}^*}{\partial r}(\theta)\int_{\underline{q}^*(\theta)}^{\overline{q}^*(\theta)} qg^*(q|\theta)dqk(\theta)d\theta\Bigg\} = 0.
\end{aligned} \tag{B.5}$$

The first term in (B.5) gives the direct utility of benefit of an increase in r (and, hence, b) to an unemployed worker at the optimum. Although higher benefits raise $\tilde{\theta}^*$ (see (B.4)) and, hence, the fraction of workers who are unemployed, this has no direct effect on worker utility as the worker of talent $\tilde{\theta}^*$ is indifferent between employment and unemployment at the optimum. However, a higher value of r reduces $\overline{q}^*$ (see (B.3)) and shifts the distribution of profit-per-worker offers. In so doing it raises the incomes and utilities of those workers who work. This is captured by the second term in (B.5). The remaining terms capture the impact on (the shadow value of) the policymaker's budget of perturbations in r around the optimum. The first of these terms (on the third line) gives the adverse effect of raising $\tilde{\theta}^*$ and increasing unemployment. This effect combines the loss of tax revenues and the increased benefits paid. The next term on the fourth line gives the increased income tax revenues stemming from the higher incomes induced by the downward shift in the profit-per-worker distribution. The final term gives the loss in profit tax revenues from the same effect.

First order condition for T'. Finally, we consider the optimal determination of $\tilde{T}$. Specifically, we derive the first order condition stated in Theorem 1 for T' or, equivalently, for $\tilde{T}'$. To this end, let $\hat{T}[x] = T^*[x] + \kappa\Omega_\varepsilon^{x_0}(x)$ denote a tax perturbation around an optimal tax function T, where κ is a real number and $\Omega_\varepsilon^{x_0}(x) = \int_0^x \omega_\varepsilon^{x_0}(x')dx'$ is a function perturbing the tax amount at each income. Thus, the functions $\{\omega_\varepsilon^{x_0}\}$ perturb the marginal tax rate at each income. They will be constructed to allow us in a limiting sense to perturb the marginal tax rate "at" the single income level $x_0 > 0$. Each function $\omega_\varepsilon^{x_0}$ is assumed to be twice continuously differentiable and the family of perturbation functions $\omega_\varepsilon^{x_0}$ is assumed to satisfy for any sequence $\varepsilon_n \downarrow 0$, (i) $\lim_{n\to\infty} \int_0^\infty \omega_{\varepsilon_n}^{x_0} = 1$ and (ii) $\lim_{n\to\infty} \omega_{\varepsilon_n}^{x_0} \overset{\text{weak}}{\to} \delta_{x_0}$, where δ_{x_0} denotes the Dirac delta function and $\overset{\text{weak}}{\to}$ denotes weak convergence. Such a family satisfies for any continuous function l and sequence $\varepsilon_n \downarrow 0$, $\lim_{n\to\infty} \int_0^\infty \omega_{\varepsilon_n}^{x_0}(x)l(x)dx = l(x_0)$. In addition, note that for any $x < x_0$, $\lim_{n\to\infty} \Omega_{\varepsilon_n}^{x_0}(x) = 0$ and for $x > x_0$, $\lim_{n\to\infty} \Omega_{\varepsilon_n}^{x_0}(x) = 1$. As a concrete example consider the family $\{\omega_\varepsilon^{x_0}\}_\varepsilon$, where:

$$\omega_\varepsilon^{x_0}(x) = \begin{cases} 0 & x \leq x_0 - \varepsilon - \varepsilon^2 \\ \frac{1}{2\varepsilon} \exp\left(-\left\{\frac{x_0-\varepsilon-x}{x-x_0+\varepsilon+\varepsilon^2}\right\}^2\right) & x \in [x_0 - \varepsilon - \varepsilon^2, x_0 - \varepsilon] \\ \frac{1}{2\varepsilon} & x \in [x_0 - \varepsilon, x_0 + \varepsilon] \\ \frac{1}{2\varepsilon} \exp\left(-\left\{\frac{x-x_0-\varepsilon}{x_0+\varepsilon+\varepsilon^2-x}\right\}^2\right) & x \in [x_0 + \varepsilon, x_0 + \varepsilon + \varepsilon^2] \\ 0 & x \geq x_0 + \varepsilon + \varepsilon^2. \end{cases}$$

Suppose that l is a continuous function. Then, using the mean value theorem with $x' \in [x_0 - \varepsilon, x_0 + \varepsilon]$,

$$\begin{aligned} \int \omega_\varepsilon^{x_0}(x)l(x)dx &= \int_{x_0-\varepsilon-\varepsilon^2}^{x_0-\varepsilon} \omega_\varepsilon^{x_0}(x)l(x)dx + \int_{x_0+\varepsilon}^{x_0+\varepsilon+\varepsilon^2} \omega_\varepsilon^{x_0}(x)l(x)dx + \frac{1}{2\varepsilon}\int_{x_0-\varepsilon}^{x_0+\varepsilon} l(x)dx \\ &\leq \frac{1}{2\varepsilon}\int_{x_0-\varepsilon-\varepsilon^2}^{x_0-\varepsilon} |l(x)|dx + \frac{1}{2\varepsilon}\int_{x_0+\varepsilon}^{x_0+\varepsilon+\varepsilon^2} |l(x)|dx + \frac{1}{2\varepsilon}\int_{x_0-\varepsilon}^{x_0+\varepsilon} l(x)dx \\ &\leq \frac{\varepsilon}{2} \sup_{x\in[x_0-\varepsilon-\varepsilon^2, x_0-\varepsilon]} |l(x)| + \frac{\varepsilon}{2} \sup_{x\in[x_0+\varepsilon, x_0+\varepsilon+\varepsilon^2]} |l(x)| + l(x'). \end{aligned}$$

Given the continuity of l as $\varepsilon \to 0$ this bound converges to $l(x_0)$. Similarly, a lower bound can be computed that converges to $l(x_0)$. We conclude that for any continuous function l, $\lim_{\varepsilon\downarrow} \int \omega_\varepsilon^{x_0}(x) l(x) dx = l(x_0)$.

We first consider incomes x_0 that are above the minimum income paid to an active worker and, hence, above the income paid to the threshold type $\tilde{\theta}^*$, i.e. that are greater than $x^*(0, \tilde{\theta}^*)$.

Perturbing the objective For a small $\kappa > 0$ and $\Omega_\varepsilon^{x_0}$ in the family of perturbation functions described above, the perturbed objective for (32) is:

$$\left(\frac{\delta}{\delta+\lambda} + \frac{\lambda}{\delta+\lambda} K[\tilde{\theta}(T^* + \kappa\Omega_\varepsilon^{x_0})]\right) U(b) \\ + \frac{\lambda}{\delta+\lambda} \int_{\tilde{\theta}(T^*+\kappa\Omega_\varepsilon^{x_0})}^{\infty} \int_{\underline{q}(\theta;T^*+\kappa\Omega_\varepsilon^{x_0})}^{\overline{q}(\theta;T^*+\kappa\Omega_\varepsilon^{x_0})} U\left(\Phi(q,\theta;T^* + \kappa\Omega_\varepsilon^{x_0})\right) g(q|\theta; T^* + \kappa\Omega_\varepsilon^{x_0}) k(\theta) dq d\theta. \tag{B.6}$$

Our first goal is to differentiate this perturbed objective with respect to κ and evaluate at $\kappa = 0$. We will call this derivative the direct impact of the tax perturbation (on social welfare). We start by evaluating the impact of the tax perturbation on the utility of workers at each (q, θ), deferring for the moment the impact of the perturbation on the distribution of workers across (q, θ). First note that by the definition of Φ:

$$\Phi(q,\theta;T^* + \kappa\Omega_\varepsilon^{x_0}) = x(q,\theta;T^* + \kappa\Omega_\varepsilon^{x_0}) - T^*[x(q,\theta;T^* + \kappa\Omega_\varepsilon^{x_0})] - \kappa\Omega_\varepsilon^{x_0}[x(q,\theta;T^* + \kappa\Omega_\varepsilon^{x_0})] \\ - h\left(\frac{x(q,\theta;T^* + \kappa\Omega_\varepsilon^{x_0}) + q}{\theta}\right),$$

where $x(q,\theta;T^* + \kappa\Omega_\varepsilon^{x_0})$ is the worker's optimal income choice at $(q,\theta;T^* + \kappa\Omega_\varepsilon^{x_0})$. Hence:

$$\left.\frac{\partial\Phi}{\partial\kappa}(q,\theta;T^* + \kappa\Omega_\varepsilon^{x_0})\right|_{\kappa=0} = -\Omega_\varepsilon^{x_0}[x^*(q,\theta)]$$

And so:

$$\lim_{\varepsilon\downarrow 0} \frac{\lambda}{\delta+\lambda} \frac{\partial}{\partial\kappa} \int_{\tilde{\theta}^*}^{\infty} \int_{\underline{q}^*(\theta)}^{\overline{q}^*(\theta)} U\left(\Phi(q,\theta;T^* + \kappa\Omega_\varepsilon^{x_0})\right) g^*(q|\theta) k(\theta) dq d\theta \Bigg|_{\kappa=0} \\ = -\frac{\lambda}{\delta+\lambda} \int_{\underline{\theta}^*(x_0)}^{\infty} \int_{\underline{q}^*(\theta)}^{\hat{q}^*(\theta,x_0)} U'\left(\Phi^*(q,\theta)\right) g^*(q|\theta) k(\theta) dq d\theta, \tag{B.7}$$

where for $\theta \in [\underline{\theta}^*(x_0), \overline{\theta}^*(x_0)]$, $\hat{q}^*(\theta, x_0) = \min\{\tilde{q}^*(\theta, x_0), \overline{q}^*(\theta)\}$ and for $\theta > \overline{\theta}^*(x_0)$, $\hat{q}^*(\theta, x_0) = \overline{q}^*(\theta)$. Next consider the impact of the tax perturbation on the distribution of profit offers made and, hence, the utility of workers:

$$\frac{\lambda}{\delta+\lambda}\frac{\partial}{\partial\kappa}\int_{\tilde{\theta}^*}^{\infty}\int_{\underline{q}(\theta;T^*+\kappa\Omega_\varepsilon^{x_0})}^{\overline{q}(\theta;T^*+\kappa\Omega_\varepsilon^{x_0})} U\left(\Phi^*(q,\theta)\right) g(q|\theta;T^*+\kappa\Omega_\varepsilon^{x_0})k(\theta)dqd\theta\bigg|_{\kappa=0}$$

$$\begin{aligned} &= -\frac{\lambda}{\delta+\lambda}\int_{\underline{\theta}^*(x_0)}^{\infty} U\left(\Phi^*(\underline{q}^*(\theta),\theta)\right)\frac{\partial \underline{q}^*}{\partial\kappa}(\theta)g^*(\underline{q}^*(\theta)|\theta)k(\theta)d\theta \\ &\quad + \frac{\lambda}{\delta+\lambda}\int_{\underline{\theta}^*(x_0)}^{\infty} U\left(\Phi^*(\overline{q}^*(\theta),\theta)\right)\frac{\partial \overline{q}^*}{\partial\kappa}(\theta)g^*(\overline{q}^*(\theta)|\theta)k(\theta)d\theta \\ &\quad + \frac{\lambda}{\delta+\lambda}\int_{\underline{\theta}^*(x_0)}^{\infty}\int_{\underline{q}^*(\theta)}^{\overline{q}^*(\theta)} U\left(\Phi^*(q,\theta)\right)\frac{\partial g^*}{\partial\kappa}(q|\theta)k(\theta)dqd\theta. \end{aligned} \quad \text{(B.8)}$$

Define $\mathcal{E}_{\overline{q}}^*(\theta) := -\frac{1-T^{*\prime}[x_0]}{\overline{q}^*(\theta)}\frac{\partial\overline{q}^*}{\partial\kappa}(\theta)$. Note that $\overline{q}(\theta;T^*+\kappa\Omega_\varepsilon^{x_0})$ is defined implicitly by:

$$x(\overline{q}(\theta;T^*+\kappa\Omega_\varepsilon^{x_0}),\theta;T^*+\kappa\Omega_\varepsilon^{x_0}) - T^*[x(\overline{q}(\theta;T^*+\kappa\Omega_\varepsilon^{x_0}),\theta;T^*+\kappa\Omega_\varepsilon^{x_0})]$$
$$-\kappa\Omega_\varepsilon^{x_0}[x(\overline{q}(\theta;T^*+\kappa\Omega_\varepsilon^{x_0}),\theta;T^*+\kappa\Omega_\varepsilon^{x_0})] - h\left(\frac{x(\overline{q}(\theta;T+\kappa\Omega_\varepsilon^{x_0}),\theta;T^*+\kappa\Omega_\varepsilon^{x_0})+\overline{q}(\theta;T+\kappa\Omega_\varepsilon^{x_0})}{\theta}\right) = b.$$

Differentiating with respect to κ and setting κ equal to 0 gives:

$$\frac{\partial\overline{q}^*}{\partial\kappa}(\theta) := \frac{\partial\overline{q}}{\partial\kappa}(\theta;T^*+\kappa\Omega_\varepsilon^{x_0})\bigg|_{\kappa=0} = -\frac{\theta\Omega_\varepsilon^{x_0}[x^*(\overline{q}^*(\theta),\theta)]}{h'\left(\frac{x^*(\overline{q}^*(\theta),\theta)+\overline{q}^*(\theta)}{\theta}\right)}.$$

Thus, in the limit as $\varepsilon\downarrow 0$, $\mathcal{E}_{\overline{q}}^*(\theta)$ is zero for θ less than $\overline{\theta}^*(x_0)$.

The definitions of $\mathcal{E}_{\overline{q}}^*(\theta)$ and $\underline{q}^*$ imply that the term in the second line of (B.8) can be written as:

$$\begin{aligned} &-\frac{\lambda}{\delta+\lambda}\int_{\underline{\theta}^*(x_0)}^{\infty} U\left(\Phi^*(\underline{q}^*(\theta),\theta)\right)\frac{\partial\underline{q}^*}{\partial\kappa}(\theta)g^*(\underline{q}^*(\theta)|\theta)k(\theta)d\theta \\ &\quad = \frac{\lambda}{\delta+\lambda}\int_{\underline{\theta}^*(x_0)}^{\infty} U\left(\Phi^*(\underline{q}^*(\theta),\theta)\right)\frac{\mathcal{E}_{\overline{q}}^*(\theta)}{1-T^{*\prime}[x_0]}\underline{q}^*(\theta)g^*(\underline{q}^*(\theta)|\theta)k(\theta)d\theta. \end{aligned} \quad \text{(B.9)}$$

Similarly, the term in the third line of (B.8) can be written as:

$$\begin{aligned} &\frac{\lambda}{\delta+\lambda}\int_{\underline{\theta}^*(x_0)}^{\infty} U\left(\Phi^*(\overline{q}^*(\theta),\theta)\right)\frac{\partial\overline{q}^*}{\partial\kappa}(\theta)g^*(\overline{q}^*(\theta)|\theta)k(\theta)d\theta \\ &\quad = -\frac{\lambda}{\delta+\lambda}\int_{\underline{\theta}^*(x_0)}^{\infty} U\left(\Phi^*(\overline{q}^*(\theta),\theta)\right)\frac{\mathcal{E}_{\overline{q}}^*(\theta)}{1-T^{*\prime}[x_0]}\overline{q}^*(\theta)g^*(\overline{q}^*(\theta)|\theta)k(\theta)d\theta. \end{aligned} \quad \text{(B.10)}$$

Next recall that:

$$g(q|\theta;T^*+\kappa\Omega_\varepsilon^{x_0}) = \frac{\delta}{2\lambda}\left(\frac{\overline{q}(\theta;T^*+\kappa\Omega_\varepsilon^{x_0})}{q}\right)^{\frac{1}{2}}\frac{1}{q}.$$

And, hence,

$$\frac{\partial g^*}{\partial \kappa}(q|\theta) := \frac{\partial g}{\partial \kappa}(q|\theta; T^* + \kappa \Omega_\varepsilon^{x_0})\Big|_{\kappa=0} = -\frac{1}{2} g^*(q|\theta) \frac{\mathcal{E}_{\overline{q}}^*(\theta)}{1 - T^{*\prime}[x_0]}.$$

Thus, the final term in (B.8) satisfies:

$$\begin{aligned}
&\frac{\lambda}{\delta + \lambda} \int_{\tilde{\theta}^*}^{\infty} \int_{\underline{q}^*(\theta)}^{\overline{q}^*(\theta)} U\left(\Phi^*(q,\theta)\right) \frac{\partial g^*}{\partial \kappa}(q|\theta) k(\theta) dq d\theta \\
&\quad = -\frac{1}{2} \frac{\lambda}{\delta + \lambda} \int_{\tilde{\theta}^*}^{\infty} \frac{\mathcal{E}_{\overline{q}}^*(\theta)}{1 - T^{*\prime}[x_0]} \left\{ \int_{\underline{q}^*(\theta)}^{\overline{q}^*(\theta)} U\left(\Phi^*(q,\theta)\right) g^*(q|\theta) dq \right\} k(\theta) d\theta. \qquad \text{(B.11)}
\end{aligned}$$

Integrating the inner integral in (B.11) by parts:

$$\begin{aligned}
&\int_{\underline{q}^*(\theta)}^{\overline{q}^*(\theta)} U\left(\Phi^*(q,\theta)\right) g^*(q|\theta) dq \\
&\quad = U\left(\Phi^*(q,\theta)\right) g^*(q|\theta) q \Big]_{\underline{q}^*(\theta)}^{\overline{q}^*(\theta)} - \int_{\underline{q}^*(\theta)}^{\overline{q}^*(\theta)} \frac{\partial}{\partial q} \left\{ U\left(\Phi^*(q,\theta)\right) g^*(q|\theta) \right\} q dq \\
&\quad = U\left(\Phi^*(q,\theta)\right) g^*(q|\theta) q \Big]_{\underline{q}^*(\theta)}^{\overline{q}^*(\theta)} + \int_{\underline{q}^*(\theta)}^{\overline{q}^*(\theta)} U'\left(\Phi^*(q,\theta)\right) (1 - T^{*\prime}[x^*(q,\theta)]) q g^*(q|\theta) dq \\
&\qquad\qquad - \int_{\underline{q}^*(\theta)}^{\overline{q}^*(\theta)} U\left(\Phi^*(q,\theta)\right) \frac{\partial g^*}{\partial q}(q|\theta) q dq \\
&\quad = U\left(\Phi^*(q,\theta)\right) g^*(q|\theta) q \Big]_{\underline{q}^*(\theta)}^{\overline{q}^*(\theta)} + \int_{\underline{q}^*(\theta)}^{\overline{q}^*(\theta)} U'\left(\Phi^*(q,\theta)\right) (1 - T^{*\prime}[x^*(q,\theta)]) q g^*(q|\theta) dq \\
&\qquad\qquad + \frac{3}{2} \int_{\underline{q}^*(\theta)}^{\overline{q}^*(\theta)} U\left(\Phi^*(q,\theta)\right) g^*(q|\theta) dq.
\end{aligned}$$

Consequently,

$$\begin{aligned}
\frac{1}{2} \int_{\underline{q}^*(\theta)}^{\overline{q}^*(\theta)} U\left(\Phi^*(q,\theta)\right) g^*(q|\theta) dq = &-U\left(\Phi^*(q,\theta)\right) g^*(q|\theta) q \Big]_{\underline{q}^*(\theta)}^{\overline{q}^*(\theta)} \\
&- \int_{\underline{q}^*(\theta)}^{\overline{q}^*(\theta)} U'\left(\Phi^*(q,\theta)\right) (1 - T^{*\prime}[x^*(q,\theta)]) q g^*(q|\theta) dq. \qquad \text{(B.12)}
\end{aligned}$$

Combining (B.8) to (B.12) gives:

$$\frac{\lambda}{\delta+\lambda}\frac{\partial}{\partial\kappa}\int_{\tilde{\theta}^*}^{\infty}\int_{\underline{q}(\theta;T^*+\kappa\Omega_\varepsilon^{x_0})}^{\overline{q}(\theta;T^*+\kappa\Omega_\varepsilon^{x_0})} U\left(\Phi^*(q,\theta)\right) g(q|\theta;T^*+\kappa\Omega_\varepsilon^{x_0})k(\theta)dqd\theta\Big|_{\kappa=0}$$
$$=\frac{\lambda}{\delta+\lambda}\int_{\overline{\theta}^*(x_0)}^{\infty}\frac{\mathcal{E}^*_{\overline{q}}(\theta)}{1-T^{*\prime}[x_0]}\int_{\underline{q}^*(\theta)}^{\overline{q}^*(\theta)} U'\left(\Phi^*(q,\theta)\right)(1-T^{*\prime}[x^*(q,\theta)])qg^*(q|\theta)dqk(\theta)d\theta. \tag{B.13}$$

Putting everything together:

$$\lim_{\varepsilon\downarrow 0}\frac{\lambda}{\delta+\lambda}\frac{\partial}{\partial\kappa}\int_{\tilde{\theta}^*}^{\infty}\int_{\underline{q}(\theta;T^*+\kappa\Omega_\varepsilon^{x_0})}^{\overline{q}(\theta;T^*+\kappa\Omega_\varepsilon^{x_0})} U\left(\Phi(q,\theta;T^*+\kappa\Omega_\varepsilon^{x_0})\right) g(q|\theta;T^*+\kappa\Omega_\varepsilon^{x_0})k(\theta)dqd\theta\Big|_{\kappa=0}$$
$$=\frac{\lambda}{\delta+\lambda}\int_{\underline{\theta}^*(x_0)}^{\infty}\int_{\underline{q}^*(\theta)}^{\hat{q}^*(\theta,x_0)} U'\left(\Phi^*(q,\theta)\right) g^*(q|\theta)k(\theta)dqd\theta$$
$$+\frac{\lambda}{\delta+\lambda}\int_{\overline{\theta}^*(x_0)}^{\infty}\frac{\mathcal{E}^*_{\overline{q}}(\theta)}{1-T^{*\prime}[x_0]}\int_{\underline{q}^*(\theta)}^{\overline{q}^*(\theta)} U'\left(\Phi^*(q,\theta)\right)(1-T^{*\prime}[x^*(q,\theta)])qg^*(q|\theta)dqk(\theta)d\theta \tag{B.14}$$

Perturbing Tax Revenues We next consider the impact of the perturbation on tax revenues. Perturbed tax revenues are given by:

$$\frac{\lambda}{\delta+\lambda}\int_{\tilde{\theta}^*}^{\infty}\int_{\underline{q}(\theta;T^*+\kappa\Omega_\varepsilon^{x_0})}^{\overline{q}(\theta;T^*+\kappa\Omega_\varepsilon^{x_0})}\{T^*[x(q,\theta;T^*+\kappa\Omega_\varepsilon^{x_0})]+\kappa\Omega_\varepsilon^{x_0}[x(q,\theta;T^*+\kappa\Omega_\varepsilon^{x_0})]\}g(q|\theta;T^*+\kappa\Omega_\varepsilon^{x_0})dqk(\theta)d\theta$$
$$+\frac{\lambda}{\delta+\lambda}\int_{\tilde{\theta}(T^*+\kappa\Omega_\varepsilon^{x_0})}^{\infty}\int_{\underline{q}(\theta;T^*+\kappa\Omega_\varepsilon^{x_0})}^{\overline{q}(\theta;T^*+\kappa\Omega_\varepsilon^{x_0})} qg(q|\theta;T^*+\kappa\Omega_\varepsilon^{x_0})dqk(\theta)d\theta.$$

From (28):

$$-\frac{\partial x(q,\theta;T^*+\kappa\Omega_\varepsilon^{x_0})}{\partial\kappa}\Big|_{\kappa=0}=\frac{x^*(q,\theta)\omega_\varepsilon^{x_0}(x^*(q,\theta))}{1-T^{*\prime}[x^*(q,\theta)]}\mathscr{E}^*(q,\theta). \tag{B.15}$$

And so:

$$\lim_{\varepsilon\downarrow 0}-\frac{\partial x(q,\theta;T^*+\kappa\Omega_\varepsilon^{x_0})}{\partial\kappa}\Big|_{\kappa=0}=\begin{cases}\frac{x_0}{1-T^{*\prime}[x_0]}\mathscr{E}^*(q,\theta) & \text{if } x^*(q,\theta)=x_0\\ 0 & \text{otherwise}\end{cases}. \tag{B.16}$$

Define:

$$\overline{\mathscr{E}}(x_0;T)=\int_{\underline{\theta}(x_0;T)}^{\overline{\theta}(x_0;T)}\mathscr{E}(\tilde{q}(x_0,\theta;T),\theta;T)\frac{g(\tilde{q}(x_0,\theta;T)|\theta;T)k(\theta)}{\int_{\underline{\theta}(x_0;T)}^{\overline{\theta}(x_0;T)} g(\tilde{q}(x_0,\theta;T)|\theta;T)k(\theta)}d\theta$$

to be the average tax elasticity amongst workers earning x_0 at tax function T and let $J[\cdot;T]$ and $j(\cdot;T)$ denote the induced labor income distribution and density. Note that:

$$1 - J[x_0; T] = \lim_{\varepsilon \downarrow 0} \int_{\underline{\theta}(T)}^{\infty} \int_{\underline{q}(\theta;T)}^{\overline{q}(\theta;T)} \Omega_{\varepsilon}^{x_0}(x(q,\theta;T))g(q|\theta;T)k(\theta)dqd\theta. \tag{B.17}$$

Proceeding analogously to the derivation of (B.14),

$$\begin{aligned}
&\lim_{\varepsilon \downarrow 0} \frac{\partial}{\partial \kappa} \int_{\tilde{\theta}(T^*)}^{\infty} \int_{\underline{q}(\theta;T^*+\kappa\Omega_{\varepsilon}^{x_0})}^{\overline{q}(\theta;T^*+\kappa\Omega_{\varepsilon}^{x_0})} \{T^*[x(q,\theta;T^*+\kappa\Omega_{\varepsilon}^{x_0})] + \kappa\Omega_{\varepsilon}^{x_0}(x(q,\theta;T^*+\kappa\Omega_{\varepsilon}^{x_0}))\}g(q|\theta;T^*+\kappa\Omega_{\varepsilon}^{x_0})k(\theta)dqd\theta\Big|_{\kappa=0} \\
&\quad = (1 - J[x_0; T^*]) - \frac{x_0 T^{*\prime}[x_0]}{1 - T^{*\prime}[x_0]} \overline{\mathscr{E}}(x_0;^* T)j(x_0;T^*) \\
&\quad + \int_{\overline{\theta}(x_0;T^*)}^{\infty} \frac{\mathcal{E}_{\overline{q}}(\theta;T^*)}{1 - T^{*\prime}[x_0]} \int_{\underline{q}(\theta;T^*)}^{\overline{q}(\theta;T^*)} T^{*\prime}[x(q,\theta;T^*)] \left\{ \frac{x(q,\theta;T^*)}{x(q,\theta;T^*) + q} \frac{\mathscr{E}(q,\theta;T^*)}{\mathcal{E}(q,\theta;T^*)} \right\} qg(q|\theta;T^*)dqk(\theta)d\theta.
\end{aligned} \tag{B.18}$$

Finally, turning to profit taxes:

$$\begin{aligned}
&\lim_{\varepsilon \downarrow 0} \frac{\partial}{\partial \kappa} \int_{\tilde{\theta}(T^*)}^{\infty} \int_{\underline{q}(\theta;T^*+\kappa\Omega_{\varepsilon}^{x_0})}^{\overline{q}(\theta;T^*+\kappa\Omega_{\varepsilon}^{x_0})} qg(q|\theta;T^*+\kappa\Omega_{\varepsilon}^{x_0})k(\theta)dqd\theta\Big|_{\kappa=0} \\
&\qquad = -\frac{\lambda}{\delta + \lambda} \int_{\overline{\theta}(x_0;T^*)}^{\infty} \frac{\mathcal{E}_{\overline{q}}(\theta;T^*)}{1 - T^{*\prime}[x_0]} \int_{\underline{q}(\theta;T^*)}^{\overline{q}(\theta;T^*)} qg(q|\theta;T^*)dqk(\theta)d\theta.
\end{aligned} \tag{B.19}$$

The optimal resource multiplier $\mu^* = E^*[U'(\Phi)]$ and so combining everything, the first order condition for the marginal tax rate $T'[x_0]$ is:

$$\begin{aligned}
&- \int_{\underline{\theta}(x_0;T^*)}^{\infty} \int_{\underline{q}(\theta;T^*)}^{\hat{q}(\theta,x_0;T^*)} U'(\Phi(q,\theta;T^*))\, g(q|\theta;T^*)k(\theta)dqd\theta \\
&+ \int_{\overline{\theta}(x_0;T^*)}^{\infty} \frac{\mathcal{E}_{\overline{q}}(\theta;T^*)}{1 - T^{*\prime}[x_0]} \int_{\underline{q}(\theta;T^*)}^{\overline{q}(\theta;T^*)} U'(\Phi(q,\theta;T^*))\,(1 - T^{*\prime}[x(q,\theta;T^*)])qg(q|\theta;T^*)dqk(\theta)d\theta \\
&+ E^*[U'(\Phi)] \Bigg\{ (1 - J[x_0;T^*]) - \frac{x_0 T^{*\prime}[x_0]}{1 - T^{*\prime}[x_0]} \overline{\mathscr{E}}(x_0;T^*)j(x_0;T^*) \\
&\quad + \int_{\overline{\theta}(x_0;T^*)}^{\infty} \frac{\mathcal{E}_{\overline{q}}(\theta;T^*)}{1 - T^{*\prime}[x_0]} \int_{\underline{q}(\theta;T^*)}^{\overline{q}(\theta;T^*)} T^{*\prime}[x(q,\theta;T^*)] \left\{ \frac{x(q,\theta;T^*)}{x(q,\theta;T^*) + q} \frac{\mathscr{E}(q,\theta;T^*)}{\mathcal{E}(q,\theta;T^*)} \right\} qg(q|\theta;T^*)dqk(\theta)d\theta \\
&\quad - \int_{\overline{\theta}(x_0;T^*)}^{\infty} \frac{\mathcal{E}_{\overline{q}}(\theta;T^*)}{1 - T^{*\prime}[x_0]} \int_{\underline{q}(\theta;T^*)}^{\overline{q}(\theta;T^*)} qg(q|\theta;T^*)dqk(\theta)d\theta \Bigg\} = 0.
\end{aligned} \tag{B.20}$$

We next consider incomes x_0 that are below the minimum income paid (to an active worker) (and, hence, above the income paid to the threshold type $\tilde{\theta}(T^*)$), i.e. that are below $x(0, \tilde{\theta}(T^*); T^*)$.

First note that:

$$\left.\frac{\partial\tilde{\theta}}{\partial\kappa}\right|_{\kappa=0}=\frac{\Omega_{\varepsilon}^{x_0}(x(0,\tilde{\theta}(T);T))}{h'\left(\frac{x(0,\tilde{\theta}(T);T)}{\tilde{\theta}(T)}\right)\frac{x(0,\tilde{\theta}(T);T)}{\tilde{\theta}(T)^2}}.$$

Thus, perturbations to the tax function at incomes below $x(0,\tilde{\theta}(T);T)$ perturb $\tilde{\theta}(T)$ (by raising the level of taxes at $x(0,\tilde{\theta}(T);T)$). This does not impact the social objective since $\tilde{\theta}(T)$ workers are indifferent between working and not working. However, it does reduce revenues by:

$$-\frac{\lambda}{\delta+\lambda}k[\tilde{\theta}(T)]\Big(T[x(0,\tilde{\theta}(T);T)]+b\Big)\left.\frac{\partial\tilde{\theta}}{\partial\kappa}\right|_{\kappa=0}.$$

Other impacts are computed along the same lines as for incomes $x_0>x(0,\tilde{\theta}(T);T)$. The overall effect evaluated at the optimum is:

$$\begin{aligned}
&-\int_{\tilde{\theta}(T^*)}^{\infty}\int_{\underline{q}(\theta;T^*)}^{\overline{q}(\theta;T^*)}U'\left(\Phi(q,\theta;T^*)\right)g(q|\theta;T^*)k(\theta)dqd\theta\\
&\quad+\int_{\tilde{\theta}(T^*)}^{\infty}\frac{\mathcal{E}_{\overline{q}}^{*}(\theta;T^*)}{1-T^{*\prime}[x_0]}\int_{\underline{q}(\theta;T^*)}^{\overline{q}(\theta;T^*)}U'\left(\Phi(q,\theta;T^*)\right)(1-T^{*\prime}[x(q,\theta;T^*)])qg(q|\theta;T^*)dqk(\theta)d\theta\\
&\quad\quad+E^*[U'(\Phi)]\Bigg\{(1-J[x(0,\tilde{\theta}(T^*);T^*);T^*])-k[\tilde{\theta}(T^*)]\frac{T^*[x(0,\tilde{\theta}(T^*);T^*)]+b}{h'\left(\frac{x(0,\tilde{\theta}(T^*);T^*)}{\tilde{\theta}(T^*)}\right)\frac{x(0,\tilde{\theta}(T^*);T^*)}{\tilde{\theta}(T^*)^2}}\\
&\quad\quad\quad+\int_{\tilde{\theta}(T^*)}^{\infty}\frac{\mathcal{E}_{\overline{q}}(\theta;T^*)}{1-T^{*\prime}[x_0]}\int_{\underline{q}(\theta;T^*)}^{\overline{q}(\theta;T^*)}T'[x(q,\theta;T^*)]\left\{\frac{x(q,\theta;T^*)}{x(q,\theta;T^*)+q}\frac{\mathscr{E}(q,\theta;T^*)}{\mathcal{E}(q,\theta;T^*)}\right\}qg(q|\theta;T^*)dqk(\theta)d\theta\\
&\quad\quad\quad\quad-\int_{\tilde{\theta}(T^*)}^{\infty}\frac{\mathcal{E}_{\overline{q}}(\theta;T^*)}{1-T^{*\prime}[x_0]}\int_{\underline{q}(\theta;T^*)}^{\overline{q}(\theta;T^*)}qg(q|\theta;T^*)dqk(\theta)d\theta\Bigg\}=0.
\end{aligned}\tag{B.21}$$

Comparing optimal nonlinear to optimal affine tax formulas The optimal nonlinear equation (B.20) is readily reorganized to more closely resemble the affine optimal equation (19):

$$
\begin{aligned}
&-\left\{\int_{\underline{\theta}(x_0;T^*)}^{\infty}\int_{\underline{q}(\theta;T^*)}^{\hat{q}(\theta,x_0;T^*)} U'\left(\Phi(q,\theta;T^*)\right) g(q|\theta;T^*)k(\theta)dqd\theta - E^*[U'(\Phi)](1-J[x_0;T^*])\right\}\\
&\quad + E^*[U'(\Phi)]\Bigg\{-\frac{x_0 T^{*\prime}[x_0]}{1-T^{*\prime}[x_0]}\overline{\mathscr{E}}(x_0;T^*)j(x_0;T^*)\\
&\qquad + \int_{\overline{\theta}(x_0;T^*)}^{\infty}\frac{\mathcal{E}_{\overline{q}}(\theta;T^*)}{1-T^{*\prime}[x_0]}\int_{\underline{q}(\theta;T^*)}^{\overline{q}(\theta;T^*)} T^{*\prime}[x(q,\theta;T^*)]\mathscr{R}(q,\theta;T^*)qg(q|\theta;T^*)dqk(\theta)d\theta\Bigg\}\\
&\quad + \Bigg\{\int_{\overline{\theta}(x_0;T^*)}^{\infty}\frac{\mathcal{E}_{\overline{q}}(\theta;T^*)}{1-T^{*\prime}[x_0]}\int_{\underline{q}(\theta;T^*)}^{\overline{q}(\theta;T^*)} U'\left(\Phi(q,\theta;T^*)\right)(1-T^{*\prime}[x(q,\theta;T^*)])qg(q|\theta;T^*)dqk(\theta)d\theta\\
&\qquad - E^*[U'(\Phi)]\int_{\overline{\theta}(x_0;T^*)}^{\infty}\frac{\mathcal{E}_{\overline{q}}(\theta;T^*)}{1-T^{*\prime}[x_0]}\int_{\underline{q}(\theta;T^*)}^{\overline{q}(\theta;T^*)}(1-T^{*\prime}[x(q,\theta)])qg(q|\theta;T^*)dqk(\theta)d\theta\Bigg\}\\
&\quad - E^*[U'(\Phi)]\left\{\int_{\overline{\theta}(x_0;T^*)}^{\infty}\frac{\mathcal{E}_{\overline{q}}(\theta;T^*)}{1-T^{*\prime}[x_0]}\int_{\underline{q}(\theta;T^*)}^{\overline{q}(\theta;T^*)} T^{*\prime}[x(q,\theta;T^*)]qg(q|\theta;T^*)dqk(\theta)d\theta\right\} = 0,
\end{aligned}
\tag{B.22}
$$

where $\mathscr{R}(q,\theta;T^*) := \frac{x(q,\theta;T^*)}{x(q,\theta;T^*)+q}\frac{\mathscr{E}(q,\theta;T^*)}{\mathcal{E}(q,\theta;T^*)}$. The first line contains the analogue of the covariance between marginal utility and labor earnings from the affine case. The second and third lines contain the analogue of the second term form the affine case: it gives the impact on labor earnings and, hence, revenues from the tax perturbation. This term includes the usual intensive margin income response and the profit channel income response. The fourth and fifth lines give the analogue of the covariance between marginal utility and the profit adjustment. As in the affine case, it captures redistributive consequences of the profit channel. The final line adjusts for the fact that workers only receive the after-tax profit adjustment.

Our affine tax discussion was cast in the language of inter- and intra-talent market redistribution. In the affine setting, inter-talent market redistribution is associated with the application of a common tax rate across incomes. Mechanically, adjustment of this tax rate upwards has a larger adverse effect on the after-tax income of high talent workers who earn more. To attract workers bottom of the job ladder firms must correspondingly reduce their profit-per-worker offers more in high relative to low talent markets. The profit squeeze is stronger in such markets and, hence, implies redistribution from low to high talents. In the nonlinear setting, the tax rate is optimally adjusted at each dollar income level separately. An adjustment at income x_0 has no effect on the after-tax income of those earning less than x_0, but the same mechanical effect on the after-tax income of all those earning more than x_0. Thus, there is discrete inter-talent market redistribution from those in talent markets with minimal income below x_0 to those in talent markets with minimal income above x_0. Intra-talent market redistribution is present in the affine case because the profit squeeze dissipates at high income levels within a talent market. This redistribution is also embedded into the nonlinear tax equation (e.g. in the second term of (33)). In the nonlinear case, perturbation of the marginal tax at x_0 induces intra-talent market redistribution in markets with minimal income exceeding x_0: amongst this population, there is more profit diversion towards the incomes of those at the bottom of job ladders. However, such intra-talent market redistribution is quantitatively small in the nonlinear setting because (i) at low incomes, optimal tax rates are high implying that there is little diversion of profit to after-tax income and (ii) at high incomes the value of such redistribution is small. In the nonlinear setting, it is more natural to cast the discussion in terms of redistribution across income earners rather than redistribution within and across talent markets.

C Data

Our main data source is the Current Population Survey (CPS) administered by the US Census Bureau and the US Bureau of Labor Statistics. We focus on the March release of the 2016 survey which provides information for the calendar year 2015 and use March supplement sample weights to produce our estimates. The total number of observations in the raw sample is 185,487. We focus on population of working age individuals and so we drop people who are younger than 25 or older than 65 years of age. This reduces the sample size to 97,168.

Our main variable of interest is labor income. We measure this as the sum of wage and salary income earned during the last calendar year. We compute total hours worked from data on average hours worked per week and total number of weeks worked. Following Heathcote, Perri and Violante (2010), we drop those who work a very small number of (less than 100) hours and those who earned very small annual amounts (less than $250). We also drop those giving inconsistent answers to questions about labor income and hours worked (i.e. those who claim they received labor income, but did not work or vice versa). We calculate the hourly wage rate for the remaining working agents in our sample by dividing total labor income earned by total hours worked, and drop individuals whose hourly wage rate is below half of the federal minimum wage.

In addition to working agents, our sample contains agents who have not worked at all during the last calendar year. These agents are asked their reasons for not working. In our model agents who do not work have very low talent. To align non-workers in the data with those in the model, we retain those giving disability, sickness or inability to find work despite searching as reasons for not working, while dropping those giving taking care of home or family, going to school, or retirement as reasons. We are left with 68,749 working agents and 8,120 people who did not work during 2015. We identify the latter group of people with those below the active talent threshold $\tilde{\theta}$ implied by current US policy (and our assumed preference and labor market transition prameters)

D Numerical Algorithm for Affine Tax Analysis

Derivations used in the Algorithm For a given policy $P \equiv (b, L, \tau)$, workers in active talent market $\theta \geq \tilde{\theta}$ solve:

$$\max_x L + (1-\tau)x - \frac{(\frac{x+q}{\theta})^{1+\gamma}}{1+\gamma}.$$

This implies worker income and surplus policy functions:

$$x(q, \theta) = [(1-\tau)\theta]^{1/\gamma}\theta - q$$

and

$$y(\theta) \equiv x(q, \theta) + q = [(1-\tau)\theta]^{1/\gamma}\theta.$$

A talent market θ is active if a worker receiving a profit offer $q = 0$ can obtain a utility of at least b. In active talent markets, the maximal profit offer $\overline{q}(\theta)$ that can be made (and accepted with positive probability) is:

$$L + (1-\tau)\{y(\theta) - \overline{q}(\theta)\} - \frac{1}{1+\gamma}\left(\frac{y(\theta)}{\tilde{\theta}}\right)^{1+\gamma} = b.$$

The activity threshold $\tilde{\theta}$ is such that $\overline{q}(\tilde{\theta}) = 0$. Hence, extending $\overline{q}$ from $[\tilde{\theta}, \overline{\theta}]$ to $\underline{\theta}, \overline{\theta}]$, we have:

$$\bar{q}(\theta) = -\frac{b-L}{1-\tau} + v(\theta) < 0, \quad \text{for } \theta < \tilde{\theta}, \tag{D.1}$$
$$\bar{q}(\tilde{\theta}) = -\frac{b-L}{1-\tau} + v(\tilde{\theta}) = 0,$$
$$\bar{q}(\theta) = -\frac{b-L}{1-\tau} + v(\theta) > 0, \quad \text{for } \theta > \tilde{\theta},$$

where $v(\theta) = y(\theta) - \frac{(\frac{y(\theta)}{\theta})^{1+\gamma}}{(1+\gamma)(1-\tau)}$. Substituting v into and rearranging the policymaker's budget constraint gives L as a function of b and τ:

$$L = \frac{1}{A}\Big[-Bb + \frac{\tilde{\lambda}}{\tilde{\delta}+\tilde{\lambda}}\tau\int_{\tilde{\theta}}^{\bar{\theta}} y(\theta)k(\theta)d\theta + \frac{\tilde{\delta}\tilde{\lambda}}{(\tilde{\delta}+\tilde{\lambda})^2}(1-\tau)(1-K(\tilde{\theta}))\int_{\tilde{\theta}}^{\bar{\theta}} v(\theta)k(\theta)d\theta\Big] \tag{D.2}$$

where the constants A and B satisfy:

$$A = \Big[(1-K(\tilde{\theta}))\frac{\tilde{\lambda}}{\tilde{\delta}+\tilde{\lambda}} - (1-K(\tilde{\theta}))\frac{\tilde{\delta}\tilde{\lambda}}{(\tilde{\delta}+\tilde{\lambda})^2}\Big]$$
$$B = \Big[K(\tilde{\theta}) + (1-K(\tilde{\theta}))\frac{\tilde{\delta}}{\tilde{\delta}+\tilde{\lambda}} + (1-K(\tilde{\theta}))\frac{\tilde{\delta}\tilde{\lambda}}{(\tilde{\delta}+\tilde{\lambda})^2}\Big],$$

with $\tilde{\delta} := 1 - \rho(1-\delta)$ and $\tilde{\lambda} = \rho\lambda$.

Algorithm The algorithm combines an outer grid search over b and τ values and an inner step in which an equilibrium and the corresponding social welfare is computed at b and τ points on the grid. In successive runs progressively finer grids are applied to smaller neighborhoods of b and τ around the calculated optimum. The inner equilibrium calculation uses a grid of θ values. It has the following sub-steps.

Substep 0. For a given b and τ, set $\tilde{\theta} = \theta_1$. Set the substep counter $k = 1$ and proceed to the next substep.

Substep k. Using (D.2), compute L. Then, check if $\bar{q}(\theta_k) > 0$ using (D.1). If so, then set $\tilde{\theta} = \theta_k$, evaluate $\underline{q}(\theta)$ and $g(\cdot|\theta)$ for all $\theta \geq \tilde{\theta}$ using the formulas in the paper, compute social welfare and end the inner step. If not, set $k = k + 1$ and repeat sub step k.

E Numerical Algorithm for Nonlinear Tax Analysis

Recall the policy problem:

$$\sup_{T,b} \left(\frac{\delta}{\delta+\lambda} + \frac{\lambda}{\delta+\lambda} K[\tilde{\theta}(T)]\right) U(b) + \frac{\lambda}{\delta+\lambda} \int_{\tilde{\theta}(T)}^{\infty} \int_{\underline{q}(\theta;T)}^{\bar{q}(\theta;T)} U\left(\Phi(q,\theta;T)\right) g(q|\theta;T) dq k(\theta) d\theta. \tag{E.1}$$

subject to:

$$\mathcal{G} + \left(\frac{\delta}{\delta+\lambda} + \frac{\lambda}{\delta+\lambda} K[\tilde{\theta}(T)]\right) b = \frac{\lambda}{\lambda+\delta} \int_{\tilde{\theta}(T)}^{\infty} \int_{\underline{q}(\theta;T)}^{\bar{q}(\theta;T)} \{\tilde{q} + T[x(q,\theta;T)]\} g(q|\theta;T) dq k(\theta) d\theta,$$

Numerical Procedure

Step 0 Set parameters of the problem. These should include a large maximal income $\bar{x}$ and a large domain $\mathcal{D} := [0, \bar{q}] \times [\underline{\theta}, \bar{\theta}]$. Select an initial premium paid to the unemployed $r_1 = b_1 + T_1[0]$. Fix a tax amount $T_1[0]$ and derivative $T_1' : [0, \bar{x}] \to \mathbb{R}_+$. Hence, for $x \in [0, \bar{x}]$, $T_1[x] - T[0] = \int_0^x T_1'[x'] dx$. Proceed to Step 1.

Step $n > 0$

Substep a Approximate x_n on $\mathcal{D}$. At each $(q, \theta) \in \mathcal{D}$, set:

$$x_n(q,\theta) = \theta h'^{-1}\left(\theta\{1 - T_n'[x_n(q,\theta)]\}\right) - q.$$

Substep b Find $\tilde{\theta}_n$, where:

$$x_n(0,\tilde{\theta}_n) - \{T_n[x_n(0,\tilde{\theta}_n)] - T_n[0]\} - h\left(\frac{x_n(0,\tilde{\theta}_n)}{\tilde{\theta}_n}\right) = r_n. \tag{E.2}$$

Substep c On a domain $[\tilde{\theta}_n, \overline{\theta}]$, set $\overline{q}_n$:

$$x_n(\overline{q}_n(\theta),\theta) - h\left(\frac{x_n(\overline{q}_n(\theta),\theta) + \overline{q}_n(\theta)}{\theta}\right) = r_n. \tag{E.3}$$

and $\underline{q}_n(\theta) = \left(\frac{\delta}{\delta+\lambda}\right)^2 \overline{q}_n(\theta)$.

Substep e Define various functions used to construct integrals:

- $g_n(q|\theta) = \frac{1-\rho(1-\delta)}{2\rho\lambda}\sqrt{\frac{\overline{q}_n(\theta)}{q}}\frac{1}{q}$
- $\tilde{q}_n(\theta,x) = \theta(h')^{-1}(\{1-T_n'[x]\}\theta) - x$.
- $\underline{\theta}(x)$, where: $\tilde{q}_n(\underline{\theta}_n(x),x) = \underline{q}_n(\underline{\theta}_n(x)) = (\frac{\delta}{\delta+\lambda})^2\overline{q}_n(\underline{\theta}_n(x)))$.
- $\overline{\theta}_n(x) = \min\left(\overline{\theta}, \frac{1}{1-T_n'[x]}h'(h^{-1}\{x - T_n[x] - b\})\right)$.
- $\hat{q}_n(\theta,x) = \tilde{q}_n(\theta,x)$ if $\theta \in [\underline{\theta}(x), \overline{\theta}(x)]$ and $\hat{q}_n(\theta,x) = \overline{q}_n(\theta)$ if $\theta \in [\overline{\theta}(x), \overline{\theta}]$.
- $\Phi_n(q,\theta) = x_n(q,\theta) - T_n[x_n(q,\theta)] - h\left(\frac{x_n(q,\theta)+q}{\theta}\right)$.
- $\mathscr{E}_n(q,\theta) = \dfrac{1}{\frac{x_n(q,\theta)}{x_n(q,\theta)+q}\frac{\frac{x_n(q,\theta)+q}{\theta}h''\left(\frac{x_n(q,\theta)+q}{\theta}\right)}{h'\left(\frac{x_n(q,\theta)+q}{\theta}\right)} + \frac{T_n''[x_n(q,\theta)]x_n(q,\theta)}{1-T_n'[x_n(q,\theta)]}}$.
- $\mathcal{E}_n(q,\theta) = \frac{\theta}{x_n(q,\theta)}\frac{h'(x_n(q,\theta)/\theta)}{h''(x_n(q,\theta)/\theta)}$.
- $\frac{\mathcal{E}_{\overline{q},n}(\theta)}{1-T_n'[x_0]} = \frac{\theta}{h'\left(\frac{x_n(\overline{q}_n(\theta),\theta)+\overline{q}_n(\theta)}{\theta}\right)}\frac{1}{\overline{q}_n(\theta)}$, for $\theta \geq \overline{\theta}_n(x)$.

Substep f Update the marginal tax function $T'_{n+1}[\hat{x}]$ at each $\hat{x}$ on a grid of incomes from $[\tilde{x}_{n+1}, \overline{x}]$ according to:

$$\frac{x_0 T'_{n+1}[\hat{x}]}{1-T'_{n+1}[\hat{x}]}\overline{\mathscr{E}_n}(\hat{x})j_n(\hat{x}) = -\frac{1}{E[U'(\Phi_n)]}\int_{\underline{\theta}_n(\hat{x})}^{\overline{\theta}}\int_{\underline{q}_n(\theta)}^{\hat{q}_n(\theta,\hat{x})} U'\left(\Phi_n(q,\theta)\right) g_n(q|\theta)k(\theta)dqd\theta$$
$$+\frac{1}{E[U'(\Phi_n)]}\int_{\overline{\theta}_n(\hat{x})}^{\overline{\theta}}\frac{\mathcal{E}_{\overline{q},n}(\theta)}{1-T'_n[\hat{x}]}\int_{\underline{q}_n(\theta)}^{\overline{q}_n(\theta)} U'\left(\Phi_n(q,\theta)\right)(1-T'_n[x_n(q,\theta)])qg_n(q|\theta)dqk(\theta)d\theta$$
$$+(1-J_n[x_0]) - \int_{\overline{\theta}_n(x_0)}^{\overline{\theta}}\frac{\mathcal{E}_{\overline{q},n}(\theta)}{1-T'_n[x_0]}\int_{\underline{q}_n(\theta)}^{\overline{q}_n(\theta)} qg_n(q|\theta)dqk(\theta)d\theta$$
$$+\int_{\overline{\theta}_n(x_0)}^{\overline{\theta}}\frac{\mathcal{E}_{\overline{q},n}(\theta)}{1-T'_n[x_0]}\int_{\underline{q}_n(\theta)}^{\overline{q}_n(\theta)} T'_n[x_n(q,\theta)]\left\{\frac{x_n(q,\theta)}{x_n(q,\theta)+q}\frac{\mathscr{E}_n(q,\theta)}{\mathcal{E}_n(q,\theta)}\right\} qg_n(q|\theta)dqk(\theta)d\theta,$$

where:

$$1-J_n[x] := \int_{\underline{\theta}_n(x)}^{\overline{\theta}_n(x)}\int_{\underline{q}_n(\theta)}^{\tilde{q}_n(\theta,x)} g_n(q|\theta)dqk(\theta)d\theta + \int_{\overline{\theta}_n(x)}^{\overline{\theta}}\int_{\underline{q}_n(\theta)}^{\overline{q}_n(\theta)} g_n(q|\theta)dqk(\theta)d\theta$$

and $j_n(\cdot)$ is the corresponding income density and:

$$\overline{\mathscr{E}}_n(x_0)j_n(x_0) = \int_{\underline{\theta}_n(x_0)}^{\overline{\theta}_n(x_0)} \mathscr{E}_n(\tilde{q}_n(x_0,\theta),\theta)g_n(\tilde{q}_n(x_0,\theta)|\theta)k(\theta)d\theta.$$

Substep g Rearrange (B.20) and use this to update the premium r_{n+1} according to:

$$k[\tilde{\theta}_n]\frac{r_{n+1}-T_n[\tilde{x}_n]+T_n[0]}{h'\left(\frac{\tilde{x}_n}{\tilde{\theta}_n}\right)\frac{\tilde{x}_n}{\tilde{\theta}_n^2}}$$
$$= -\frac{1}{E_n[U'(\Phi_n)]}\int_{\tilde{\theta}_n}^{\infty}\int_{\underline{q}_n(\theta)}^{\overline{q}_n(\theta)} U'\left(\Phi_n(q,\theta)\right) g_n(q|\theta)k(\theta)dqd\theta$$
$$+\frac{1}{E_n[U'(\Phi_n)]}\int_{\tilde{\theta}_n}^{\infty}\frac{\mathcal{E}_{\overline{q},n}(\theta)}{1-T'_n[\tilde{x}_n]}\int_{\underline{q}_n(\theta)}^{\overline{q}_n(\theta)} U'\left(\Phi_n(q,\theta)\right)(1-T'_n[x_n(q,\theta)])qg_n(q|\theta)dqk(\theta)d\theta$$
$$+(1-J_n[\tilde{x}_n]) + \int_{\tilde{\theta}_n}^{\infty}\frac{\mathcal{E}_{\overline{q},n}(\theta)}{1-T'_n[\tilde{x}_n]}\int_{\underline{q}_n(\theta)}^{\overline{q}_n(\theta)} T'_n[x(q,\theta)]\left\{\frac{x_n(q,\theta)}{x_n(q,\theta)+q}\frac{\mathscr{E}_n(q,\theta)}{\mathcal{E}_n(q,\theta)}\right\} qg_n(q|\theta)dqk(\theta)d\theta$$
$$-\int_{\tilde{\theta}_n}^{\infty}\frac{\mathcal{E}_{\overline{q},n}(\theta)}{1-T'_n[\tilde{x}_n]}\int_{\underline{q}_n(\theta)}^{\overline{q}_n(\theta)} qg_n(q|\theta)dqk(\theta)d\theta.$$

Substep h Update the zero income tax amount to $T_{n+1}[0]$ using the budget constraint according to:

$$\begin{aligned}\mathcal{G}_{n+1} &+ \Big(\frac{\delta}{\delta+\lambda} + \frac{\lambda}{\delta+\lambda} K[\tilde{\theta}_n]\Big) r_{n+1} \\ &= \frac{\lambda}{\lambda+\delta} \int_{\tilde{\theta}_n}^{\infty} \int_{\underline{q}_n(\theta)}^{\overline{q}_n(\theta)} \{q + T_n[x_n(q,\theta)] - T_n[0]\}\, g_n(q|\theta) dq k(\theta) d\theta + T_{n+1}[0],\end{aligned}$$

where $\mathcal{G}_{n+1}$ is set to a specified fraction of the total surplus:

$$\frac{\lambda}{\lambda+\delta} \int_{\tilde{\theta}_n}^{\infty} \int_{\underline{q}_n(\theta)}^{\overline{q}_n(\theta)} \{q + x_n(q,\theta)\}\, g_n(q|\theta) dq k(\theta) d\theta.$$

Substep l Calculate *test* $:= \sup_x |T_{n+1}(x) - T_n(x)|$. If test $< \varepsilon$ finish. Otherwise, set $n = n+1$ and return to Step n.

F Connecting Income and Talent Distributions in Frictional Economies

The Pareto coefficient of the income and talent distributions are defined respectively by the functions $\frac{j(x)x}{1-J[x]}$ and $\frac{k(\theta)\theta}{1-K[\theta]}$. The former coefficient plays an important role in the optimal nonlinear tax formula presented in this paper. Proposition F 1 connects these Pareto coefficients in the frictional economy by an explicit formula. An implication of this formula is that if tax functions are affine (or asymptotically affine), then a talent distribution with a Paretian right tail implies an income distribution with a limiting right tail Pareto coefficient that is independent of the degree of labor market frictions.

To derive and express the Pareto coefficient formula some notation is needed. Define $t = \frac{\lambda}{\delta}$ to be the labor market frictions ratio. Recall that $\tilde{q}(\theta, x) = (1-\tau)^{\frac{1}{\gamma}} \theta^{\frac{1+\gamma}{\gamma}} - x$, $\overline{q}(\theta) = -\frac{r}{1-\tau} + \frac{\gamma}{1+\gamma}(1-\tau)^{\frac{1}{\gamma}} \theta^{\frac{1+\gamma}{\gamma}}$ and, using the definition of t, $\underline{q}(\theta) = \left(\frac{\delta}{\delta+\lambda}\right)^2 \overline{q}(\theta) = \left(\frac{1}{1+t}\right)^2 \overline{q}(\theta)$. Further recall that $\underline{q}(\underline{\theta}(x)) = \tilde{q}(\underline{\theta}(x), x)$. It is convenient to extend the previous definition to allow for values $s \neq t$ and to define $\underline{\theta}(s, x)$ such that:

$$\begin{aligned}\tilde{q}(\underline{\theta}(s,x), x) &= (1-\tau)^{\frac{1}{\gamma}} \underline{\theta}(s,x)^{\frac{1+\gamma}{\gamma}} - x \\ &= \left(\frac{1}{1+s}\right)^2 \left\{ -\frac{r}{1-\tau} + \frac{\gamma}{1+\gamma}(1-\tau)^{\frac{1}{\gamma}} \underline{\theta}(s,x)^{\frac{1+\gamma}{\gamma}} \right\} = \underline{q}(\underline{\theta}(s,x); s).\end{aligned}$$

Proposition F 1. *If taxes are affine, then the Pareto coefficients of the income and* *talent distributions are related by:*

$$\frac{j[x]x}{1-J[x]} = \frac{\gamma}{1+\gamma}\frac{\int_0^t \frac{x}{x-\left(\frac{1}{1+s}\right)^2 r}\frac{k(\underline{\theta}(s,x))\underline{\theta}(x,s)}{1-K[\underline{\theta}(x,s)]}(1-K[\underline{\theta}(x,s)])ds}{\int_0^t (1-K[\underline{\theta}(x,s)])ds}.$$

In particular, if the talent distribution has a Paretian right tail with coefficient α_θ, then the income distribution is asymptotically Paretian with coefficient $\alpha_x = \frac{\gamma}{1+\gamma}\alpha_\theta$. In particular, the limiting right tail Pareto coefficient of income is independent of the frictions parameter λ/δ in this case.

Proof. Recall the expression for J:

$$J[x] = K[\underline{\theta}(x)] + \int_{\underline{\theta}(x)}^{\overline{\theta}(x)} \left\{ \int_{\tilde{q}(\theta;x)}^{\overline{q}(\theta)} g(q|\theta)dq \right\} k(\theta)d\theta,$$

where:

$$g(q|\theta) = \frac{\delta}{2\lambda}\frac{\overline{q}(\theta)^{\frac{1}{2}}}{q^{\frac{3}{2}}}.$$

Hence, the income density is given by:

$$j[x] = \int_{\underline{\theta}(x)}^{\overline{\theta}(x)} \frac{\delta}{2\lambda}\frac{\overline{q}(\theta)^{\frac{1}{2}}}{\tilde{q}(\theta,x)^{\frac{3}{2}}}k(\theta)d\theta,$$

Recall also that: $\tilde{q}(\theta,x) = (1-\tau)^{\frac{1}{\gamma}}\theta^{\frac{1+\gamma}{\gamma}} - x$, $\overline{q}(\theta) = -\frac{r}{1-\tau} + \frac{\gamma}{1+\gamma}(1-\tau)^{\frac{1}{\gamma}}\theta^{\frac{1+\gamma}{\gamma}}$, $\underline{q}(\theta) = \left(\frac{\delta}{\delta+\lambda}\right)^2 \overline{q}(\theta)$, $\overline{q}(\overline{\theta}(x)) = \tilde{q}(\overline{\theta}(x),x)$ and $\underline{q}(\underline{\theta}(x)) = \tilde{q}(\underline{\theta}(x),x)$. It is convenient to define $t := \frac{\lambda}{\delta}$. Thus, the income density can be written as:

$$j[x] = \frac{1}{2t}\int_{\underline{\theta}(t;x)}^{\overline{\theta}(x)} \frac{\overline{q}(\theta)^{\frac{1}{2}}}{\tilde{q}(\theta,x)^{\frac{3}{2}}}k(\theta)d\theta,$$

where the dependence of $\underline{\theta}$ on t is made explicit in the notation. (Note that $\overline{\theta}$ does not depend on t). Using the fact that $\underline{\theta}(0,x) = \overline{\theta}(x)$, after a change of variables the income density is given by:

$$j[x] = -\frac{1}{2t}\int_0^t \frac{\overline{q}(\underline{\theta}(s,x))^{\frac{1}{2}}}{\tilde{q}(\underline{\theta}(s,x),x)^{\frac{3}{2}}}k(\underline{\theta}(s,x))\underline{\theta}'(s)ds.$$

Using: $\tilde{q}(\underline{\theta}(s,x),x) = \underline{q}(\underline{\theta}(s,x)) = \left(\frac{1}{1+s}\right)^2\overline{q}(\underline{\theta}(s,x))$, the expression for the density reduces to:

$$j[x] = -\frac{1}{2t}\int_0^t \frac{1}{\left(\frac{1}{1+s}\right)^3\overline{q}(\underline{\theta}(s,x))}k(\underline{\theta}(s,x))\frac{\partial\underline{\theta}}{\partial t}(x,s)ds.$$

After some algebra:

$$\frac{\partial \underline{\theta}}{\partial t}(x,s) = -2\frac{\gamma}{1+\gamma}\frac{\underline{\theta}(x,s)}{x-\left(\frac{1}{1+s}\right)^2 r}\frac{\overline{q}(\underline{\theta}(x,s))}{(1+s)^3}.$$

Combining terms:

$$j[x] = \frac{\gamma}{1+\gamma}\frac{1}{t}\int_0^t \frac{\underline{\theta}(x,s)}{x-\left(\frac{1}{1+s}\right)^2 r}k(\underline{\theta}(s,x))ds.$$

Thus,

$$j[x]x = \frac{\gamma}{1+\gamma}\frac{1}{t}\int_0^t \frac{x}{x-\left(\frac{1}{1+s}\right)^2 r}k(\underline{\theta}(s,x))\underline{\theta}(x,s)ds. \tag{F.1}$$

In particular, note that as $t \to \infty$, the right hand side converges to:

$$\frac{\gamma}{1+\gamma}k(\underline{\theta}(x))\underline{\theta}(x),$$

where $\underline{\theta}(x) := (1-\tau)^{\frac{1}{\gamma}}x^{\frac{1+\gamma}{\gamma}}$. Next reconsider:

$$J[x] = K[\underline{\theta}(x)] + \int_{\underline{\theta}(x)}^{\overline{\theta}(x)}\left\{\int_{\tilde{q}(\theta;x)}^{\overline{q}(\theta)}\frac{\delta}{2\lambda}\frac{\overline{q}(\theta)^{\frac{1}{2}}}{q^{\frac{3}{2}}}dq\right\}k(\theta)d\theta.$$

Define $M(\theta)$ to be the value of the inner right hand side integral:

$$M(\theta) := \int_{\tilde{q}(\theta;x)}^{\overline{q}(\theta)}\frac{\delta}{2\lambda}\frac{\overline{q}(\theta)^{\frac{1}{2}}}{q^{\frac{3}{2}}}dq = 2\left[\left(\frac{\overline{q}(\theta)}{\tilde{q}(\theta,x)}\right)^{\frac{1}{2}} - 1\right]$$

and let $m(\theta) = M'(\theta)$. We may compactly write $J[x]$ as:

$$J[x] = K[\underline{\theta}(x)] + \frac{1}{2t}\int_{\underline{\theta}(x)}^{\overline{\theta}(x)} M(\theta)k(\theta)d\theta$$

and, after integrating by parts, obtain:

$$J[x] = -\frac{1}{2t}\int_{\underline{\theta}(x)}^{\overline{\theta}(x)} m(\theta)K[\theta]d\theta.$$

Next solve for m:

$$m(\theta) = -\left(\frac{\overline{q}(\theta)^{\frac{1}{2}}}{\tilde{q}(\theta,x)^{\frac{3}{2}}}\right)\left\{\frac{1+\gamma}{\gamma} - \frac{\tilde{q}(\theta,x)}{\overline{q}(\theta)}\right\}(1-\tau)^{\frac{1}{\gamma}}\theta^{\frac{1}{\gamma}}.$$

Thus,

$$1 - J[x] = 1 - \frac{1}{2t}\int_{\underline{\theta}(x)}^{\overline{\theta}(x)} \left(\frac{\overline{q}(\theta)^{\frac{1}{2}}}{\tilde{q}(\theta,x)^{\frac{3}{2}}}\right)\left\{\frac{1+\gamma}{\gamma} - \frac{\tilde{q}(\theta,x)}{\overline{q}(\theta)}\right\}(1-\tau)^{\frac{1}{\gamma}}\theta^{\frac{1}{\gamma}}K[\theta]d\theta.$$

We again change variables and re-express the right hand integral as:

$$\frac{1}{2t}\int_0^t \frac{\partial\underline{\theta}}{\partial t}(x,s)\left(\frac{1}{\frac{\overline{q}(\underline{\theta}(x,s),x)}{(1+s)^3}}\right)\left\{\frac{1+\gamma}{\gamma} - \left(\frac{1}{1+s}\right)^2\right\}(1-\tau)^{\frac{1}{\gamma}}\underline{\theta}(s)^{\frac{1}{\gamma}}K[\underline{\theta}(s)]d\theta.$$

Substituting for $\frac{\partial\underline{\theta}}{\partial t}(x,s)$ reduces the previous expression to:

$$-\frac{1}{t}\int_0^t K[\underline{\theta}(s,x)]ds.$$

Hence,

$$1 - J[x] = \frac{1}{t}\int_0^t (1 - K[\underline{\theta}(x,s)])ds.$$

Combining this with (F.1) gives:

$$\frac{j[x]x}{1-J[x]} = \frac{\gamma}{1+\gamma}\frac{\int_0^t \frac{x}{x-\left(\frac{1}{1+s}\right)^2 r}k(\underline{\theta}(s,x))\underline{\theta}(x,s)ds}{\int_0^t (1-K[\underline{\theta}(x,s)])ds}.$$

Or:

$$\frac{j[x]x}{1-J[x]} = \frac{\gamma}{1+\gamma}\frac{\int_0^t \frac{x}{x-\left(\frac{1}{1+s}\right)^2 r}\frac{k(\underline{\theta}(s,x))\underline{\theta}(x,s)}{1-K[\underline{\theta}(x,s)]}(1-K[\underline{\theta}(x,s)])ds}{\int_0^t (1-K[\underline{\theta}(x,s)])ds}. \tag{F.2}$$

As x becomes large, the ratio $\frac{x}{x-\left(\frac{1}{1+s}\right)^2 r}$ converges to 1 and the right hand side of (F.2) converges to:

$$\frac{\gamma}{1+\gamma}\frac{\int_0^t \frac{k(\underline{\theta}(s,x))\underline{\theta}(x,s)}{1-K[\underline{\theta}(x,s)]}(1-K[\underline{\theta}(x,s)])ds}{\int_0^t (1-K[\underline{\theta}(x,s)])ds}.$$

Thus at large x, the income Pareto coefficient is (approximately) a scaled average of talent Pareto coefficients. In particular, if the talent distribution has a Pareto tail with coefficient α_θ, then the income distribution is also asymptotically Paretian with limiting coefficient $\alpha_x = \frac{\gamma}{1+\gamma}\alpha_\theta$. □

Appendix References

Heathcote, J., F. Perri, G. Violante (2010). Unequal we stand: An empirical analysis of economic inequality on the United States, 1967-2006. *Review of Economic Dynamics 13*(1), 15–51.