Tax Design with Labor Market Frictions*

Christopher Sleet[†]

University of Rochester

Hakkı Yazıcı[‡]

University of Bristol

This Version: April 15, 2025

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. Increases in income tax rates squeeze ex post firm profits, suppressing profit tax revenues and vacancy creation and redistributing differentially to workers on job ladders and across talent markets. These effects shape and modify conventional optimal tax formulas. 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 frictions create a motive for moderately lower marginal income tax rates, but a higher tax wedge between work and non-work.

JEL classification: D31, H21, H24, J60, J62, J65.
Keywords: Taxation, frictional labor markets, redistribution.

*We thank seminar participants at Aarhus, Bocconi, Bristol, CMU, Carlos III de Madrid, CERGE-EI, CREI, CREST, Essex, EUI, Harvard, Koç, LMU, Queen Mary University London, Rice, Sabanci, St Andrews, Tinbergen Institute, Universitat Autonoma de Barcelona, and audiences at various conferences and workshops. We especially thank Laurence Ales, Javier Fernandez-Blanco, Melvyn Coles, Davide Debortoli, Piero Gottardi, Nathan Hendren, Marek Kapička, Per Krusell, Iacopo Morchio, Chris Moser, Nicola Pavoni, Edouard Schaal, Ctirad Slavík and Stefanie Stantcheva for helpful comments and suggestions.

† csleet@ur.rochester.edu. Department of Economics, University of Rochester, Rochester, NY 14610, USA.

‡ hakki.yazici@bristol.ac.uk. 12 Priory Rd, Bristol BS8 1TU, United Kingdom.

1 Introduction

A worker's pay depends upon the surplus that she generates in production and the share of that surplus that she captures as income. She may be poor because she has low marketable talent or because she is matched with a firm that extracts most of the surplus that she produces (or both). The opportunity to search and (re)match with a firm that pays more and extracts less creates job ladders for equally talented workers. We explore the optimal design of tax and benefit policies in the face of both the ex post risk of climbing, falling off or getting stuck on the job ladder and the ex ante risk of being born with more or less talent. Our framework blends salient elements of the canonical frictionless optimal tax model (effort choice and talent variation) with those of frictional search models (on and off the job search, vacancy creation). In this setting, we evaluate tax incidence, derive optimal tax equations and quantitatively evaluate tax designs.

The canonical optimal tax model supposes a frictionless labor market in which all worker income variation is attributable to talent heterogeneity and intensive margin effort choice. Workers immediately find work and are paid their marginal product. In this setting, tax design trades off redistributive benefit against distortion on the effort margin. If firm technologies exhibit decreasing returns to scale, general equilibrium effects are present: Higher income tax rates discourage labor supply, raise wages and reduce both profits and profit tax revenues. However, assuming the policymaker can tax profits, terms describing such general equilibrium "profit squeeze" effects cancel from optimal tax equations.

Search frictions modify the transmission from income tax rates to worker and firm equilibrium income distributions. Higher labor income tax rates deter work and compel firms to pay more and extract less as they compete for workers. Income is shifted from the profit to the labor earnings base, where it is (optimally) taxed at a lower rate, thus suppressing tax revenues. However, such equilibrium effects impact the extensive rather than the intensive return to working. They no longer cancel from optimal tax equations. In addition, tax-induced equilibrium profit squeezes suppress vacancy and job creation. Absent direct tax or subsidization of vacancies, the desirability of such suppression depends on the efficiency of job creation at the underlying equilibrium. Complicating all of these effects, labor earnings are not uniformly enhanced by the profit squeeze: within talent markets, those at the bottom benefit more; across talent markets, the higher talented are the prime beneficiaries.

To elucidate these impacts of tax variation and derive optimal tax designs for frictional settings in the most transparent way, we focus in the main text on affine income tax (and benefit) designs and proceed in steps. Elaborations of formulas for nonlinear tax designs are supplied in the online appendix. We begin our analysis with a model that embeds a simple Burdett and Mortensen (1998) job ladder structure into an otherwise standard public finance framework. The model features heterogeneous workers distributed across talent markets, an intensive labor supply margin, taxes and search both on and off the job. Workers experience "market luck" unrelated to their talent: job destruction shocks consign them temporarily to unemployment, while randomly arriving opportunities to climb job ladders permit them to find less extractive employers.[1] We initially treat the job finding rate as a parameter. 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. It is, thus, constrained to select a tax function applied to the incomes of those who are in work and a benefit paid to those who are not. Firm profit is optimally taxed at 100%. Within this framework firms are naturally modeled as extracting a component of the surplus generated by a worker. We call this a "job price". In each talent market, tax and benefit policy determines a minimal utility that firms must deliver to workers and a corresponding maximal "job price" that they can extract. In choosing a job price below this maximum, firms must trade off profit extracted from each worker against the number of workers attracted and retained. Some firms make high job price offers, but attract and retain a small number of (lower earning) workers. Others make low job price offers, but attract and retain many (higher earning) workers. Thus, within-talent market job ladders and frictional wage dispersion emerge.

Increments to the income tax rate directly lower the return to work relative to inactivity. This, in turn, suppresses the maximal and all other job prices that firms can extract within each talent market. Such suppression shifts income from the profit to the labor earnings tax base. Since the latter is optimally taxed at a lower rate than the former this tax base shifting dampens the additional revenue generated by an income tax increment. However, in contrast to the frictionless model, because tax-induced general equilibrium effects modify job prices

[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.

and the return to a job, rather than the return to effective labor, they do not cancel from optimal tax equations.[2] To the extent that the revenues generated by income tax rate increments are dampened by general equilibrium effects in the frictional model, a force for lower optimal income tax rates is introduced. However, the tax-induced suppression of job prices is not uniform across workers. Within talent markets, it dissipates over the length of the job ladder implying intra-talent market redistribution from high to low earners. On the other hand, it strengthens with talent, implying inter-talent market redistribution from low to high earning talents. The overall impact of the profit squeeze depends on the balance of these effects. If the first redistributive effect is large, then a policymaker seeking to redistribute from high to low earners may favor greater income taxation.

In the environments described so far matching is exogenous and the supply of vacancies fixed. In frictional models with endogenous matching, a fraction of job price revenue is used by firms to pay vacancy costs and finance job creation. Firms internalize the costs of vacancy creation, but their assessment of the benefits (an expected job price, exclusive of the negative externalities they create for other firms) deviates from that of the planner (additional expected tax revenues and utility to previously unemployed job finders, inclusive of externalities). Failure to internalize the tax and worker benefits of additional jobs causes firms to under-post vacancies, failure to internalize the externalities linked to congestion on the new jobs margin and poaching from other firms on the pre-existing jobs margin causes firms to over-post. A tax-induced profit squeeze suppresses vacancy and, hence, job creation. Such suppression is desirable if vacancy creation is excessive and undesirable otherwise. To incorporate this aspect we utilize a Burdett-Mortensen model augmented with endogenous matching. In this setting, optimal tax formulas are modified to include the net marginal benefit or cost associated with vacancy suppression.

To assess the quantitative implications of labor frictions for optimal policy, we calibrate our model with endogenous matching to US income distribution data. This requires inverting the map 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. In frictionless models, such inversion is straightforward. In frictional models it is significantly compli-

[2] In particular, they do not elicit a stimulus to effective labor via higher pre-tax wages and an off-setting boost to revenues as occurs in the frictionless model.

cated by the intra-talent market dispersion of earnings along job ladders. We develop a procedure for inverting the distributional map in our frictional setting. The procedure draws on techniques for solving integral equations and embeds an integral inversion step inside a larger fixed point problem. We then compute an optimal affine income tax and benefit system for the calibrated frictional model. To explore the implications of abstracting from labor market frictions, we recalibrate and recompute optimal policy under the assumption that the data was generated by a frictionless model. We find that a utilitarian policymaker in the calibrated frictional economy sets the marginal income tax rate to 31.7% and the employment-non employment tax wedge to $6912 per annum in 2017 dollars.[3] In the frictionless model, the optimal marginal income tax rate increases moderately to 35.7%, while the employment-non employment tax wedge falls to $3360. We decompose the various countervailing forces underpinning these results. Specifically, we evaluate how, under alternative assumptions about frictions, varied forces contribute to the overall social marginal benefit of adjusting tax function parameters as they sweep over intervals containing the optimum. With respect to the marginal income tax rate, we find that the main distinction between environments is the lower redistributive benefit of tax rate increments in the frictional economy. Although the marginal costs (loss of tax revenues and worker rents) and benefits (correction of externalities amongst firms) from suppressing vacancies via the income tax rate in the frictional setting are separately significant the two largely offset and together are not large. In contrast, the marginal benefit of raising the employment-non employment tax wedge is greater under the assumption of frictions and this is underpinned by the benefit of suppressing vacancies in this case.

The remainder of the paper proceeds as follows. After a brief literature review, Section 2 considers affine tax design in a benchmark frictionless model. The section highlights the cancellation of general equilibrium effect terms from optimal tax equations and the irrelevance of the "frictionless" profit squeeze for the structure of these equations. Section 3 introduces a basic Burdett-Mortensen model with an intensive effort margin, elaborates tax incidence on firm and worker incomes, and derives optimal tax equations for this case. In Section 4 tax design in a frictional model with endogenous matching and job creation is considered. Section 5 calculates an optimal affine tax and benefit scheme for a calibrated

[3]The employment-non employment tax wedge $\tau_0 = T_0 + b$ is the sum of the affine income tax function's intercept at zero income, T_0, and the unemployment benefit, b.

frictional model with endogenous matching and job creation. It contrasts results with those obtained under the assumption that the data is generated by a frictionless model. Section 6 concludes. Proofs, calibration details and elaboration of results for the nonlinear income tax case are provided in 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. These papers and most other contributions to the optimal income tax literature adopt a frictionless specification of the labor market in which there is a single market (and price) for effective labor. Scheuer and Werning (2016) extend this analysis to settings with a non-linear equilibrium pricing function for (superstar) labor and show that standard optimal tax formulas apply. These results hold independently of the presence of firm profits, provided profit taxation is unconstrained. Standard formulas for taxes are modified when profit taxation is constrained (Munk (1978)) or when there are multiple (frictionless) labor markets and the policymaker is constrained to use a single tax function that cannot condition on the labor market in which income is earned (Stiglitz (1982), Rothschild and Scheuer (2013), Ales and Sleet (2015), Sachs, Tsyvinski, and Werquin (2020)).

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. Redistributive considerations are omitted. Hungerbühler et al (2006) augments the framework of Boone and Bovenberg (2002) with variation in worker talent (but omits search effort) and reinstates redistributional concerns. In this setting, workers are distributed over endowments of effective labor, which they costlessly supply to their employer if matched. In both Boone and Bovenberg (2002) and Hungerbühler et al (2006), income taxes are transmitted to job prices and firm quasi-rents via a Nash bargain. In the latter optimal (nonlinear) income 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). Golosov et al (2013) consider policy design in a model with directed search and no ex ante talent heterogeneity. Our paper complements 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.[4] It is also partly motivated by the observation of Hornstein et al (2011) that large unemployment to work transitions and the relatively short duration of unemployment is inconsistent with 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 Affine tax design in a frictionless economy

As a precursor to analysis of the income-tax induced profit squeeze in frictional settings, we first review its operation and implications for tax design in a benchmark frictionless setting. In this setting, we assume a decreasing returns to scale production technology. Wages are then determined endogenously with pre-tax profits accruing to firm owners. Increments to labor income taxes reduce labor supply, raise wages, and lower firm profit. This squeeze spills over to the profit tax base. However, terms describing it cancel from optimal tax equations leaving standard (exogenous wage) tax formulas intact. This result does not survive the introduction of frictions.

Environment Workers are partitioned across talents $\theta \in \Theta = (\underline{\theta}, \overline{\theta}) \subset \mathbb{R}_+$ according to a distribution K with density k. A worker of talent θ who consumes c and supplies effective labor z obtains payoff $U(c, z, \theta)$, with $U : \mathbb{R}^2_+ \times \mathbb{R} \to \mathbb{R}$ increasing in c, decreasing in z, and strictly concave and twice continuously differentiable in (c, z). In addition, U has a marginal rate of substitution $-\frac{\partial U}{\partial c} \Big/ \frac{\partial U}{\partial z}$ that is decreasing in θ and, hence, satisfies the usual single crossing property. Let $x = wz$ denote a worker's pre-tax income, with w the pre-tax wage, and let $T[x] = T_0 + \tau x$ denote an affine income tax. It is convenient to formulate the pol-

[4] 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.

icy problem in terms of after-tax quantities and then back out implications for taxes. To that end let $\xi = -T_0$ denote a worker's after-tax lump sum or virtual income and $\omega = (1-\tau)w$ the after-tax wage. Let $\psi = (\xi, \omega)$ denote the corresponding after-tax income-wage pair. The labor supply problem of a θ-talent worker is then:

$$z(\theta;\psi) := \arg\max_z U(\xi + \omega z, z, \theta). \tag{1}$$

Competitive firms operate a technology $F : \mathbb{R}_+ \to \mathbb{R}_+$ that converts effective labor to output, with F increasing, concave and differentiable in effective labor. Given a wage w, a representative competitive firm hires effective labor to maximize profits $F(z) - wz$. The policymaker faces an exogenous revenue requirement: $G \in \mathbb{R}_+$.

We adopt the conventional assumption of 100% taxation of profit, as in Diamond and Mirrlees (1971).[5] It is easy to verify that $\psi = (\xi, \omega)$ is consistent with a competitive equilibrium in which workers maximize utility, firms maximize profits, profits are taxed at 100% and the policymaker's budget is in surplus if and only if the following resource constraint holds at the utility maximizing labor supply function $z(\cdot;\psi)$:

$$-\xi + F\left(\int_\Theta z(\theta;\psi)k(\theta)d\theta\right) - \omega \int_\Theta z(\theta;\psi)k(\theta)d\theta \geq G. \tag{2}$$

Given such a ψ, the corresponding income tax parameters (T_0, τ) and equilibrium wage w are recovered from definitions and the firm optimality condition:

$$T_0 = -\xi \qquad \text{and} \qquad \frac{\partial F}{\partial z}\left(\int_\Theta z(\theta;\psi)k(\theta)d\theta\right) = w = \frac{\omega}{1-\tau}. \tag{3}$$

Policy problem and optimal tax equations A policymaker maximizes a Γ-Pareto weighted integral of worker utilities. We impose no restrictions on Γ at this point, but later restrict attention to non-increasing weighting functions. The policymaker's choice problem, cast in terms of $\psi = (\xi, \omega)$, is:

$$\max_{\psi=(\xi,\omega)} \int_\Theta \Gamma(\theta) U(\xi + \omega z(\theta;\psi), z(\theta;\psi), \theta)\, k(\theta)d\theta \tag{4}$$

[5] This assumption simplifies, but is not essential for our analysis. Our results proceed if profits are distributed amongst households and profit taxes are optimally set to levels below 100%. The essential requirement is that the policymaker has no incentive to use income taxation to extract resources from firm owners by manipulating pre-tax wages.

subject to the resource constraint (2). Suppose for now that all workers make an interior effective labor effort choice (i.e. all work) and that each $z(\cdot;\psi)$ is locally smooth at the optimum. Omitting explicit dependence of variables on policy ψ and formatting the policymaker's first order condition with respect to ω in terms of incomes x implies:

$$-\int_{\mathbb{R}_+}\left\{\frac{M(x)}{\Lambda}-\frac{\tau}{1-\tau}\eta(x)-1\right\}xh(x)dx=\frac{\tau}{1-\tau}\int_{\mathbb{R}_+}x\mathcal{E}^c_{x,\omega}(x)h(x)dx, \tag{5}$$

where Λ is the optimal multiplier on the resource constraint and, for income x, $M(x)$ is the marginal social welfare weight, $-\frac{\eta(x)}{1-\tau}$ is the impact of an extra unit of lump sum income on pre-tax earnings, $h(x)$ is the earnings density, and $\mathcal{E}^c_{x,\omega}(x)$ is the compensated earnings elasticity. This formulation replicates the classic organization of Diamond (1975), with the left hand side giving the redistributive benefit and right hand side the marginal deadweight loss associated with a reduction in the after-tax price of labor. The form of the optimal tax equation (5) is not affected by the presence of profit or endogenous prices, though these elements will generally affect the values of h, η and $\mathcal{E}^c_{x,w}$ via their impact on allocations.[6]

Abstracting from income effects and using $\omega=(1-\tau)w$, expression (5) can be recast as a first order condition for τ:

$$\left(-1+\frac{1-\tau}{w}\frac{\partial w}{\partial\tau}\right)\left\{\int_{\mathbb{R}_+}\left\{\frac{M(x)}{\Lambda}-1\right\}xh(x)dx+\frac{\tau}{1-\tau}\int_{\mathbb{R}_+}x\mathcal{E}^c_{x,\omega}(x)h(x)dx\right\}=0. \tag{6}$$

The term $\left(-1+\frac{1-\tau}{w}\frac{\partial w}{\partial\tau}\right)$ in (6) combines the direct negative effect of a tax rate increment on after-tax wages (-1) with the indirect general equilibrium effect $\left(\frac{1-\tau}{w}\frac{\partial w}{\partial\tau}\right)$. Thus, general equilibrium effects modify the transmission from tax rates to after-tax wages, but not the requirement that the latter are set to balance redistributive and incentive costs and benefits. Under standard conditions, these general equilibrium effects involve a profit squeeze: a higher tax rate depresses labor supply, raises pre-tax wages, and reduces profit. Income is correspondingly shifted from the high tax profit to the low tax labor earnings base. This reduces tax revenues and affects the distribution of after-tax incomes: funding for lump sum transfers falls, while pre-tax wages rise, the latter benefiting higher earners. The term $\frac{1-\tau}{w}\frac{\partial w}{\partial\tau}\int\left\{\frac{M(x)}{\Lambda}-1\right\}xh(x)dx$ in (6), describes the welfare impli-

[6]This logic extends to much more complex environments with non-linear tax systems and competitive, but non-linear pre-tax pricing of effective labor, see Scheuer and Werning (2017).

cations of this redistribution. The general equilibrium increase in pre-tax wages also elicits a labor supply response that mitigates the direct behavioral impact of taxes on revenues. This is described by $\frac{1-\tau}{w}\frac{\partial w}{\partial \tau}\frac{\tau}{1-\tau}\int x\mathcal{E}^c_{x,\omega}(x)h(x)dx$ in (6). Collectively, these terms are a scaled version of the welfare impact of an after-tax wage perturbation and so cancel from (6) to yield (5). In later frictional settings, general equilibrium effects will also imply firm-level profit squeezes. However, they will not operate via pre-tax wages and, in particular, they will not induce a mitigating labor supply response. Such effects will not cancel from optimal tax equations in these settings.

Exogenous and endogenous inactivity; Adding benefits for non-work To facilitate comparison with the frictional economy that follows, it is useful to adjust the benchmark model in several ways. First, assume that a fraction of agents $1-\mu$ at each talent are exogenously assigned to inactivity. In our basic frictional model to follow this assignment is attributed to the presence of frictions. The second adjustment is an extension of the tax system to include a benefit for non-work b. The tax-benefit system is then described by a triple (b, T_0, τ). It is useful to reformat this as a basic income b paid to all agents, a tax $\tau_0 = T_0 + b$ paid by agents conditional on working, but independent of hours worked and a marginal income tax τ as before. The tax τ_0 operates on the extensive margin deterring participation, while τ operates on the intensive margin. The third adjustment is to explicitly incorporate an extensive margin decision not to work. Under our assumptions a talent inactivity threshold $\tilde{\theta} \in [\underline{\theta}, \overline{\theta})$ emerges below which workers do not work and above which they do. An interior threshold satisfies:

$$U\left(b - \tau_0 + \omega z(\tilde{\theta};\psi), z(\tilde{\theta};\psi), \tilde{\theta}\right) = U(b, 0, \tilde{\theta}). \tag{7}$$

The right hand side of (5) is then augmented to include the tax revenues obtained by activating the marginal talent through a higher after-tax wage: $\frac{\tau_0+\tau\tilde{x}}{\omega}\frac{\tilde{\theta}k(\tilde{\theta})}{1-K(\tilde{\theta})}\mathcal{E}_{\tilde{\theta},\omega}$, with $\mathcal{E}_{\tilde{\theta},\omega}$ the (negative) elasticity of the threshold $\tilde{\theta}$ with respect to ω and $\tilde{x} = wz(\tilde{\theta}, \psi)$. Specifically,

$$\begin{aligned} &-\int_{\mathbb{R}_+}\left\{\frac{M(x)}{\Lambda} - \frac{\tau}{1-\tau}\eta(x) - 1\right\} xh(x)dx \\ &\qquad = \int_{\mathbb{R}_+}\frac{\tau}{1-\tau}x\mathcal{E}^c_{x,\omega}(x)h(x)dx + \frac{\tau_0+\tau\tilde{x}}{1-\tau}\frac{\tilde{\theta}k(\tilde{\theta})}{1-K(\tilde{\theta})}\mathcal{E}_{\tilde{\theta},\omega}, \end{aligned} \tag{8}$$

with h the earnings density conditional on employment. A related optimal tax equation holds for the participation tax τ_0:

$$-\int_{\mathbb{R}_+}\left\{\frac{M(x)}{\Lambda}-\frac{\tau}{1-\tau}\eta(x)-1\right\}h(x)dx=\frac{\tau_0+\tau\tilde{x}}{1-\tau}\frac{\tilde{\theta}k(\tilde{\theta})}{1-K(\tilde{\theta})}\mathcal{E}_{\tilde{\theta},\tau_0},\tag{9}$$

where $\mathcal{E}_{\tilde{\theta},\tau_0}$ is the (negative) elasticity of threshold $\tilde{\theta}$ with respect to τ_0.

3 Affine income tax design in frictional economies with exogenous matching

Frictions prevent workers from immediately finding work and, by impeding competition, from extracting all of the surplus created when they do. Some surplus accrues to firms as profit or as payment towards an upfront job creation cost. Surplus division is determined in general equilibrium and is shaped by tax policy since this impacts a worker's return to work relative to its outside options. Moreover surplus division may be heterogeneous reflecting a worker's position on the job ladder or relative bargaining power. Heterogeneity in surplus division confounds policymaker attempts to control both a worker's lump sum income and after-tax wage with an (affine) income tax.

To focus analysis on the profit squeeze channel, we first introduce results via a basic Burdett-Mortensen (bBM) model. The model features on and off the job search, elastic labor supply, tax policy, and exogenous job matching. It is the kernel of many contributions to the structural search literature and, hence, is a natural framework to focus upon. In macroeconomics, it has been used to reconcile the short duration of unemployment with the extent of frictional income dispersion. In later sections we augment the bBM model with endogenous matching and vacancy posting.

3.1 Environment

Time is continuous and attention restricted to steady state equilibria and time invariant policy choices. Policy consists of a benefit b paid to unemployed workers and an affine tax function $T[x]=T_0+\tau x$ applied to employed workers. It will (again) be convenient to summarize a policy as a tuple $\mathcal{P}=(b,\tau_0,\tau)$, where b is a basic income received by all, $\tau_0=T_0+b$ is a tax contingent on working and τ

is the marginal income tax. Workers and firms trade effective labor for income in frictional labor markets segmented by talent θ. On-the-job search in combination with frictional matching creates a steady state equilibrium distribution of workers over job prices.

Workers Employed and unemployed workers are partitioned across talent specific markets θ. Firms own simple linear technologies that map effective labor z one-for-one into output. They make take-it-or-leave-it "job price" offers q to workers. A worker of talent θ that locates and accepts a job price offer q and supplies effective labor z earns pre-tax income $x = z - q$. Given the tax policy described above, the worker selects effective labor z to solve:[7]

$$V(q, \theta; \mathcal{P}) := \max_{z} U\left(b - \tau_0 + (1-\tau)(z-q), z, \theta\right). \tag{10}$$

Remark 1. In the standard bBM model on-the-job effective labor is abstracted from and the firm is modeled as posting an income rather than a job price offer. This is equivalent to the firm posting a job price and the worker receiving the residual surplus as income. In our extended model with effective labor, job price posting by firms and effective labor selection by workers aligns more closely with the frictionless framework from the preceding section. The function $V(\cdot, \theta; \mathcal{P})$ in (10) defines the after-tax Pareto frontier of job prices and worker utilities available to matched firm/worker pairs of talent θ given policy $\mathcal{P}$. Thus, firm job price posting and worker effective labor selection is equivalent to the posting of an income/effective labor contract by the firm that secures a point on this Pareto frontier. Formulated in this way, the Burdett-Mortensen framework supplies a particular model of job price and, hence, surplus division that attains the after-tax Pareto frontier. Other, e.g. bargaining, models supply alternative division arrangements.

Firms are distinguished by the job prices that they charge and, hence, the extractiveness of their behavior. Equilibrium variation in extractiveness generates an offer distribution of job prices $F[q|\theta, \mathcal{P}]$ within each talent market given $\mathcal{P}$. A worker in talent market θ makes contact with a (new) firm with probability λ in-

[7] Our maintained assumption is that workers and firms match, agree on a surplus division and then workers supply effective labor and produce. Surplus is divided in line with the worker-firm agreement. This contrasts with models in which workers exert effort in skill formation or search prior to matching. In addition, we abstract from informational problems between workers and firms.

dependent of employment status or talent. Conditional on contact, this firm and its job price q are drawn from $F[\cdot|\theta,\mathcal{P}]$. Acceptance of q followed by selection of effective labor implies a worker payoff $V(q,\theta;\mathcal{P})$. An employed worker will accept a job at a newly encountered firm (and climb the job ladder) if that firm offers a lower job price and, hence, higher $V(q,\theta;\mathcal{P})$ than that provided by its current employer. An unemployed worker will accept if the implied payoff exceeds that from unemployment: $V(q,\theta;\mathcal{P}) > U(b,0,\theta)$.[8] Central to subsequent analysis is the maximum possible job price in each talent market given a policy: $\bar{q}(\theta;\mathcal{P})$. This job price makes a worker indifferent between employment and unemployment. It satisfies:

$$V(\bar{q}(\theta;\mathcal{P}),\theta;\mathcal{P}) := \max_z U\left(b-\tau_0+(1-\tau)(z-\bar{q}(\theta;\mathcal{P})),z,\theta\right) = U(b,0,\theta). \tag{11}$$

Expression (11) connects the maximal job price in each talent market to policy. As elaborated below, this connection is central to the transmission of tax policy to firm profit in this environment.

Jobs are destroyed and matched workers returned to unemployment at an exogenous rate δ. The ratio λ/δ parameterizes frictions in the bBM model. In the limit as $\lambda/\delta \to \infty$, the model approaches the frictionless one.

Firms Firms post and commit to a job price q in a talent market θ. Let $N(q,\theta;\mathcal{P})$ denote the steady state population of workers at a firm posting job price q in talent market θ given policy $\mathcal{P}$. Total firm profits at such a firm are:

$$\pi(q,\theta;\mathcal{P}) = N(q,\theta;\mathcal{P})q. \tag{12}$$

Firms select a θ (the talent market in which they post) and a q (the job price that they post in that talent market) to maximize $\pi(\cdot,\cdot;\mathcal{P})$.

3.2 Equilibria and tax incidence

No firm enters a market in which the maximal job price that a worker will accept is negative. Non-negativity of a job price defines an activity threshold as in (7). We refer to talent markets $\theta \in (\tilde{\theta},\overline{\theta}]$ as *active* and talent markets $\theta \in [\underline{\theta},\tilde{\theta}]$ as *inactive*. All workers are unemployed and, effectively, out of the labor force in in-

[8]Note that since contact rates are independent of employment status, the decision to accept or reject a job offer is static and depends only on current utility (or job price).

active talent markets. The employment rate in these markets is $\mu = 0$. Equating steady state flows of workers into and out of employment in active talent markets yields the employment rate:

$$\mu = \frac{\lambda/\delta}{1 + \lambda/\delta}. \tag{13}$$

In addition, equating steady state worker flows into and out of firms charging a job price of less than q in talent market θ implies a distribution of workers over job prices of:

$$G[q|\theta;\mathcal{P}] = \frac{1 + \lambda/\delta}{1 + (\lambda/\delta)F[q|\theta;\mathcal{P}]} F[q|\theta;\mathcal{P}]. \tag{14}$$

Finally, equating flows at firms charging a job price q in market θ implies a steady state employment at such firms of:

$$N(q,\theta;\mathcal{P}) = \frac{\lambda/\delta}{(1 + (\lambda/\delta)F[q|\theta;\mathcal{P}])^2} \frac{k(\theta)}{m(\theta)}, \tag{15}$$

where $m(\theta)$ is the density of firms in active market θ. It is convenient to express equilibrium job prices q as a function of worker talent θ and quantile $i \in [0,1]$ in $G[q|\theta;\mathcal{P}]$. That is to formulate job prices as $q(i,\theta;\mathcal{P})$, where this function is defined implicitly by $i = G[q(i,\theta;\mathcal{P}),\theta;\mathcal{P}]$. Combining (12), (14) and (15) yields the explicit equilibrium *job price function* of this form:

$$q(i,\theta;\mathcal{P}) = \left(\frac{1}{1 + (\lambda/\delta)(1-i)}\right)^2 \overline{q}(\theta;\mathcal{P}). \tag{16}$$

We interpret the quantile argument i as indexing the extractiveness of the firm with which the worker is matched. Thus, the job price function (16) relates job prices to the extent of frictions λ/δ, the extractiveness index i, and the maximal job price $\overline{q}(\theta;\mathcal{P})$. Taxes and benefits perturb the entire θ talent market job price function via the maximal job price $\overline{q}(\theta;\mathcal{P})$. In particular, from (11), perturbations to tax rates are transmitted to the maximal job price in talent market θ according to:

$$\frac{\partial \overline{q}(\theta;\mathcal{P})}{\partial \tau} = -\frac{\underline{x}(\theta;\mathcal{P})}{1-\tau} \quad \text{and} \quad \frac{\partial \overline{q}(\theta;\mathcal{P})}{\partial \tau_0} = -\frac{1}{1-\tau}, \tag{17}$$

where $\underline{x}(\theta;\mathcal{P})$ is the income earned by a θ talent accepting job price $\bar{q}(\theta;\mathcal{P})$ given $\mathcal{P}$ and, hence, is the lowest income earned by a θ talent. The direct impact of an increase in the income tax rate τ is to reduce a worker's utility. However, the outside option of unemployment sets a floor on utility from work. To continue to attract workers a firm charging $\overline{q}(\theta;\mathcal{P})$ must reduce its job price to fully compensate the worker for the additional tax $\underline{x}(\theta;\mathcal{P})$ that it pays. Consequently, $\frac{\partial\overline{q}(\theta;\mathcal{P})}{\partial\tau} = -\frac{\underline{x}(\theta;\mathcal{P})}{1-\tau}$. A similar logic underpins the expression $\frac{\partial\overline{q}(\theta;\mathcal{P})}{\partial\tau_0}$. Tax induced perturbations to maximal job prices propagate through the job price ladder within a talent market. For the case of τ :

$$\frac{\partial q(i,\theta;\mathcal{P})}{\partial\tau} = \left(\frac{1}{1+(\lambda/\delta)(1-i)}\right)^2 \frac{\partial\overline{q}(\theta;\mathcal{P})}{\partial\tau}. \tag{18}$$

Intuitively, firms cutting their job price from $\overline{q}(\theta;\mathcal{P})$ in response to a tax rate increment reduce recruitment and retention at other firms previously charging job prices slightly below $\overline{q}(\theta;\mathcal{P})$. These other firms respond by lowering their job prices, which elicits a round of job price cuts at firms charging job prices slightly below theirs. In this way the impact of a tax rise is transmitted through the entire job price distribution according to (18).

3.3 Implications for optimal tax design

In the frictionless setting, the policymaker could directly control the after-tax earnings schedule of all workers and it was possible to formulate the policymaker's problem in terms of after-tax lump sum income (possibly conditioned on employment status) and an after-tax effective labor price. In the current frictional model, a worker's after-tax lump sum income conditioned on employment, but before exertion of effective labor is: $b - \tau_0 - (1-\tau)q$. While the policymaker can influence the distribution of job prices through taxation, it now lacks the tax instruments to fully control them and, hence, after-tax lump sum income conditioned on employment. It is convenient now to write the policymaker's problem directly as a function of those variables that it can fully control, taxes and benefits $\mathcal{P} = (b, \tau_0, \tau)$.[9] We continue to assume 100% taxation of profit. The policymaker's problem is:

$$\begin{aligned} \max_{\mathcal{P}} \int_{\underline{\theta}}^{\overline{\theta}} \Gamma(\theta) U(b,0,\theta) k(\theta) d\theta \\ + \mu \int_{\tilde{\theta}(\mathcal{P})}^{\overline{\theta}} \int_0^1 \Gamma(\theta) \left\{ V(q(i,\theta;\mathcal{P}),\theta;\mathcal{P}) - U(b,0,\theta) \right\} di \; k(\theta) d\theta, \end{aligned} \tag{19}$$

[9] Worker effective labor is correspondingly redefined with $z(i,\theta;\mathcal{P})$ standing for $z(\theta; b-\tau_0-(1-\tau)q(i,\theta;T), 1-\tau)$.

subject to the resource constraint:

$$-b+\mu\int_{\tilde{\theta}(\mathcal{P})}^{\overline{\theta}}\left\{\tau\int_0^1\{z(i,\theta;\mathcal{P})-q(i,\theta;\mathcal{P})\}di+\tau_0\right\}k(\theta)d\theta\geq G.$$

Throughout the remainder of the paper we assume that Γ is non-increasing and that the policymaker weakly favors low θ types. Deriving the first order condition for τ, reformatting in terms of incomes x and suppressing the policy argument $\mathcal{P}$ in notation gives:[10]

$$-\int_{\mathbb{R}_+}E\Big[\left\{\frac{M}{\Lambda}+\frac{\tau}{1-\tau}\eta-1\right\}\Big\{x+\underbrace{(1-\tau)\frac{\partial q}{\partial\tau}}_{\substack{\text{Profit squeeze}\\ \text{redistributive term}}}\Big\}\Big|x\Big]h(x)dx$$
$$=\frac{\tau}{1-\tau}\int_{\mathbb{R}_+}xE[\mathcal{E}^c_{x,1-\tau}|x]h(x)dx+\frac{\tau_0+\tau\tilde{x}}{1-\tau}\frac{\tilde{\theta}k(\tilde{\theta})}{1-K(\tilde{\theta})}\mathcal{E}_{\tilde{\theta},1-\tau}. \tag{20}$$

The key difference between the frictionless optimal tax equation (8) and (20) is the appearance of the "profit squeeze redistributive term" involving $(1-\tau)\frac{\partial q}{\partial\tau}$ on the left hand side of the latter. As discussed previously, income tax rate increments induce general equilibrium adjustments in job prices. The new term in (20) describes the societal benefit of such adjustment. Since $\frac{\partial q}{\partial\tau}<0$, an increment in τ depresses tax revenues by squeezing value added from the profit tax base (where it is taxed at rate 1) to the labor earnings base (where it is taxed at rate $\tau<1$). This squeeze is distributed amongst workers according to $\frac{\partial q}{\partial\tau}$. In the frictionless model, the profit squeeze operated via a rise in the pre-tax wage. It implied redistribution proportional to income and, by raising the return to effective labor, a mitigating stimulus to income tax revenues. In contrast in the current frictional model, the profit squeeze does not affect the return to effective labor. Its redistributive consequences are not proportional to income and there is no mitigating stimulus to effective labor and income tax revenues. For these reasons, general equilibrium profit squeeze effects do not factor out of (20).

Consider first the case of a Rawlsian planner who seeks to maximize income tax revenues (on behalf of a $\underline{\theta}$ type that does not work). For this policymaker (20)

[10] See Appendix A for detailed derivation.

reduces to:

$$\int_{\mathbb{R}_+} E\left[\left\{1-\frac{\tau}{1-\tau}\eta\right\}\middle| x\right] xh(x)dx - \frac{\tau}{1-\tau}\int_{\mathbb{R}_+} xE[\mathcal{E}^c_{x,1-\tau}|x]h(x)dx + \frac{\tau_0+\tau\tilde{x}}{1-\tau}\frac{\tilde{\theta}k(\tilde{\theta})}{1-K(\tilde{\theta})}\mathcal{E}_{\tilde{\theta},1-\tau}$$
$$+\int_{\mathbb{R}_+} E\left[\left\{1-\frac{\tau}{1-\tau}\eta\right\}(1-\tau)\frac{\partial q}{\partial \tau}\middle| x\right] h(x)dx = 0. \tag{21}$$

That is, the policymaker selects a tax rate that places it at the top of the Laffer curve. Equation (21) augments the optimal tax equation of a Rawlsian policymaker in a frictionless world with the term $\int_{\mathbb{R}_+} E\left[\left\{1-\frac{\tau}{1-\tau}\eta\right\}(1-\tau)\frac{\partial q}{\partial \tau}\middle| x\right] h(x)dx$ which gives the loss of revenue from the profit squeeze. Since this term is negative it introduces a force for a lower optimal income tax rate other things equal.[11]

In non-Rawlsian settings, the distributional implications of the profit squeeze for employed workers must be incorporated into optimal tax evaluations. These are not uniform across the income distribution and may work to enhance or offset the policymaker's distributional goals. In (20) they are described by:

$$\int E\left[\left\{\frac{M}{\Lambda}+\frac{\tau}{1-\tau}\eta-1\right\}(1-\tau)\frac{\partial q}{\partial \tau}\middle| x\right] h(x)dx$$
$$= E\left[\frac{M}{\Lambda}+\frac{\tau}{1-\tau}\eta-1\right] E\left[(1-\tau)\frac{\partial q}{\partial \tau}\right] + \mathrm{Cov}\left(\frac{M}{\Lambda}+\frac{\tau}{1-\tau}\eta, (1-\tau)\frac{\partial q}{\partial \tau}\right), \tag{22}$$

where the right hand side decomposes the redistributive implications of the profit squeeze into two components: the social value of transferring the sum $E\left[(1-\tau)\frac{\partial q}{\partial \tau}\right]$ from the government's budget to each employed worker and the social value of the profit squeeze implied dispersal of these dollars amongst the employed.

The no income effects case To simplify analysis and purge income effects consider worker preferences of the form: $U(c,z,\theta) = W(c - f(z,\theta))$, with W increasing, concave and differentiable and f convex in z, twice differentiable and with $-\frac{\partial^2 f(z,\theta)}{\partial z \partial \theta} > 0$ and $f(0,\theta) = 0$. Consequently, individual choice and, hence, the distribution of job prices is not impacted by income effects. The distributional

[11] Across environments with and without frictions values of the terms in the first line of (21) are modified. Thus, (21) does not imply that taxes are necessarily lower in frictional economies. Rather it highlights the need to deduct losses in revenue stemming from the profit squeeze in evaluations of the marginal benefits of income taxation in frictional settings.

consequences of the profit squeeze (22) reduce to:

$$-(1-P)\left\{E\left[\frac{M}{\Lambda}\middle|u\right]-E\left[\frac{M}{\Lambda}\middle|e\right]\right\}E\left[(1-\tau)\frac{\partial q}{\partial\tau}\right]+\text{Cov}\left(\frac{M}{\Lambda},(1-\tau)\frac{\partial q}{\partial\tau}\right),\qquad(23)$$

where $P=K(\tilde{\theta})+\mu(1-K(\tilde{\theta}))$ denotes employment and $E[\cdot|u]$ and $E[\cdot|e]$ denote expectations conditioned on unemployment and employment respectively. The first component of (23) describes the marginal benefit of profit-squeeze induced redistribution from the unemployed to the employed and the second the marginal benefit of such redistribution amongst the employed. Under our assumptions on the social criterion, the first redistribution term is negative: The squeezing of profits transfers resources from the policymaker's budget, where it is partly spent on basic income to the unemployed, towards the better off employed. This is a force for lower taxes. Redistribution amongst the employed, the second term in (23), is more complex. It combines two competing forces that render the sign of this term ambiguous. These opposing forces are revealed by the job price sensitivity expressions (17) and (18). On the one hand, tax rate increments deliver larger job price reductions to lower earners at the bottom of job ladders. These workers pay the largest job prices, face the most extractive firms, and are fully compensated through job price reduction for any tax increment. On the other hand, the size of the job price reduction is increasing in $\underline{x}(\theta;\mathcal{P})$ and, hence, is greater for more talented workers. Thus, *within* talent markets, redistribution amongst the employed via tax-induced job price adjustment favors lower earners, while *across* talent markets it favors higher earners. Our assumptions on the social criterion ensure that the first effect contributes positively to $\text{Cov}\left(\frac{M}{\Lambda},(1-\tau)\frac{\partial q}{\partial\tau}\right)$ and the social value of the profit squeeze, while the second contributes negatively.

In summary, the frictional profit squeeze redistributes from the unemployed to the employed, from lower earning talents to higher and from workers in higher positions on job ladders to those in lower.

Comparing the frictional to the frictionless economies The bBM model introduces an additional profit squeeze term into a frictionless and otherwise standard optimal tax equation. As described above this term modifies the redistributive value of a tax increment to a policymaker at a given income distribution. For a given talent distribution frictions also alter the endogenous income and marginal social welfare weight distributions. In particular, while introducing additional income dispersion *within* talent markets, by preventing workers from securing their full marginal product, they tend to depress income dispersion *across* talent markets. The first of these effects lies behind the conventional view that frictions provide a rationale for greater redistribution. However, this first effect maybe offset by the second and a counter redistributive profit squeeze. Evaluation of the overall impact of frictions for tax design requires assessment both of the newly introduced profit squeeze term and the shift in endogenous distributions implied by frictions at a given talent distribution.

3.4 Frictional models and profit squeeze

The bBM model supplies a particular structural model of the profit squeeze and its redistributive impact. However, tax-induced profit squeezes that do not operate via changes to the effective price of labor and do not cancel from optimal tax equations as in the frictionless case are a broader feature of frictional models. In Online Appendix E, we elaborate a model in which surplus division is via Nash Bargaining and firms are heterogeneous with respect to bargaining power. In this model, the distribution over firm/worker bargaining power is treated as a parameter and may be picked to match the pattern of surplus division and profit squeeze that emerges in bBM.

4 Affine tax design in frictional economies with endogenous matching

The preceding section treated worker/firm matching as exogenous and all job prices as rent accruing to a fixed population of firms. Since the latter was taxed at 100%, job prices were valued by the planner using the marginal social value of public funds and the cost of the marginal fiscal spillover from the profit to the labor income base valued accordingly. In frictional models with endogenous matching, the impact of income tax policy on job prices affects incentives to create vacancies and, hence, employment. At least some part of a firm's job price is used to finance up front vacancy costs and this part is not a rent that can be expropriated to finance public spending. The social value of job prices is now linked to the social value of a job. Consequently, an income tax-induced profit squeeze generates a tradeoff between the societal benefits of job creation and the extraction of job prices from differently earning workers. The former benefits are related to externalities in the hiring and matching process. We develop these ideas in the context of a Burdett-Mortensen model with endogenous matching. In addition to externalities familiar from the off-the-job search literature, this model introduces an additional poaching externality amongst firms. In the absence of Pigouvian taxes that directly target them, these externalities influence the value of marginal job creation and, hence, the trade offs facing the income tax designing policymaker. With an eye on matching the data, we further generalize the Burdett-Mortensen model to allow for "Godfather shocks" that compel movement down the job ladder.

4.1 Worker and firm decision making

The worker problem is modified to incorporate endogenous contact rates. Let $\lambda(\theta;\mathcal{P})$ denote the equilibrium job finding rate for workers in talent market θ given policy $\mathcal{P} = (b, \tau_0, \tau)$. As before, conditional on meeting a firm, a worker draws a job price from $F[\cdot|\theta;\mathcal{P}]$. In addition, the worker draws a Godfather shock with probability s. If the Godfather shock is not drawn, then, as before, a worker with current job price q who draws a new job price q', accepts the new job price if $q' < q$. If the Godfather shock is drawn, then the worker selects between unemployment and the new job price. It will select the latter if it is below $\overline{q}(\theta;\mathcal{P})$, the highest job price that makes the worker indifferent between unemployment and work. The Godfather shock is a simple way of reconciling the model with high to low income job to job transitions.

The economy is populated by a unit mass of firms whose problems are extended to accommodate vacancy creation. Let $\kappa(v,\theta) = \bar{\kappa}(\theta)v$, $\bar{\kappa}(\theta) > 0$, denote the cost of creating v vacancies in a talent market.[12] Let $N(q,\theta;\mathcal{P})$ be the steady state number of workers per vacancy at a firm charging job price q in talent market θ given policy $\mathcal{P}$. Firms select a talent market, vacancy level, and job price to maximize:

$$N(q,\theta;\mathcal{P})qv - \kappa(v,\theta). \tag{24}$$

The term $N(q,\theta;\mathcal{P})q$ gives the firm's ex post income per vacancy.

[12] This specializes $\kappa(v,\theta) = \bar{\kappa}(\theta)\frac{v^{1+\rho}}{1+\rho}$ to the case $\rho = 0$. Our earlier model corresponds to the limiting case in which $\rho = \infty$ and $\kappa(\theta, v) = 0$ if $v \in [0,1]$ and $\kappa(\theta, v) = \infty$ if $v \in (1,\infty)$. In this earlier case, firms can costlessly post up to a single vacancy and cannot further scale the arrival rate of workers to the firm. The case $\rho \in (0,\infty)$ is intermediate. In that case job prices are a mixture of rents and vacancy cost payments.

4.2 Equilibria and tax incidence

No firm posts vacancies in an inactive talent market defined as before. Talent markets $(\underline{\theta}(p), \overline{\theta})$, where $\underline{\theta}(\mathcal{P}) = \max(\underline{\theta}, \tilde{\theta}(\mathcal{P}))$, are active and in these firms create vacancies and post job prices in $(0, \overline{q}(\theta;\mathcal{P})]$. Equating inflows into unemployment from job destruction and outflows from matching implies a steady state employment rate in active talent market θ of:

$$\mu(\theta;\mathcal{P}) = \frac{\lambda(\theta;\mathcal{P})/\delta}{1+\lambda(\theta;\mathcal{P})/\delta}, \tag{25}$$

where the notation emphasizes the dependence of the employment μ rate on talent market and policy. To obtain the equilibrium λ and μ functions, we first derive the steady state number of workers per vacancy at a firm posting job price q in talent market θ by equating steady state flows:

$$N(q,\theta;\mathcal{P}) = \frac{\varphi(\theta;\mathcal{P})}{(1+(\hat{\lambda}(\theta;\mathcal{P})/\hat{\delta}(\theta;\mathcal{P}))F[q|\theta;\mathcal{P}])^2}, \tag{26}$$

where φ is the arrival rate of workers at a vacancy and the job destruction and finding rates are modified to accommodate Godfather shocks: $\hat{\delta}(\theta;\mathcal{P}) = \delta + s\lambda(\theta;\mathcal{P})$ and $\hat{\lambda}(\theta;\mathcal{P}) = \lambda(\theta;\mathcal{P})(1-s)$. Inserting this expression for N into (24), solving the inner optimization over q and proceeding as in the earlier section delivers the job price function:

$$q(i,\theta;\mathcal{P}) = \left(\frac{1}{1+(\hat{\lambda}(\theta;\mathcal{P})/\hat{\delta}(\theta;\mathcal{P}))(1-i)}\right)^2 \overline{q}(\theta;\mathcal{P}), \tag{27}$$

with $i \in [0,1]$ a measure of firm extractiveness. This function is modified from earlier sections by the dependence of $\hat{\lambda}(\theta;\mathcal{P})/\hat{\delta}(\theta;\mathcal{P})$ on θ and policy. The first order condition for vacancies $v(\theta;\mathcal{P})$ of a firm in talent market θ is:

$$\bar{\kappa}(\theta) = \frac{\varphi(\theta;\mathcal{P})/\hat{\delta}(\theta;\mathcal{P})}{\left(1+\hat{\lambda}(\theta;\mathcal{P})/\hat{\delta}(\theta;\mathcal{P})\right)^2}\overline{q}(\theta;\mathcal{P}). \tag{28}$$

To close the model a standard matching technology is assumed: $m(v,k(\theta)) = av^{\alpha}k(\theta)^{1-\alpha}$, with $\alpha \in [0,1)$. Solving for the implied equilibrium contact rates for workers and firms in terms of vacancies and substituting into (28) generates the following implicit expression for v:

$$(\delta + D(\theta)v(\theta;\mathcal{P})^{\alpha})^2 v(\theta;\mathcal{P})^{1-\alpha} = E(\theta)\overline{q}(\theta;\mathcal{P}), \tag{29}$$

for constants $D(\theta)$ and $E(\theta)$. Thus, policy impacts vacancies in a given talent market via its impact on the maximal job price. Tax rate increments squeeze this price, and, hence, all job prices in the market and via this squeeze reduce the returns to vacancy posting. This, in turn, reduces vacancies, worker contact rates, and impacts employment via (25). In particular, (25) implies that for each θ market the elasticity of μ with respect to $1-\tau$, $\mathcal{E}_{\mu,1-\tau}$, is :

$$\mathcal{E}_{\mu,1-\tau} = \alpha(1-\mu)\cdot\mathcal{E}_{v,\overline{q}}\cdot\mathcal{E}_{\overline{q},1-\tau},$$

with $\mathcal{E}_{v,\overline{q}}$, the elasticity of v with respect to $\overline{q}$ supplied by (29) and $\mathcal{E}_{\overline{q},1-\tau}$, the elasticity of $\overline{q}$ with respect to $1-\tau$, supplied by (11). In addition, the sensitivity of the job price function to tax perturbations is modified from (18) to:

$$\frac{\partial q(\theta)}{\partial(1-\tau)} = \left\{1 - 2\left(\frac{(\hat{\lambda}(\theta)/\hat{\delta}(\theta))(1-i)}{1+(\hat{\lambda}(\theta)/\hat{\delta}(\theta))(1-i)}\right)\mathcal{E}_{\hat{\lambda}/\hat{\delta},\overline{q}}(\theta)\right\}\frac{q(\theta)}{\overline{q}(\theta)}\frac{\partial\overline{q}(\theta)}{\partial(1-\tau)}, \tag{30}$$

with $\mathcal{E}_{\hat{\lambda}/\hat{\delta},\overline{q}}(\theta)$ the elasticity of $\hat{\lambda}/\hat{\delta}$ with respect to $\overline{q}$ in market θ. This sensitivity incorporates two channels via which tax-induced changes in $\overline{q}(\theta)$ transmit to the entire intra-talent market job price function. First there is the price cutting channel encountered in the previous section without endogenous matching and captured by the term $1\cdot\frac{q(\theta)}{\overline{q}(\theta)}\frac{\partial\overline{q}(\theta)}{\partial(1-\tau)}$ in (30). As before, an increase in taxes compels firms previously charging $\overline{q}(\theta)$ to lower their job prices to continue to attract workers. This reduces recruitment and retention at other firms previously charging job prices just below $\overline{q}(\theta)$. These firms respond by lowering their job prices, which elicits a round of price cutting at firms charging job prices lower than theirs. In this way the impact of the tax rise travels through the entire intra-talent job price distribution. A second competition dampening channel captured by the term $-2\left(\frac{(\hat{\lambda}(\theta)/\hat{\delta}(\theta))(1-i)}{1+(\hat{\lambda}(\theta)/\hat{\delta}(\theta))(1-i)}\right)\mathcal{E}_{\hat{\lambda}/\hat{\delta},\overline{q}}(\theta)\frac{q(\theta)}{\overline{q}(\theta)}\frac{\partial\overline{q}(\theta)}{\partial(1-\tau)}$ is now also operative. Downward pressure on job prices deters vacancy creation. This in turn improves worker retention and mitigates the incentive for firms to compete through lower job prices. Evaluation of $\mathcal{E}_{\hat{\lambda}/\hat{\delta},\overline{q}}(\theta)$ and of the competition dampening channel term reveals that it is dominated by the price cutting channel term.

4.2.1 The policymaker's problem

The policymaker's problem in this setting is:

$$\max_{\mathcal{P}} \int_{\underline{\theta}}^{\overline{\theta}} \Gamma(\theta)U(b,0,\theta)k(\theta)d\theta \\ + \int_{\tilde{\theta}(\mathcal{P})}^{\overline{\theta}} \mu(\theta;\mathcal{P}) \int_0^1 \Gamma(\theta)\left\{V(q(i,\theta;\mathcal{P}),\theta;\mathcal{P}) - U(b,0,\theta)\right\} di\, k(\theta)d\theta, \tag{31}$$

subject to the resource constraint:

$$-b + \int_{\tilde{\theta}(\mathcal{P})}^{\overline{\theta}} \mu(\theta;\mathcal{P}) \left\{\tau \int_0^1 \{z(i,\theta;\mathcal{P}) - q(i,\theta;\mathcal{P})\}di + \tau_0\right\} k(\theta)d\theta \geq G,$$

where now μ and q depend on θ and $\mathcal{P}$ according to (25) and (27) respectively.

Deriving the first order condition for τ and formatting in terms of incomes x gives:[13]

$$\begin{aligned} -\int_{\mathbb{R}_+} E\left[\left\{\frac{M}{\Lambda} + \frac{\tau}{1-\tau}\eta - 1\right\}\left\{x + (1-\tau)\frac{\partial q}{\partial \tau}\right\}\middle| x\right] h(x)dx \\ = \frac{\tau}{1-\tau}\int_{\mathbb{R}_+} E[\mathcal{E}^c_{x,1-\tau}|x]h(x)dx \\ + \frac{1}{1-\tau}\int_{\mathbb{R}_+} E\left[\left(\left\{\frac{\Delta U}{\Lambda} + \tau x + \tau_0\right\}\mathcal{E}_{\mu,v} + \left\{\mathcal{E}_{\mu,v} - \frac{vN(q)}{\mu}\right\}q\right)\mathcal{E}_{v,1-\tau}\middle| x\right] h(x)dx, \end{aligned} \tag{32}$$

where $\frac{\Delta U(i,\theta)}{\Lambda} := \frac{\Gamma(\theta)}{\Lambda}\{V(q(i,\theta;\mathcal{P}),\theta;\mathcal{P}) - U(b,0,\theta)\}$ and the terms $\mathcal{E}_{\mu,v} = \alpha(1-\mu)$, $\mathcal{E}_{v,1-\tau} = \mathcal{E}_{v,\overline{q}}\mathcal{E}_{\overline{q},1-\tau}$, with the latter supplied by expressions in the previous section.

Equation (32) is modified relative to the earlier expression (20) by the addition of the term in the last line. This gives the marginal cost of tax-induced vacancy suppression.[14] If firms post the socially efficient level of vacancies given the tax system, then this term is zero. In general, however, this is not the case. While externalities associated with vacancy posting are familiar from other search environments, here they are modified by the presence of taxes and on-the-job search.

[13] We omit the first order condition for τ_0, which features a similar organization of terms. For derivation of (32) see Appendix A.

[14] Equation (32) omits the term $\frac{\tau_0+\tau\tilde{x}}{1-\tau}\frac{\tilde{\theta}k(\tilde{\theta})}{1-K(\tilde{\theta})}\mathcal{E}_{\tilde{\theta},1-\tau}$ found in (20). This term describes the tax revenues lost from deactivation of the marginal talent market in the bBM model. Although higher tax rates lead to a similar deactivation in the current model, vacancy creation converges to zero as $\theta \downarrow \tilde{\theta}$. Thus, no vacancies are created in the marginal talent market and no revenues lost when it is closed.

The last line in (32) organizes the marginal cost of vacancy suppression into two parts:

1. The first, $\int_{\mathbb{R}_+} E\left[\left(\left\{\frac{\Delta U}{\Lambda}+\tau x+\tau_0\right\}\mathcal{E}_{\mu,v}\right)\mathcal{E}_{v,1-\tau}\middle| x\right]h(x)dx$, stems from the failure of firms to internalize the worker utility and tax benefits and, hence, full social value of a job. This failure suggests equilibrium under-posting of vacancies and implies a positive marginal cost from further vacancy suppression.

2. The second part, $\int_{\mathbb{R}_+} E\left[\left(\left\{\mathcal{E}_{\mu,v}-\frac{vN(q)}{\mu}\right\}q\right)\mathcal{E}_{v,1-\tau}\middle| x\right]h(x)dx$, stems from the failure of firms to fully internalize the effect of their vacancy posting on employment at other firms. This, in turn, is part a standard congestion externality on the new jobs margin and part a firm poaching externality: many hires are firm-to-firm and do not create additional jobs. Both externalities imply equilibrium over-posting of vacancies and a corresponding marginal benefit (negative marginal cost) from vacancy suppression. Formally, for any θ and q,

$$\left\{\mathcal{E}_{\mu,v}-\frac{vN(q)}{\mu}\right\}q\leq\frac{k}{\mu}\frac{\delta\lambda(v)}{(\delta+\lambda(v))^2}\left\{\frac{v\lambda_v(v)}{\lambda(v)}-1\right\}\bar{q}<0,$$

   where the first inequality follows from the definitions of $\mathcal{E}_{\mu,v}$ and $N(q)$ and the fact that all firms in a given talent market earn the same expected revenues per job and the second from the concavity of the matching function with respect to vacancies. The second component term in the last line of (32) is, consequently, negative.

Whether vacancies are above or below their socially optimal level given the tax system and whether the combination of terms in the last line of (32) is positive or negative reduces to the balance of the forces just described.

5 Quantitative analysis of optimal affine taxes

This section explores quantitative implications of our theory for optimal affine income tax determination. We first discuss how to connect the Burdett Mortensen model with endogenous matching to the data.

5.1 Calibration of the frictional model

Structural parameters The primitives of the model are parameters describing the utility function U, the exogenous job separation rate δ, the Godfather shock probability s, the matching function parameter α, the vacancy cost function κ, the talent distribution K, and exogenous revenue requirement G. We restrict the parametric form of U to be:

$$U(c,z,\theta) = \frac{1}{1-\sigma}\left(c - \frac{1}{1+\gamma}\left(\frac{z}{\theta}\right)^{1+\gamma}\right)^{1-\sigma} \tag{33}$$

and take parametric values for σ, γ and also α, δ and s from the literature. The vacancy cost function $\kappa(\theta, v) = \overline{\kappa}(\theta)v$ is linear in vacancies, with talent-contingent parameter $\overline{\kappa}(\theta)$. The talent distribution K is assumed to have a density k. We derive estimates for $\overline{\kappa}(\theta)$ and $k(\theta)$ by exploiting an analogy between the integral equations defining steady state equilibrium in our setting and similar integral equations found in discrete choice random coefficients settings.

Preferences In our baseline analysis, utility parameters are set to $\sigma = 2$ and $\gamma = 1$. We report sensitivity analysis around these values in Online Appendix H.

Job destruction rate and probability of Godfather shocks The model's time period is set to one month (but later convert tax functions to an annualized form). We select δ to be in line with the empirical job destruction 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. We set $\delta = 0.03$. We take the Godfather shock probability to be $s = 0.33$, which is within the range of estimates provided by Jolivet et al (2006) in their analysis of wage and employment transition data.

Revenue requirement The exogenous revenue requirement G is set to equal 25% of equilibrium output.

Talent density k and vacancy cost parameter $\overline{\kappa}$ It remains to determine the talent density k and the vacancy cost parameter $\overline{\kappa}$. In frictionless settings, absent bunching, the latent talent density is readily recovered from the observed income density via worker first order conditions. In the current frictional setting this recovery is complicated by job ladders, which imply a non-unique relation between talents and incomes. We proceed by first relating the observed current earnings density of the employed and last earnings density of the unemployed to the latent talent density and a function giving equilibrium job finding/destruction rate ratios by talent. We then describe how to "invert" these relations and, hence, obtain the unknown talent density and job finding/destruction rate ratios by talent.

To streamline exposition, we suppress notational dependence of variables on policy $\mathcal{P}$ in this section. Let $\underline{\theta}(x)$ denote the least talented worker earning income x. This worker will be at the top of her talent market job ladder and will pay the smallest job price $\underline{q}(\underline{\theta}(x))$. Similarly, let $\overline{\theta}(x)$ denote the most talented worker earning x. This worker will be at the bottom of her talent market job ladder and will pay the highest job price $\overline{q}(\underline{\theta}(x))$. Let h denote the density of employed workers across incomes and l the density of unemployed workers across their last earned income. Proposition 1 relates the income and talent densities.

Proposition 1. *Let $\beta : (\tilde{\theta}, \infty) \to \mathbb{R}_+$, $\beta(\theta) := \lambda(\theta)/\delta$, give the job contact-destruction rate ratio. The income densities h and l satisfy:*

$$h(x) = \int_{\tilde{\theta}}^{\overline{\theta}} \Phi(x|\theta;\beta)\mu(\theta)\frac{k(\theta)}{\mathcal{N}}d\theta, \tag{34}$$

$$l(x) = \int_{\tilde{\theta}}^{\overline{\theta}} \Phi(x|\theta;\beta)(1-\mu(\theta))\frac{k(\theta)}{\mathcal{U}}d\theta, \tag{35}$$

where $\mathcal{N} = \int_{\tilde{\theta}}^{\overline{\theta}} \mu(\theta)k(\theta)d\theta$ and $\mathcal{U} = \int_{\tilde{\theta}}^{\overline{\theta}}(1-\mu(\theta))k(\theta)d\theta$ are, respectively, the fractions of employed and unemployed agents with talents in excess of $\tilde{\theta}$, $\mu(\theta) = \frac{\beta(\theta)}{1+\beta(\theta)}$ is the talent-specific employment rate, and the kernel $\Phi(x|\theta;\beta)$ linking talents to incomes satisfies:

$$\Phi(x|\theta;\beta) = \begin{cases} \frac{1}{2\Psi(\beta(\theta))}\sqrt{\frac{-\frac{\tau_0}{1-\tau}+\frac{\gamma}{1+\gamma}\theta^{\frac{1+\gamma}{\gamma}}(1-\tau)^{\frac{1}{\gamma}}}{\{(1-\tau)^{\frac{1}{\gamma}}\theta^{\frac{1+\gamma}{\gamma}}-x\}^3}} & \theta \in [\underline{\theta}(x;\beta), \overline{\theta}(x)] \\ 0 & \textit{otherwise,} \end{cases} \tag{36}$$

with $\Psi(\beta(\theta)) := (1-s)\beta(\theta)/(1+s\beta(\theta)) = \hat{\lambda}(\theta)/\hat{\delta}(\theta)$, $\underline{\theta}(x;\beta)$ *the solution to:*

$$\underline{\theta}(x) = \left(\frac{x - \frac{\tau_0}{1-\tau}\left(\frac{1}{1+\Psi(\beta(\underline{\theta}(x)))}\right)^2}{(1-\tau)^{\frac{1}{\gamma}}\left(1 - \frac{\gamma}{1+\gamma}\left(\frac{1}{1+\Psi(\beta(\underline{\theta}(x)))}\right)^2\right)} \right)^{\frac{\gamma}{1+\gamma}}, \text{ and } \overline{\theta}(x) = \left(\frac{x - \frac{\tau_0}{1-\tau}}{(1-\tau)^{\frac{1}{\gamma}}\frac{1}{1+\gamma}} \right)^{\frac{\gamma}{1+\gamma}}. \tag{37}$$

Proof. See Appendix B. □

The mapping (34) computes the density of currently employed workers earning a given income level x by aggregating the densities of employed workers earning x in each talent market. Division by $\mathcal{N}$ normalizes to ensure that h is a probability density. The mapping (35) computes the density of currently unemployed workers who earned a given income x in their last employment. Equation (36) relates the kernel Φ linking talents to incomes to the structural parameters δ and γ, observed policy $\mathcal{P}$ and the endogenous job finding rate $\lambda(\theta)$.

Expressions (34) and (35) imply:

$$h(x) = \int_{\tilde{\theta}}^{\overline{\theta}(x)} \Phi(x|\theta;\beta)\iota(\theta)d\theta \qquad \text{and} \qquad l(x) = \int_{\tilde{\theta}}^{\overline{\theta}(x)} \Phi(x|\theta;\beta)\nu(\theta)d\theta, \tag{38}$$

where $\iota(\theta) := \frac{\beta(\theta)}{1+\beta(\theta)}\frac{k(\theta)}{\mathcal{N}}$ gives the density of employed workers across talents and $\nu(\theta) := \frac{1}{1+\beta(\theta)}\frac{k(\theta)}{\mathcal{U}}$ the density of active unemployed workers across talents. Given β and $\mathcal{P}$ and, hence, Φ and $\underline{\theta}$, and empirical estimates of h and l, the expressions in (38) are Fredholm equations of the first kind in the unknown functions ι and ν. Well known procedures for inverting such equations exist. This observation and the definitions of ι and ν motivate an algorithm for recovering k and β sketched in Algorithm 1. The algorithm embeds a Fredholm equation inversion step into an iterative search for a β function that solves a fixed point problem. Once β is obtained (and, hence, given δ and s, λ, $\hat{\lambda}$ and $\hat{\delta}$ are found), an estimate of $\overline{\kappa}$ is calculated via:[15]

$$\bar{\kappa}(\theta) = \frac{a^{\frac{1}{\alpha}}\hat{\delta}(\theta;\mathcal{P})\lambda(\theta;\mathcal{P})^{\frac{\alpha-1}{\alpha}}}{\left(\hat{\lambda}(\theta;\mathcal{P}) + \hat{\delta}(\theta;\mathcal{P})\right)^2}\overline{q}(\theta;\mathcal{P}). \tag{39}$$

Further details of the numerical procedure used in the estimation are described in Appendix C.

[15] Equation (39) is obtained by substituting equilibrium expression $\varphi(\theta;\mathcal{P}) = a^{\frac{1}{\alpha}}\lambda(\theta;\mathcal{P})^{\frac{\alpha-1}{\alpha}}$ into (28).

Algorithm 1 Recovering β and k.

Initialize. Set values for program hyper-parameters: `Maxiter`, `Criterion`, `Tol` and update parameter o. Set iteration counter $n = 1$. Select β_1.

while $n <$ `Maxiter`, `Criterion`$<$`Tol` **do**

    Given β_n and empirical proxies for h, l and $\mathcal{P}$, solve the Fredholm equations to obtain ι_n and ν_n.

    Using empirical proxies for $\mathcal{N}$ and $\mathcal{U}$, set $\beta_n' = \frac{\mathcal{N}}{\mathcal{U}} \frac{\iota_n}{\nu_n}$.

    Update $\beta_{n+1} = o\beta_n' + (1-o)\beta_n$.

    Evaluate convergence criterion `Criterion`. Set $n = n + 1$.

end while

Calculate: $k = \iota_n \frac{1+\beta_n}{\beta_n} \mathcal{N}$.

Data Our procedure for estimating k and $\overline{\kappa}$ uses two empirical earnings distributions as inputs: one for the currently employed and one (a distribution of last earnings) for the currently unemployed. We construct these two earnings distributions using data from 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 2017 survey which provides information for the calendar year 2016 and use March supplement sample weights to produce our estimates. Detailed description of the data and of our sample selection is given in Online Appendix F.

Approximating status quo U.S. policy Expressions (36) and (37) and Algorithm 1 require empirical proxies for policy. We form an affine approximation to status quo US income tax policy, $T[x] = T_0 + \tau x$, by regressing total income taxes paid (state plus federal income tax liabilities net of tax credits) on labor income. Our estimated (annualized) tax function (with standard deviations in parentheses) is:

$$\hat{T}[x] = \underset{(27.576)}{-4230} + \underset{(0.000321)}{0.338}\; x.$$

The smallest active talent $\tilde{\theta}$ is then determined by: $(1-\hat{\tau}) - \frac{\underline{x}^{\gamma}}{\tilde{\theta}^{1+\gamma}} = 0$, where $\hat{\tau}$ is the estimated marginal income tax value, 0.338, $\underline{x}$ is the smallest income in our sample and γ equals its calibrated value. In addition, once $\tilde{\theta}$ is determined, an empirical proxy for the worker's non-work payoff $\hat{b}$ is pinned down by:

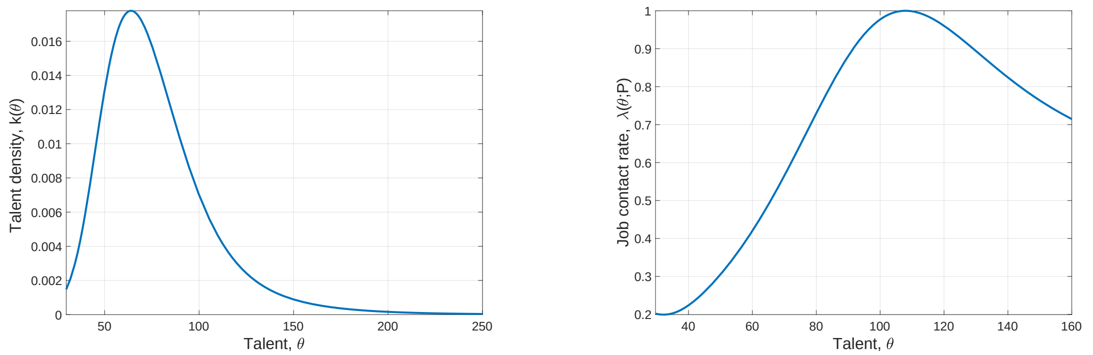
$$-\hat{T}_0 + (1-\hat{\tau})\underline{x} - \frac{1}{1+\gamma}\left(\frac{\underline{x}}{\tilde{\underline{\theta}}}\right)^{1+\gamma} = \hat{b}, \tag{40}$$

where $\hat{T}_0$ denotes the estimated value -4230. In preceding sections b was identified with the transfer made to non-employed workers and, hence, was both the value of inactivity to a worker and its direct cost 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.[16] Given the values for T, $\tilde{\theta}$ and $\underline{x}$, an empirical counterpart for b can be recovered from the data using (40). A value for b^U is not needed to calibrate the talent distribution. Given empirical values for the tax policy parameters and for b, empirical counterparts for the functions $\overline{q}$, $\tilde{q}$, $\underline{\theta}$ and $\overline{\theta}$ may be constructed using the relevant formulas from preceding sections.

Estimation results Figure 1 displays results from the estimation exercise. Figure 1(a) shows the estimated talent density, $k(\theta)$; Figure 1(b) the estimated equilibrium job rate $\lambda(\theta)$ by talent market at status quo U.S. tax policy. The last earnings density of unemployed workers is left-shifted relative to the current earnings density of employed workers. The model attributes this to lower job finding rates amongst low talents. This is reflected in Figure 1(b).

Frictionless benchmark Below we compare optimal policy in the calibrated frictional economy to that obtained from treating the data as generated by the frictionless model with extensive margin introduced in Section 2. The latter recovers optimal taxes from (8), retains utility parameters used in the frictional calibration, but recalibrates the talent distribution following the approach of Saez (2001). See Appendix D for details. We emphasize that the comparison of frictional to frictionless that we undertake is not a pure comparative static exercise in which frictions are purged. Rather we compare policy prescriptions under the assumption that the data was generated by, respectively, a frictional and a frictionless model.

[16] 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.



(a) Talent density (b) Job contact rates by talent market

Figure 1: Calibrated talent density and job contact rates.

5.2 Quantitative results

This section reports the computed optimal tax system for the calibrated frictional economy. It interprets this system through the lens of the optimal tax equation (32) and compares it to that implied by the calibrated frictionless model and (8). We assume a utilitarian objective.

5.2.1 Overall tax systems: frictional vs. frictionless

Table 1 reports optimal tax results for the calibrated frictional and frictionless economies. Comparison of the first and second columns reveals that the marginal income tax rate τ is moderately reduced (31.7% vs. 35.7%) and the annualized employment tax τ_0 raised ($6912 vs. $3360) at the frictional relative to the frictionless calibration. Overall the optimal frictional tax design im-

Table 1: Optimal Affine Tax Policy

	Frictional	Frictionless
τ	31.7%	35.7%
τ_0	6912	3360
b	9060	8580

τ_0, b: annual 2017 US $ amounts.

plies a flattening of the income tax system with less redistribution amongst the employed and slightly more redistribution to the unemployed (b is increased to

$9060 from $8580) relative to the frictionless design. Although the difference in the marginal income tax across the optimal frictional and frictionless tax designs appears moderate, redistribution towards the lowest earners is significantly reduced in the first case. Figure 2 shows that *average tax rates* for low earners are

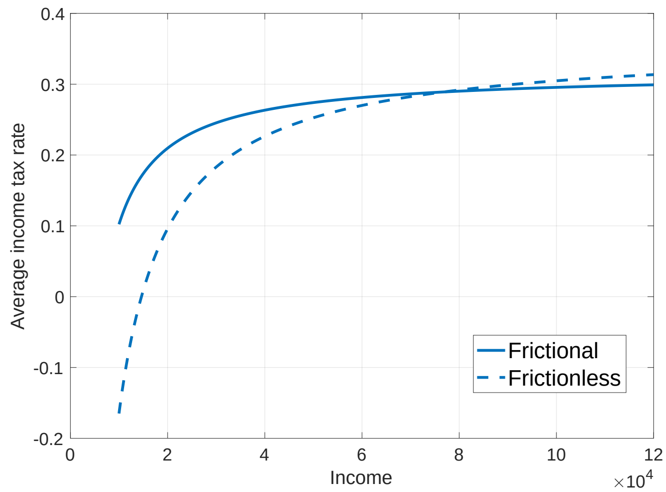


Figure 2: Average income tax rates ($\frac{\tau x+\tau_0-b}{x}$) implied by frictional and frictionless optimal tax systems.

substantially higher in the frictional case: a person earning $20,000 a year pays tax equal to 21% of their annual income under the optimal frictional scheme, but just 10% in the optimal frictionless case. The difference in relative optimal average income tax rates becomes even larger at lower incomes.

5.2.2 Marginal benefits and costs of adjusting τ

Differences in optimal tax systems across the calibrated frictional and frictionless economies are the product of various competing forces. To better understand the role of these forces in shaping optimal income tax rates, Figure 3 displays marginal benefit and cost terms from the frictional (32) and frictionless (8) optimal tax equations. These are scaled by $\partial\tau = 0.01$ and plotted as functions of τ holding τ_0, b and Λ fixed at their respective optimal values. Thus, the plots display components of the (normalized) derivative of the policymaker's Lagrangian as τ varies locally around the optimum under frictional and frictionless assumptions. The displayed impacts are expressed in terms of annual per capita 2017 US dollars accruing to the government budget. Colors indicate marginal component terms; line styles indicate environments: solid for frictional, dashed for frictionless. The horizontal axis ranges across τ values in a neighborhood of

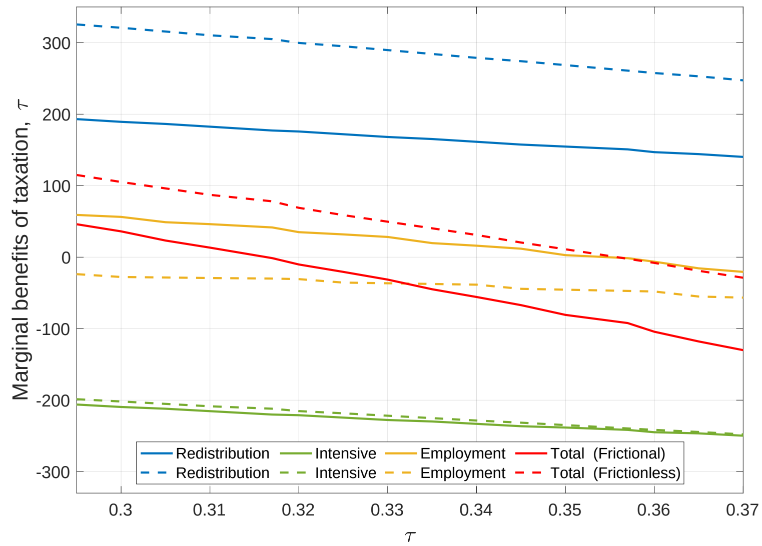


Figure 3: Marginal terms in frictional and frictionless optimal τ equations.
Note: Marginal benefits for the frictional economy shown as solid lines; Marginal benefits for the frictionless economy shown as dashed lines.

the optima from the two models.

The red solid and dashed lines show the *overall marginal benefit* of an increment in the income tax rate τ for the frictional and frictionless cases respectively. The solid line lies everywhere below the dashed (with a difference of between \$69 and \$101 annual per capita 2017 US dollars). Consistent with the results in Table 1, these lines cross zero at 31.7% (the optimal frictional marginal tax rate) and 35.7% (the optimal frictionless marginal tax rate) respectively. The green, orange and blue curves describe components of these overall marginal benefits.

1. Tax rates distort on the intensive margin. The associated marginal cost (or negative marginal benefit), given by $-\frac{\tau}{1-\tau}\int_0^\infty xE[\mathcal{E}^c_{x,1-\tau}|x]h(x)dx$, is shown by the green lines (solid and dashed) in the figure. These lines indicate that distortion on the intensive margin is the largest marginal cost (most negative benefit) of a tax increment. The lines are very similar across the two cases.

2. Marginal benefits and costs connected to employment are shown by orange lines (solid and dashed) in the figure. In the frictional economy, these are

related to the efficiency of vacancy creation. The orange solid line indicates a marginal benefit of suppressing vacancies through higher taxes that falls from $59 to -$21 annual per capita 2017 US dollars as the tax rate rises. At the optimal frictional tax rate, it is small and slightly positive, indicating that at this tax rate congestion and poaching externalities dominate, vacancy creation is excessive and there is a modest social benefit to suppressing it.[17] In the frictionless economy, employment effects operate via the (de)activation of marginal talent markets. They are always associated with marginal costs (negative marginal benefits). However, these costs are small as marginal talents are not very productive. Overall the frictional evaluation attributes a marginal benefit from raising taxes to the employment margin that moderately exceeds that from the frictionless one (holding fixed other dimensions of the tax code).

3. The main source of difference between the evaluations lies in the marginal redistributive benefit (MRB), which is positive, but significantly lower in the frictional case. This benefit is $-E\left[\left\{\frac{M}{\Lambda}-1\right\}\left\{x+(1-\tau)\frac{\partial q}{\partial \tau}\right\}\right]$ in the frictional optimal tax equation and $-E\left[\left\{\frac{M}{\Lambda}-1\right\}x\right]$ in the frictionless one and is shown by blue lines. It decreases from $193 to $140 annual per capita 2017 US Dollars over the tax range in the frictional case and from $326 to $247 in the frictionless case. The frictional MRB term incorporates the redistributive impact of the tax-induced profit squeeze: $E\left[\left\{\frac{M}{\Lambda}-1\right\}\left\{(1-\tau)\frac{\partial q}{\partial \tau}\right\}\right]$, this is not present in the frictionless case. It can itself be decomposed into a redistributive within talent market effect (higher taxes squeeze the maximal job price that is extracted by firms from those at the bottom of the job ladder, with the effect dissipating as the ladder is climbed) and a counter redistributive cross-talent market effect (job prices are squeezed more in higher talent markets, redistributing consumption from low to high talent markets). The latter effect predominates and overall the profit squeeze introduces a marginal redistributive cost that lowers the MRB in frictional models.[18] The value of the term $-E\left[\left\{\frac{M}{\Lambda}-1\right\}x\right]$, present in both MRB terms, is also modified by the presence of frictions. The covariance between incomes and marginal social welfare weights is suppressed in absolute value by frictions in part because more talented high earners pay higher job prices

[17] See Online Appendix H.1 for further decomposition in this direction.

[18] See Online Appendix H.1 for decomposition of the redistributive impact of the tax-induced profit squeeze.

necessitating that they exert higher efforts to achieve their incomes. Under our (non-separable) utility specification, this raises their marginal utilities of consumption relative to those of less talented lower earners so reducing (in absolute value) the covariance between marginal social welfare weights and incomes.

5.2.3 Marginal benefits and costs of adjusting τ_0

Table 1 indicates a frictional employment tax τ_0 that is double that in the frictionless case: \$6912 vs. \$3360 per annum. Figure 4 decomposes the marginal benefit and cost terms from the frictional optimal tax equation for τ_0 and its frictionless analogue. Again these plots hold other components of the tax code, in

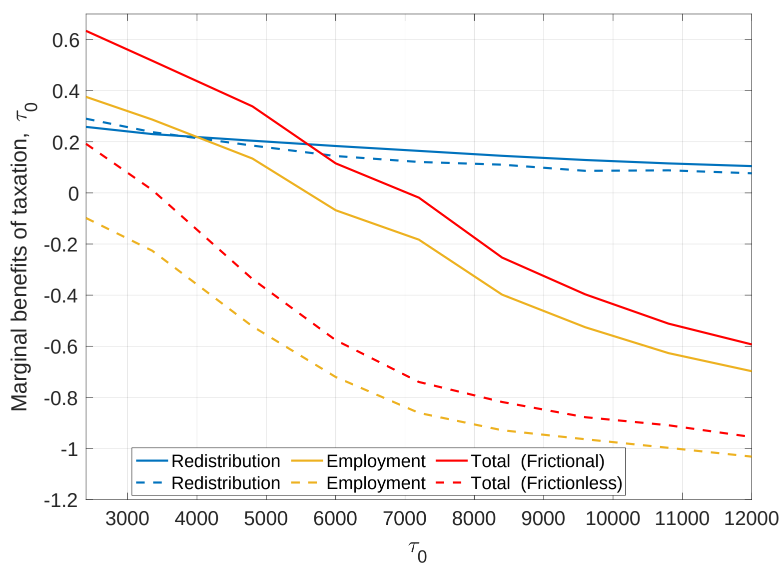


Figure 4: Marginal terms in frictional and frictionless optimal τ_0 equations.
Note: Marginal benefits for the frictional economy shown in solid lines; Marginal benefits for the frictionless economy shown in dashed lines.

this case τ and b, as well as Λ fixed at their optimal values. They give welfare effects in per capita 2017 US dollars accruing to the government budget of a dollar increment in the fixed employment tax.

The red curves give the overall marginal benefit of raising τ_0 in the two environments. The solid curve for the frictional economy lies everywhere above the dashed curve for the frictionless economy. As before different forces are at work. Across the two cases, there is a similar, moderate redistributive benefit from raising τ_0 (blue plots). Compared to the τ case and relative to welfare effects on the employment margin, redistributive benefits are no longer as salient. This is because while adjustments in τ_0 redistribute from the employed to the unemployed, they no longer mechanically redistribute amongst employed workers as is the case with adjustments to τ.[19] The largest welfare effects both across environments and τ_0 levels are now on the employment margin (orange plots). At lower τ_0 values, congestion and poaching externalities imply excessive vacancy posting in the frictional environment and there is social value in using larger τ_0 values to suppress this. For the frictionless case, higher τ_0 values closes marginal talent markets, which becomes increasing costly as τ_0 rises.

6 Conclusion

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. We analyze the implications of this for tax design in a model that blends standard public finance features with a frictional labor market. In doing so we highlight a novel frictional "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, deter vacancy creation and reduce profit tax revenues. On the other hand they raise worker incomes and income tax revenues. Such profit squeezing is socially desirable and a motive for higher income tax rates if vacancy creation is excessive and 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. Our quantitative analysis derives optimal tax designs for frictional and frictionless economies. It obtains a moderately lower marginal income tax rate and a higher employment tax for the frictional relative to the frictionless case.

We have explored the role of the profit channel in a salient, but particular fric-

[19] In the frictional case, adjustments in τ_0 generate some redistribution amongst the employed by modifying the distribution of job prices. Recall that when τ is increased, induced job price adjustment redistributes within and across talent markets. The former redistribution is towards those at the bottom of job ladders, the latter towards those in higher earning talent markets. Adjustments of τ_0 induce only the former-type of redistribution.

tional 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 the share of pre-tax surplus collected by firms and differentially distributes it across workers. Such differential distributional might occur because workers are heterogeneous with respect to mobility or replaceability, firms are heterogeneous with respect to their bargaining ability or, as here, firms trade off size against extractiveness.

References

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

C. Bontemps, J.-M. Robin and G. 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. *Journal of Economic Theory 126*(1), 165-193.

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

Diamond, P. (1975). A many person Ramsey tax rule. *Journal of Public Economics 4*(1), 335-342.

Diamond, P. and J. Mirrlees (1971). Optimal taxation and public production II: Tax rules. *American Economic Review 61*(3), 261-278.

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

Hornstein, A., P. Krusell, and G. Violante. (2011). Frictional wage dispersion in search models: A quantitative assessment. *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. *Review of Economic Studies 73*(3), 743-767.

Jolivet, G., F. Postel-Vinay, and J.-M. Robin. (2006). The empirical content of the job search model, labor mobility and wage distributions in Europe and the U.S. *Contributions to Economic Analysis 275*, 269-308.

E. Lehmann, 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.

J. Lise, C. Meghir and J.-M. Robin. (2016). Matching, sorting and wages. *Review of Economic Dynamics 19*(11), 63-87.

Mirrlees, J. (1971). An exploration in the Theory of Optimum Income Taxation. *Review of Economic Studies 38*(2), 175-208.

Munk, K. (1978). Optimal taxation and pure profit. *Scandinavian Journal of Economics 80*(1), 1-19.

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. (1978). Redistributive taxation in the Roy model. *Quarterly Journal of Economics 128*(2), 623-668.

Sachs, D., A. Tsyvinski and N. Werquin. (2020). Nonlinear tax incidence and optimal taxation in general equilibrium. *Econometrica 88*(2), 469-493.

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

Scheuer, F. and I. Werning (2017). The taxation of superstars. *Quarterly Journal of Economics 132*(1), 211-270.

Scheuer, F. and I. Werning (2016). Mirrlees meets Diamond-Mirrlees. NBER WP 22076.

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

Stiglitz, J. (1982). Self-selection and Pareto efficient taxation. *Journal of Public Economics 17*(2), 213-240.

Appendices

A Elaboration of optimal tax equations from text

A.1 Derivation of optimal tax equation (20)

First differentiate the objective (19) with respect to τ and use the definitions of M, ω and ξ:

$$\int_{\Theta}\int_{[0,1]}\{M(i,\theta)-\Lambda\}\left\{z(i,\theta)-q(i,\theta)+(1-\tau)\frac{\partial q}{\partial\tau}(i,\theta)\right\}dik(\theta)d\theta \tag{A.1}$$
$$+\Lambda\tau\int_{\Theta}\int_{[0,1]}\frac{\partial z(i,\theta)}{\partial\omega}dik(\theta)d\theta$$
$$+\Lambda\tau\int_{\Theta}\int_{[0,1]}\frac{\partial z(i,\theta)}{\partial\xi}\left\{-q(i,\theta)+(1-\tau)\frac{\partial q}{\partial\tau}(i,\theta)\right\}dik(\theta)d\theta+\frac{\tau_0+\tau\tilde{x}}{1-\tau}\frac{\tilde{\theta}k(\tilde{\theta})}{1-K(\tilde{\theta})}\mathcal{E}_{\tilde{\theta},1-\tau}=0.$$

Then apply Slutsky's equation:

$$\int_{\Theta}\int_{[0,1]}\{M(i,\theta)-\Lambda\}\left\{z(i,\theta)-q(i,\theta)+(1-\tau)\frac{\partial q}{\partial\tau}(i,\theta)\right\}dik(\theta)d\theta \tag{A.2}$$
$$+\Lambda\tau\int_{\Theta}\int_{[0,1]}\frac{\partial z^c(i,\theta)}{\partial\omega}dik(\theta)d\theta$$
$$+\Lambda\tau\int_{\Theta}\int_{[0,1]}\frac{\partial z(i,\theta)}{\partial\xi}\left\{z(i,\theta)-q(i,\theta)+(1-\tau)\frac{\partial q}{\partial\tau}(i,\theta)\right\}dik(\theta)d\theta+\frac{\tau_0+\tau\tilde{x}}{1-\tau}\frac{\tilde{\theta}k(\tilde{\theta})}{1-K(\tilde{\theta})}\mathcal{E}_{\tilde{\theta},1-\tau}=0.$$

Finally, change variables to $x=z-q$, re-express in terms of elasticities and rearrange:

$$-\int_{\mathbb{R}_+}E\left[\left\{\frac{M}{\Lambda}+\frac{\tau}{1-\tau}\eta-1\right\}\left\{x+(1-\tau)\frac{\partial q}{\partial\tau}\right\}\middle|\,x\right]h(x)dx$$
$$=\frac{\tau}{1-\tau}\int_{\mathbb{R}_+}xE[\mathcal{E}^c|x]h(x)dx+\frac{\tau_0+\tau\tilde{x}}{1-\tau}\frac{\tilde{\theta}k(\tilde{\theta})}{1-K(\tilde{\theta})}\mathcal{E}_{\tilde{\theta},1-\tau}. \tag{A.3}$$

□

A.2 Derivation of the first order condition for $-\tau_0$ in the frictional bBM economy

Differentiating (19) this time with respect to $-\tau_0$:

$$\int_\Theta \int_{[0,1]} \{M(i,\theta) - \Lambda\} \left\{1 + (1-\tau)\frac{\partial q}{\partial \tau_0}(i,\theta)\right\} dik(\theta)d\theta \tag{A.4}$$
$$+ \Lambda\tau \int_\Theta \int_{[0,1]} \frac{\partial z(i,\theta)}{\partial \xi} \left\{1 + (1-\tau)\frac{\partial q}{\partial \tau_0}(i,\theta)\right\} dik(\theta)d\theta + \frac{\tau_0 + \tau\tilde{x}}{\tau_0}\frac{\tilde{\theta}k(\tilde{\theta})}{1-K(\tilde{\theta})}\mathcal{E}_{\tilde{\theta},\tau_0} = 0.$$

Change variables to $x = z - q$, use the definition of the income effect η and rearrange to get:

$$-\int_0^\infty E\left[\left\{\frac{M}{\Lambda} + \frac{\tau}{1-\tau}\eta - 1\right\}\left\{1 + (1-\tau)\frac{\partial q}{\partial \tau_0}\right\}\middle| x\right] h(x)dx = \frac{\tau_0 + \tau\tilde{x}}{\tau_0}\frac{\tilde{\theta}k(\tilde{\theta})}{1-K(\tilde{\theta})}\mathcal{E}_{\tilde{\theta},\tau_0}. \tag{A.5}$$

A.3 Derivation of optimal tax equation (32)

The introduction of endogenous matching and vacancy posting modifies the planner's problem as (A.6).

$$\max_{\mathcal{P}} \int_{\underline{\theta}}^{\overline{\theta}} \Gamma(\theta)U(b,0,\theta)k(\theta)d\theta \tag{A.6}$$
$$+ \int_{\tilde{\theta}(\mathcal{P})}^{\overline{\theta}} \mu(\theta;\mathcal{P}) \int_0^1 \Gamma(\theta)\left\{V(q(i,\theta;\mathcal{P}),\theta;\mathcal{P}) - U(b,0,\theta)\right\} di\; k(\theta)d\theta,$$

subject to the resource constraint:

$$-b + \int_{\tilde{\theta}(\mathcal{P})}^{\overline{\theta}} \mu(\theta;\mathcal{P}) \left\{\tau \int_0^1 \{z(i,\theta;\mathcal{P}) - q(i,\theta;\mathcal{P})\}di + \tau_0\right\} k(\theta)d\theta \geq G.$$

Proceeding similarly to the derivation of (A.3) delivers:

$$-\int_0^\infty E\left[\left\{\frac{M}{\Lambda} + \frac{\tau}{1-\tau}\eta - 1\right\}\left\{x + (1-\tau)\frac{\partial q}{\partial \tau}\right\}\middle| x\right] h(x)dx$$
$$= \frac{\tau}{1-\tau}\int_0^\infty xE[\mathcal{E}^c_{x,1-\tau}|x]h(x)dx$$
$$+ \frac{1}{1-\tau}\int_0^\infty E\left[\left\{\frac{\Delta U}{\Lambda} + \tau x + \tau_0\right\}\mathcal{E}_{\mu,1-\tau} - (1-\tau)\frac{\partial q}{\partial(1-\tau)}\middle| x\right] h(x)dx, \tag{A.7}$$

where $\frac{\Delta U(i,\theta)}{\Lambda} := \frac{\Gamma(\theta)}{\Lambda}\{V(q(i,\theta;\mathcal{P}),\theta;\mathcal{P}) - U(b,0,\theta)\}$. The last line in (A.7) is obtained by augmenting the policymaker's first order condition with the impact of variation in τ on μ, changing variables to x in the integration and adding $E\left[\frac{\partial q}{\partial(1-\tau)}\right]$ to both sides of the equality.[20] Then using the fact that under the linear vacancy cost assumption, all job prices are directed to the payment of vacancy costs:

$$\int_\Theta E\left[q|\,\theta\right]\mu(\theta)k(\theta)d\theta - \int_\Theta \kappa(\theta, v(\theta))d\theta = 0.$$

Differentiating this last expression with respect to $1-\tau$ and rearranging:

$$\int_\Theta E\left[\frac{\partial q}{\partial(1-\tau)}\middle|\,\theta\right]\mu(\theta)k(\theta)d\theta = -\int_\Theta E\left[q|\,\theta\right]\frac{\partial \mu}{\partial(1-\tau)}(\theta)k(\theta)d\theta + \int_\Theta \frac{\partial \kappa(\theta, v(\theta))}{\partial v}\frac{\partial v(\theta)}{\partial(1-\tau)}d\theta.$$

Substituting for $\frac{\partial \kappa(\theta,v(\theta))}{\partial v}$ from the firms' vacancy posting first order conditions, using $\frac{\partial \mu}{\partial(1-\tau)} = \frac{\partial \mu}{\partial v}\frac{\partial v}{\partial(1-\tau)}$ and changing variables in the integration:

$$\begin{aligned}\int_0^\infty \left[\frac{\partial q}{\partial(1-\tau)}\middle|\,x\right]h(x)dx =& \int_0^\infty E\left[\frac{1}{\mu}\left\{\frac{\partial \mu}{\partial v} - N(q)\right\}q\frac{\partial v}{\partial 1-\tau}\middle|\,x\right]h(x)dx \\ =& \frac{1}{1-\tau}\int_0^\infty E\left[\left\{\mathcal{E}_{\mu,v} - \frac{vN(q)}{\mu}\right\}q\mathcal{E}_{v,1-\tau}\middle|\,x\right]h(x)dx. \qquad \text{(A.8)}\end{aligned}$$

Substitution of this into (A.7) gives (32).

B Proofs for Section 5

Proof of Proposition 1. Assuming preferences (33) and affine taxes, it follows from worker optimization (10) that a worker's pre-tax income choice satisfies $x(i,\theta) = \theta^{\frac{1+\gamma}{\gamma}}(1-\tau)^{\frac{1}{\gamma}} - q(i,\theta)$. Let $\underline{\theta}(x)$ denote the least talented worker earning income x. This worker must be at the top of her job market ladder and be paying the smallest job price $\underline{q}(\underline{\theta}(x))$ in talent market $\underline{\theta}(x)$. Substituting for this job price into the worker's optimal income choice, using the definition of β and inverting yields the expression for $\underline{\theta}(x)$ in (37). Let $\overline{\theta}(x)$ denote the most talented worker earning income x. This worker must be at the bottom of her job market ladder and be paying the largest job price $\overline{q}(\overline{\theta}(x))$ in talent market $\overline{\theta}(x)$. Substituting for this job price into the worker's optimal income choice and inverting yields the expression for $\overline{\theta}(x)$ in (37).

[20] Note that while $\tilde{\theta}$ depends on $\mathcal{P}$ induced variation in $\tilde{\theta}$ has no impact on the objective because in this case no vacancy posting occurs in the marginal talent market.

For talents between $\underline{\theta}(x)$ and $\overline{\theta}(x)$, there is a threshold $i(\theta, x)$ such that workers with $i \in (i(\theta, x), 1]$ earn less than x and those with $i \in [0, i(\theta, x))$ earn more. For such θ, $i(\theta, x)$ satisfies:

$$x = -\left\{\frac{-\tau_0}{1-\tau} + \frac{\gamma}{1+\gamma}\theta^{\frac{1+\gamma}{\gamma}}(1-\tau)^{\frac{1}{\gamma}}\right\}\left(\frac{1}{1+\frac{\hat{\lambda}(\theta)}{\hat{\delta}}(1-i(\theta,x))}\right)^2 + (1-\tau)^{\frac{1}{\gamma}}\theta^{\frac{1+\gamma}{\gamma}}. \quad \text{(B.1)}$$

Inverting (B.1) gives:

$$i(\theta, x) = \frac{\hat{\delta} + \hat{\lambda}(\theta)}{\hat{\lambda}(\theta)} - \frac{\hat{\delta}}{\hat{\lambda}(\theta)}\sqrt{\frac{-\frac{\tau_0}{1-\tau} + \frac{\gamma}{1+\gamma}\theta^{\frac{1+\gamma}{\gamma}}(1-\tau)^{\frac{1}{\gamma}}}{(1-\tau)^{\frac{1}{\gamma}}\theta^{\frac{1+\gamma}{\gamma}} - x}}.$$

Let h denote the income density of working agents. The talent and income distributions are related by:

$$\int_0^x h(x')dx' = \frac{\int_{\tilde{\theta}}^{\underline{\theta}(x)} \tilde{k}(\theta)d\theta}{\int_{\tilde{\theta}}^{\overline{\theta}} \tilde{k}(\theta)d\theta} + \int_{\underline{\theta}(x)}^{\overline{\theta}(x)} [1 - i(\theta, x)] \frac{\tilde{k}(\theta)}{\int_{\tilde{\theta}}^{\overline{\theta}} \tilde{k}(\theta')d\theta'} d\theta,$$

where $\tilde{k}(\theta) = \mu(\theta)k(\theta)$, the first right hand side term consists of all employed workers with talent below $\underline{\theta}(x; \mathcal{P})$ (all of whom earn less than income x) and the second consists of employed workers between $\underline{\theta}(x)$ and $\overline{\theta}(x;)$, some of whom earn less than x. All employed workers with talent above $\overline{\theta}(x)$ earn more than x. The term $1 - i(\theta, x)$ can be interpreted as the rung on the job ladder of a θ worker earning x. Differentiating both sides of the previous equation relates the income density to the talent density by:

$$h(x) = -\int_{\underline{\theta}(x)}^{\overline{\theta}(x)} \frac{\partial i}{\partial x}(\theta, x) \frac{\tilde{k}(\theta)}{\int_{\tilde{\theta}}^{\overline{\theta}} \tilde{k}(\theta')d\theta'} d\theta, \quad \text{(B.2)}$$

and $\Phi(x|\theta) = -\frac{\partial i}{\partial x}(\theta) = \frac{\hat{\delta}}{2\hat{\lambda}(\theta)}\sqrt{\frac{\frac{-\tau_0}{1-\tau} + \frac{\gamma}{1+\gamma}\theta^{\frac{1+\gamma}{\gamma}}(1-\tau)^{\frac{1}{\gamma}}}{\{(1-\tau)^{\frac{1}{\gamma}}\theta^{\frac{1+\gamma}{\gamma}} - x\}^3}}$ implying (34). The expression for l is derived similarly. □

C Numerical estimation procedure

Inversion of Fredholm equations Our numerical algorithm requires inversion of the Fredholm equations (38). In general Fredholm equations of the first kind lead to ill posed problems in which solutions are highly sensitive to parameters. Our procedure for recovering estimates of ι and ν given empirical proxies for h and u and the model implied $\Phi(\cdot|\cdot;\beta)$ builds on Fox et al (2016).[21] First select two families of parametric basis functions $\{\Upsilon_r\}_{r=1}^R$ and $\{\Xi_r\}_{r=1}^R$. Candidate densities ι and ν are approximated as mixtures of these basis functions:

$$\hat{\iota}(\theta;a) = \sum_{r=1}^R a_r \Upsilon_r(\theta) \qquad \text{and} \qquad \hat{\nu}(\theta;b) = \sum_{r=1}^R b_r \Xi_r(\theta),$$

where $a = \{a_r\}$ and $b = \{b_r\}$ belong to R-simplices. Let $\{x_i\}_{i=1}^I$ denote a grid of incomes and let $\hat{h}_i$ denote the fraction of agents with incomes between x_i and x_{i+1} (with $x_{I+1} = \infty$). Let $\hat{h}_i^u$ denote the fraction of unemployed workers whose last recorded income was between x_i and x_{i+1}. Unemployed workers with no previous income are treated as having talents below $\tilde{\theta}$.

The numerical procedure starts from an initial guess for the weights a^0 and b^0, which deliver initial ι^0 and ν^0 and, hence, $\beta^0 = \frac{\iota^0}{\nu^0}\frac{\mathcal{N}}{\mathcal{U}}$. On the j-th iteration, β^j is given and used to construct $\Phi(x|\theta;\beta^j)$ and $\underline{\theta}(x;\beta^j)$ and, hence, approximated components for the current income distribution of the employed:

$$m_{i,r}^j = \int_{x_i}^{x_{i+1}} \int_{\underline{\theta}(x;\beta^j)}^{\overline{\theta}(x)} \Phi(x|\theta;\beta^j)\Upsilon_r(\theta) d\theta dx$$

and last income distribution of the unemployed:

$$n_{i,r}^j = \int_{x_i}^{x_{i+1}} \int_{\underline{\theta}(x;\beta^j)}^{\overline{\theta}(x)} \Phi(x|\theta;\beta^j)\Xi_r(\theta) d\theta dx.$$

Updated coefficients $a^{j+1} = \{a_r^{j+1}\}$ and $b^{j+1} = \{b_r^{j+1}\}$ are obtained via the minimizations:

$$a^{j+1} = \arg\min_{a\in\Delta^R} \sum_{i=1}^I \left(\hat{h}_i - \sum_{r=1}^R a_r m_{i,r}^j\right)^2 \qquad \text{and} \qquad b^{j+1} = \arg\min_{b\in\Delta^R} \sum_{i=1}^I \left(\hat{h}_i^u - \sum_{r=1}^R b_r n_{i,r}^j\right)^2.$$

[21] Fox et al (2016) are concerned with the estimation of latent densities in discrete choice random coefficients settings. However, the mathematical structure of the discrete choice random coefficient model is a Fredholm equation of the first kind.

Solutions to the (inverted) Fredholm equations are then approximated as $\iota^{j+1} = \sum_{r=1}^{R} a_r^{j+1} \Upsilon_r(\theta)$ and $\nu^{j+1} = \sum_{r=1}^{R} b_r^{j+1} \Xi_r(\theta)$. From these $k^{j+1} = \mathcal{N}\frac{1+\beta^j}{\beta^{j+1}}\iota^{j+1}$ and an updated $\beta^{j+1} = \frac{\iota^{j+1}}{\nu^{j+1}}\frac{\mathcal{N}}{\mathcal{U}}$ are obtained. If the Euclidian distance between a^j and b^j and a^{j+1} and b^{j+1} is smaller than a pre-specified threshold, then convergence is achieved. The iteration is repeated until convergence. Estimates for k and β over the interval $[\tilde{\theta}(\mathcal{P}), \overline{\theta})$ combined with values for δ and α, permit recovery of κ over this range using the formula in the main text.

Choice of basis functions We first invert the empirical distribution of earnings among employed using the relationship between earnings and talent implied by the model without frictions, that is $x(\theta) = \theta^{1+1/\gamma}(1-\tau)^{1/\gamma}$, under the estimated status-quo tax policy, $\hat{\mathcal{P}}$. This delivers a talent distribution. We then fit a lognormal distribution to this talent distribution using Matlab's *lognfit* command. This produces a lognormal fit with a mean, $\mathcal{M}_E$, and a variance, $\mathcal{S}_E$, together with 95% confidence interval bounds for mean and variance. We construct a lower mean $\underline{\mathcal{M}}_E = M_E - 20 * (\mathcal{M}_E - \underline{\mathcal{M}}'_E)$ and a higher mean $\overline{\mathcal{M}}_E = \mathcal{M}_E + 20 * (\mathcal{M}_E - \underline{\mathcal{M}}'_E)$, where $\underline{\mathcal{M}}'_E$ is the lower bound on the 95% confidence interval around $\mathcal{M}_E$. Identically, we construct $\underline{\mathcal{S}}_E$ and $\overline{\mathcal{S}}_E$. We set the family of basis functions $(\Upsilon_r)_{r=1}^{9}$ to the $3 \times 3 = 9$ combination of lognormal distributions generated this way truncated at $\tilde{\theta}(\hat{\mathcal{P}})$ using Matlab's *truncate* command. We construct the basis functions $(\Xi_r)_{r=1}^{9}$ using an identical procedure, this time starting from the empirical distribution of last earnings among unemployed.

D Benchmark frictionless economy

D.1 Optimal tax problem

The benchmark frictionless economy is identical to the frictionless model with extensive margin introduced in Section 2. The policymaker chooses policy $\mathcal{P} = (b, \tau_0, \tau)$ to solve:

$$\max_{\mathcal{P}} \left\{ \int_{\underline{\theta}}^{\tilde{\theta}(\mathcal{P})} \Gamma(\theta)k(\theta)d\theta + \int_{\tilde{\theta}(\mathcal{P})}^{\overline{\theta}} (1-\mu(\theta))\Gamma(\theta)k(\theta)d\theta \right\} U(b,0,\theta)$$
$$+ \int_{\tilde{\theta}(\mathcal{P})}^{\overline{\theta}} \mu(\theta)\Gamma(\theta)U(b-\tau_0+(1-\tau)z(\theta;\mathcal{P}), z(\theta;\mathcal{P}), \theta)k(\theta)d\theta, \tag{D.1}$$

subject to the government budget constraint:

$$-b+\tau_0 \int_{\tilde{\theta}(\mathcal{P})}^{\overline{\theta}} \mu(\theta)k(\theta)d\theta + \tau \int_{\tilde{\theta}(\mathcal{P})}^{\overline{\theta}} \mu(\theta)z(\theta;\mathcal{P})\ k(\theta)d\theta \geq G,$$

where $\mu(\theta)$ gives the exogenous fraction of inactive workers in each talent market.

D.2 Calibration

The primitives of the model are parameters describing the utility function U, the talent distribution k, and the exogenous government spending G. We restrict the parametric form of U to be the same as in the case with frictions given by (33). We set $\sigma = 2$ and $\gamma = 1$ also to be in line with the frictional case. G is set to equal 25% of GDP in the optimal tax problem, again to be in line with the frictional model.

Following Saez (2001), we calibrate the talent distribution, k, by inverting the empirical distribution of earnings using the one-to-one mapping between talents and earnings given by the model under status-quo policy, $\hat{\mathcal{P}}$. As in the frictional calibration, we construct this earnings distribution using data from the Current Population Survey, March release of the 2017 survey which provides information for the calendar year 2016 and use March supplement sample weights to produce our estimates. With frictions, we match the distribution of the employed over current earnings and the unemployed over last earnings separately. In the frictionless case, we pool these distributions together to construct the empirical earnings distribution. The smallest active talent $\tilde{\theta}(\hat{\mathcal{P}})$ is determined by: $(1-\hat{\tau}) - \frac{\underline{x}^\gamma}{\tilde{\theta}(\hat{\mathcal{P}})^{1+\gamma}} = 0$, where $\hat{\tau} = 0.338$, $\underline{x}$ is the smallest income in our sample and $\gamma = 1$. Since $\underline{x}$ is identical in the frictional and frictionless earnings distributions, $\tilde{\theta}(\hat{\mathcal{P}})$ is also identical across the two environments.

Appendix References

Fox, J., K. Kim, C. Yang (2016). A simple nonparametric approach to estimating the distribution of random coefficients in structural models. *Journal of Econometrics 195*(2), 236–254.

Online Appendices

E Incidence of labor taxation under Nash Bargaining

In this section, we demonstrate that frictional profit squeeze effects are not specific to the particular frictional model we assume in the main body of the paper. To do so, we investigate the impact of taxation on surplus sharing between workers and firms under Nash bargaining.

Recall the problem of a θ talent employed worker who needs to pay job price q to a firm. Assuming the worker faces policy $\mathcal{P} = (b, \tau_0, \tau)$, the worker solves:

$$V(q, \theta; \mathcal{P}) = \max_{z} U(b - \tau_0 + (1-\tau)(z-q), z, \theta), \tag{E.1}$$

with an associated first-order optimality condition

$$(1-\tau)U_c(b - \tau_0 + (1-\tau)(z-q), z, \theta) + U_z(b - \tau_0 + (1-\tau)(z-q), z, \theta) = 0. \tag{E.2}$$

In what follows, we assume utility is separable between consumption and effort, i.e., $U_{cz} = 0$.

Suppose the job price q is determined according to a Nash bargaining protocol between the worker and the firm:

$$\max_{q} \Big(V(q, \theta; \mathcal{P}) - U(b, 0, \theta) \Big)^{\eta} q^{1-\eta}, \tag{E.3}$$

with η being the exogenously given bargaining power of the worker. For $\eta \in (0,1)$, the first-order condition with respect to q is given by

$$\frac{(1-\tau)qU_c(b - \tau_0 + (1-\tau)(z-q), z, \theta)}{V(q, \theta; \mathcal{P}) - U(b, 0, \theta)} - \frac{1-\eta}{\eta} = 0. \tag{E.4}$$

Equation (E.4) implicitly defines q as a function of τ. We next establish that, under a plausible assumption on preferences, $\frac{\partial q}{\partial \tau} < 0$. Consequently, a higher tax rate on labor income squeezes firm profits, as it does in our workhorse Burdett-Mortensen model. We then show that we can exactly replicate the distribution of job prices in the Burdett-Mortensen model in the bargaining model through careful selection of bargaining weights.

Toward this goal, denote the left-hand side of (E.4) by $F(\tau, q)$. Then, by the implicit function theorem, we have $\frac{\partial q}{\partial \tau} = -\frac{F_\tau}{F_q}$. Notice that

$$F_q = \frac{(1-\tau)U_c + q(1-\tau)^2 U_{cc}\left(\frac{\partial z}{\partial q} - 1\right)}{V(q,\theta;\mathcal{P}) - U(b,0,\theta)} + \frac{(1-\tau)^2 q U_c^2}{(V(q,\theta;\mathcal{P}) - U(b,0,\theta))^2} > 0. \tag{E.5}$$

The inequality follows from the fact that $\frac{\partial z}{\partial q} < 1$, which, in turn, follows from totally differentiating (E.2) with respect to q. This implies that if $F_\tau > 0$, then $\frac{\partial q}{\partial \tau} < 0$. As shown below, a sufficient condition for $F_\tau > 0$ is the assumption that $\frac{\partial z}{\partial \tau} < 0$, meaning that the substitution effect of a proportional tax change dominates the income effect. To see this, observe that

$$F_\tau = \frac{-qU_c + q(1-\tau)U_{cc}\left(-(z-q) + \frac{\partial z}{\partial \tau}(1-\tau)\right)}{V(q,\theta;\mathcal{P}) - U(b,0,\theta)} + \frac{(1-\tau)qU_c^2(z-q)}{(V(q,\theta;\mathcal{P}) - U(b,0,\theta))^2}. \tag{E.6}$$

Totally differentiating (E.2) with respect to τ and using that expression in (E.6) implies:

$$F_\tau = \frac{(1-\tau)qU_c^2(z-q)}{(V(q,\theta;\mathcal{P}) - U(b,0,\theta))^2} - U_{zz}\frac{\partial z}{\partial \tau}\frac{q}{V(q,\theta;\mathcal{P}) - U(b,0,\theta)}. \tag{E.7}$$

Hence, a sufficient condition for $F_\tau > 0$ is $\frac{\partial z}{\partial \tau} < 0$. Since:

$$\frac{\partial z}{\partial \tau} = \frac{(1-\tau)U_{cc}(z-q) + U_c}{U_{cc}(1-\tau)^2 + U_{zz}}, \tag{E.8}$$

it follows that if $|U_{cc}|$ is sufficiently small, then $\frac{\partial z}{\partial \tau} < 0$. This happens, for instance, when there are no income effects, i.e., $U_{cc} = 0$. Alternatively, if $u(c) = c^{1-\sigma}/(1-\sigma)$ and $\sigma \leq 1$, then $\frac{\partial z}{\partial \tau} < 0$ if $b - \tau_0 < 0$.

In the basic Burdett-Mortensen model, the job price function characterizes frictional inequality conditional on talent as $q(i,\theta;\mathcal{P}) = (1-\phi(i))\overline{q}(\theta;\mathcal{P})$, where $\phi(i) = 1 - \left(\frac{1}{1+\lambda/\delta(1-i)}\right)^2$, $\overline{q}(\theta;\mathcal{P})$ is the maximal job price that sets $V(\overline{q}(\theta;\mathcal{P}),\theta;\mathcal{P}) = U(b,0,\theta)$ and i is the percentile of the worker in the job price distribution with $i = 1$ and $i = 0$ denoting the bottom and the top of the job ladder, respectively. We can generate the same distribution of job prices in the Nash bargaining model by carefully choosing the distribution of bargaining weights in each talent market as follows. Set the bargaining weight of those at the bottom of the ladder in any talent market to $\eta(1) = 0$, which delivers job price $\overline{q}(\theta;\mathcal{P})$. The bargaining weight of those at percentile i is defined implicitly by the solution to the equation:

$$\frac{(1-\tau)q(i,\theta;\mathcal{P})U_c(-\tau_0+(1-\tau)(z-q(i,\theta;\mathcal{P})),z,\theta)}{V(q(i,\theta;\mathcal{P}),\theta;T)-U(b,0,\theta)}-\frac{1-\eta(i)}{\eta(i)}=0. \tag{E.9}$$

If agent utility is quasilinear, then one can explicitly calculate the distribution of bargaining weights that would deliver identical job price distributions across the two models at all tax systems. To see this, notice that, if utility function is of the $U(c,z,\theta)=c-v(z/\theta)$, then (E.4) delivers q in closed form:

$$q=(1-\eta)\overline{q}(\theta;\mathcal{P}), \tag{E.10}$$

where $\overline{q}(\theta;\mathcal{P})=\frac{-\tau_0}{1-\tau}+\frac{\gamma}{1+\gamma}(1-\tau)^{1/\gamma}\theta^{1+1/\gamma}$ is the maximal job price paid by workers who have no bargaining power. Then, setting $\eta(i)=\phi(i)$ ensures that the distribution of job prices is identical across the two economies. This implies that the profit squeeze effect is also identical: $\frac{\partial q(i,\theta;\mathcal{P})}{d\tau}=(1-\phi(i))\frac{\partial\overline{q}(\theta;\mathcal{P})}{d\tau}$ in both economies.

F 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 2017 survey which provides information for the calendar year 2016 and use March supplement sample weights to produce our estimates. The total number of observations in the raw sample is 185,914. We focus on population of working age individuals and so we drop people who are younger than 18 or older than 65 years of age. This reduces the sample size to 112,667. We also drop those giving inconsistent answers to questions about labor income and hours worked last year (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. These adjustments reduce the sample size further to 106,882 observations.

Of the people in this group, there are 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 80,423 working agents and 8,520 people who did not work during 2016. We identify the latter group of people with those below the active talent threshold $\tilde{\theta}$ implied by status quo US policy (and our assumed preference and labor market transition parameters).

Recall that our procedure for estimating k and $\kappa(\cdot)$ uses two earnings distributions as inputs: one for currently employed and one (distribution of last earnings) for currently unemployed. We construct these two earnings distributions using the remaining 80,423 people who were active the previous calendar year. Following Heathcote, Perri and Violante (2010), we drop those who work very little, less than two full weeks (80 hours) last year. We are left with 79,841 observations. Of this group, 73,329 are employed at the time of the survey and 2,286 were unemployed (the remaining were not in the labor force). The empirical counterparts of employed and unemployed earnings distributions come from these two groups of workers. We measure labor earnings 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.

Status-quo income tax estimation The affine approximation to status-quo US income tax policy, $T[x] = T_0 + \tau x$, is estimated by regressing total income taxes paid on labor income. We construct total income taxes paid by summing up CPS variables stataxac and fedtaxac for last calendar year. The former variable reports state level income tax liability after tax credits are deducted while the latter variable reports the federal income tax liability, again after deducting tax credits.

G Numerical algorithm for affine tax analysis

G.1 Equilibrium characterization

Assuming preferences (33) and a given policy $\mathcal{P}$, it follows from worker optimization (10) that a worker of type $\theta \geq \tilde{\theta}$ who is at quantile i of the job ladder in their talent market supplies effective labor $z(\theta;\mathcal{P}) = \theta^{\frac{1+\gamma}{\gamma}}(1-\tau)^{\frac{1}{\gamma}}$ and earns pre-tax income $x(i,\theta;\mathcal{P}) = \theta^{\frac{1+\gamma}{\gamma}}(1-\tau)^{\frac{1}{\gamma}} - q(i,\theta;\mathcal{P})$, where $q(i,\theta;\mathcal{P})$ is defined in (16) in the main text.

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 $\overline{q}(\theta; \mathcal{P})$ is given by:

$$-T_0 + (1-\tau)\{z(\theta; \mathcal{P}) - \overline{q}(\theta; \mathcal{P})\} - \frac{1}{1+\gamma}\left(\frac{z(\theta; \mathcal{P})}{\theta}\right)^{1+\gamma} = b.$$

The activity threshold $\tilde{\theta}$ is such that $\overline{q}(\tilde{\theta}; \mathcal{P}) = 0$. Extend the function $\overline{q}$ from $[\tilde{\theta}, \overline{\theta}]$ onto $[\underline{\theta}, \overline{\theta}]$ using:

$$\bar{q}(\theta; \mathcal{P}) = -\frac{b + T_0}{1-\tau} + \bar{z}(\theta; \mathcal{P}) \tag{G.1}$$

with $\bar{z}(\theta; \mathcal{P}) = z(\theta; \mathcal{P}) - \frac{(\frac{z(\theta; \mathcal{P})}{\theta})^{1+\gamma}}{(1+\gamma)(1-\tau)}$.

A firm in talent market θ chooses v and a job price $q \in [0, \overline{q}(\theta; \mathcal{P})]$ to maximize profits, i.e.,

$$\max_{q,v} N(q, \theta; \mathcal{P}) q v - \bar{\kappa}(\theta) v. \tag{G.2}$$

For a firm creating vacancies v and setting a job price $q \in [0, \overline{q}(\theta; \mathcal{P})]$ in market θ, steady-state employment is given by the expression:

$$N(q, \theta; \mathcal{P}) = \frac{\hat{\delta}\varphi(\theta; \mathcal{P})}{(\hat{\delta} + \hat{\lambda}(\theta; \mathcal{P}) F[q|\theta; \mathcal{P}])^2}. \tag{G.3}$$

Thus, the firm's problem (G.2) can be decomposed into an inner maximization over revenues per vacancy:

$$R(\theta; \mathcal{P}) := \max_{q \in [0, \overline{q}(\theta; \mathcal{P})]} \frac{\hat{\delta}\varphi(\theta; \mathcal{P}) q}{(\hat{\delta} + \hat{\lambda}(\theta; \mathcal{P}) F[q|\theta; \mathcal{P}])^2} \tag{G.4}$$

and an outer maximization over vacancies:

$$\max_{v \in \mathbb{R}_+} R(\theta; \mathcal{P}) v - \bar{\kappa}(\theta) v. \tag{G.5}$$

The vacancy problem (G.5) implies that in equilibrium $R(\theta; \mathcal{P}) = \bar{\kappa}(\theta)$. This, together with an evaluation of the firm's revenue per vacancy at $\overline{q}(\theta; \mathcal{P})$, yields the following expression:

$$\bar{\kappa}(\theta) = \frac{\hat{\delta}\bar{q}(\theta;\mathcal{P})}{\left(\hat{\delta}+\hat{\lambda}(\theta;\mathcal{P})\right)^2}\varphi(\theta;\mathcal{P}). \tag{G.6}$$

It is easy to show that under the assumption of linearity of vacancy posting costs, we have

$$\lambda(\theta;\mathcal{P}) = a\left(\frac{v(\theta;\mathcal{P})}{k(\theta)}\right)^{\alpha} \qquad \text{and} \qquad \varphi(\theta;\mathcal{P}) = a\left(\frac{k(\theta)}{v(\theta;\mathcal{P})}\right)^{1-\alpha}. \tag{G.7}$$

Substituting these into (G.6) yields:

$$\bar{\kappa}(\theta) = \frac{a^{\frac{1}{\alpha}}\hat{\delta}\bar{q}(\theta;\mathcal{P})}{\left(\hat{\delta}+\hat{\lambda}(\theta;\mathcal{P})\right)^2}\lambda(\theta;\mathcal{P})^{\frac{\alpha-1}{\alpha}}. \tag{G.8}$$

G.2 Government budget constraint

An important simplification of the analysis is that we can rewrite the government's budget constraint in a way that does not depend on profit distributions. This allows us to express T_0 as a function of the parameters of the model, $\tau, b, \tilde{\delta}, \gamma, k(\theta)$, and endogenous variables $\tilde{\theta}(p)$ and $\mu(\theta;\mathcal{P})$.

To do so, we can manipulate the government budget constraint to compute τ_0 for given b and τ as follows. Government's budget constraint is:

$$\begin{aligned} &G + b\Big[K(\tilde{\theta}(\mathcal{P})) + \int_{\tilde{\theta}(\mathcal{P})}^{\bar{\theta}}(1-\mu(\theta;\mathcal{P}))k(\theta)d\theta\Big] - T_0\int_{\tilde{\theta}(\mathcal{P})}^{\bar{\theta}}\mu(\theta;\mathcal{P})k(\theta)d\theta \\ &= \int_{\tilde{\theta}(\mathcal{P})}^{\bar{\theta}}\mu(\theta;\mathcal{P})\int_0^1 \tau x(q(i,\theta;\mathcal{P}),\theta)dik(\theta)d\theta. \end{aligned} \tag{BC}$$

Using $x(q(i,\theta;\mathcal{P}),\theta) = z(\theta;\mathcal{P}) - q(i,\theta;\mathcal{P})$, we can rewrite, the government budget constraint as:

$$\begin{aligned} &G + b\Big[K(\tilde{\theta}(\mathcal{P})) + \int_{\tilde{\theta}}^{\bar{\theta}}(1-\mu(\theta;\mathcal{P}))k(\theta)d\theta\Big] - T_0\int_{\tilde{\theta}(\mathcal{P})}^{\bar{\theta}}\mu(\theta;\mathcal{P})k(\theta)d\theta \\ &= \tau\int_{\tilde{\theta}(\mathcal{P})}^{\bar{\theta}}\mu(\theta;\mathcal{P})z(\theta;\mathcal{P})k(\theta)d\theta - \tau\int_{\tilde{\theta}(\mathcal{P})}^{\bar{\theta}}\mu(\theta;\mathcal{P})\int_0^1 q(i,\theta;\mathcal{P})dik(\theta)d\theta. \end{aligned} \tag{G.9}$$

Notice that the term that follows the first integral sign in the second term on the second line of (G.9) equals total job price payment in the economy which further equals total cost of vacancy posting since firms make zero profit in equilibrium. Therefore, we have

$$\int_{\tilde{\theta}(\mathcal{P})}^{\bar{\theta}} \mu(\theta;\mathcal{P}) \int_0^1 q(i,\theta;\mathcal{P}) di k(\theta) d\theta = \int_{\tilde{\theta}(p)}^{\bar{\theta}} \bar{\kappa}(\theta) v(\theta;\mathcal{P}) d\theta, \tag{G.10}$$

where $v(\theta;\mathcal{P})$ is the aggregate level of vacancy posting in talent market θ and $\bar{\kappa}(\theta)v(\theta;\mathcal{P})$ is the total vacancy posting cost in that talent market under the assumption of linear vacancy posting. Notice that (G.7) implies that $\lambda(\theta;\mathcal{P})k(\theta) = \varphi(\theta;\mathcal{P})v(\theta;\mathcal{P})$. Using this latter and (G.6), and recalling that $\mu(\theta;\mathcal{P}) = \frac{\lambda(\theta;\mathcal{P})}{\lambda(\theta;\mathcal{P})+\delta}$, we have that aggregate vacancy posting cost equals:

$$\int_{\tilde{\theta}(\mathcal{P})}^{\bar{\theta}} \bar{\kappa}(\theta) v(\theta) d\theta = \int_{\tilde{\theta}(\mathcal{P})}^{\bar{\theta}} \mu(\theta;\mathcal{P}) \frac{\hat{\delta}}{\hat{\delta} + \hat{\lambda}(\theta;\mathcal{P})} \overline{q}(\theta;\mathcal{P}) k(\theta) d\theta. \tag{G.11}$$

Thus, we can rewrite (G.9) as:

$$\begin{aligned} & G + b\Big[K(\tilde{\theta}(\mathcal{P})) + \int_{\tilde{\theta}(\mathcal{P})}^{\bar{\theta}} (1-\mu(\theta;\mathcal{P}))k(\theta)d\theta\Big] - T_0 \int_{\tilde{\theta}(\mathcal{P})}^{\bar{\theta}} \mu(\theta;\mathcal{P})k(\theta)d\theta \\ & = \tau \int_{\tilde{\theta}(\mathcal{P})}^{\bar{\theta}} \mu(\theta;\mathcal{P}) z(\theta;\mathcal{P}) k(\theta) d\theta - \tau \int_{\tilde{\theta}(\mathcal{P})}^{\bar{\theta}} \mu(\theta;\mathcal{P}) \frac{\hat{\delta}}{\hat{\delta} + \hat{\lambda}(\theta;\mathcal{P})} \overline{q}(\theta;\mathcal{P}) k(\theta) d\theta. \end{aligned} \tag{G.12}$$

Plugging $\overline{q}(\theta;\mathcal{P}) = -\frac{b+\tau_0}{1-\tau} + \bar{z}(\theta)$, where recall that $\bar{z}(\theta) = z(\theta) - \frac{(\frac{z(\theta)}{\theta})^{1+\gamma}}{(1+\gamma)(1-\tau)}$, into (G.12) and leaving τ_0 alone, we get:

$$T_0 = -\frac{1}{A}\Big[-G - Bb + \tau \int_{\tilde{\theta}(\mathcal{P})}^{\bar{\theta}} \mu(\theta;\mathcal{P}) z(\theta;\mathcal{P}) k(\theta) d\theta - \tau \int_{\tilde{\theta}(p)}^{\bar{\theta}} \mu(\theta;\mathcal{P}) \frac{\hat{\delta}}{\hat{\delta} + \hat{\lambda}(\theta;\mathcal{P})} \bar{z}(\theta;\mathcal{P}) k(\theta) d\theta\Big] \tag{G.13}$$

where

$$A := \int_{\tilde{\theta}(\mathcal{P})}^{\bar{\theta}} \mu(\theta;\mathcal{P}) \Big[1 + \frac{\tau}{1-\tau} \frac{\hat{\delta}}{\hat{\delta} + \hat{\lambda}(\theta;\mathcal{P})}\Big] k(\theta) d\theta,$$

$$B := K(\tilde{\theta}(\mathcal{P})) + \int_{\tilde{\theta}(\mathcal{P})}^{\bar{\theta}} \left((1-\mu(\theta;\mathcal{P})) - \frac{\tau}{1-\tau} \mu(\theta;\mathcal{P}) \frac{\hat{\delta}}{\hat{\delta} + \hat{\lambda}(\theta;\mathcal{P})} \right) k(\theta) d\theta.$$

G.3 Algorithm

The algorithm has two main steps. In the first, we compute equilibrium for a given choice of (b, τ). In the second step, we construct a grid for b and τ and do a grid search to find the b and τ that maximizes social welfare. Below, we provide the sub-steps of the first equilibrium computation step.

Substep 1. For a given b and τ, guess T_0^1. Set the substep counter $k = 1$ and proceed to the next substep.

Substep 2. Use (G.1) to compute the activity threshold $\tilde{\theta}$. Use

$$\bar{\kappa}(\theta) = \frac{a^{\frac{1}{\alpha}}\hat{\delta}\bar{q}(\theta;\mathcal{P})}{\left(\hat{\delta} + \hat{\lambda}(\theta;\mathcal{P})\right)^2}\lambda(\theta;\mathcal{P})^{\frac{\alpha-1}{\alpha}} \tag{G.14}$$

to compute $\lambda(\theta;\mathcal{P})$ and that to compute $\mu(\theta;\mathcal{P}) = \frac{\lambda(\theta;\mathcal{P})}{\delta+\lambda(\theta;\mathcal{P})}$. Using (G.13), compute $\hat{T}_0^k$. Calculate $test := |T_0^k - \hat{T}_0^k|$. If test$< \varepsilon$, then equilibrium T_0 is found, proceed to Step 3. Otherwise, set $k = k+1$, $T_0^{k+1} = \eta T_0^k + (1-\eta)\hat{T}_0^k$ and repeat Substep 2 until convergence.

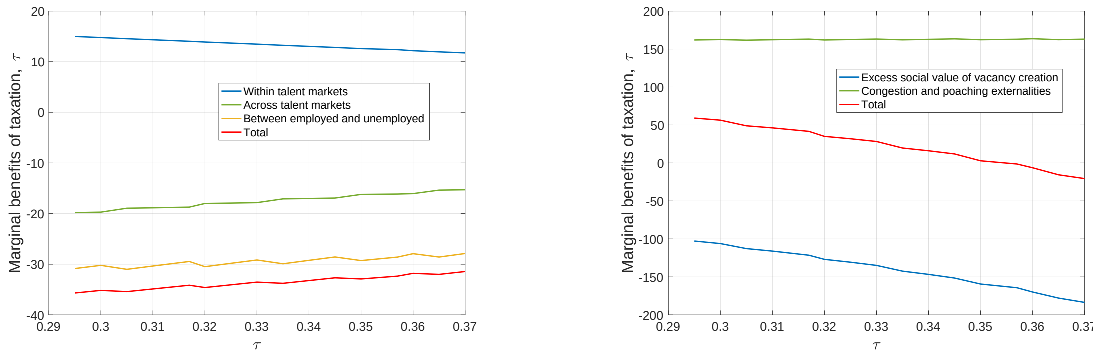
Substep 3. Use (G.1) to compute activity threshold $\tilde{\theta}(p)$, the maximal job price $\bar{q}(\theta;\mathcal{P})$, and (16) to compute the distribution of job prices in each talent market $q(i,\theta;\mathcal{P})$, for all $\theta \geq \tilde{\theta}$. Compute social welfare under given policy.

H Additional quantitative results

This appendix reports some additional quantitative results.

H.1 Decomposition of optimal marginal benefits terms

Subsection 5.2 quantitatively evaluates marginal benefit and cost terms for the frictonal economy optimal income tax equation. Figure H.1 provides further decomposition of these terms. Recall that $-E\left[\left\{\frac{M}{\Lambda} - 1\right\}\left\{x + (1-\tau)\frac{\partial q}{\partial \tau}\right\}\right]$ is the marginal redistributive benefit of an income tax rate increment, with the term $-E\left[\left\{\frac{M}{\Lambda} - 1\right\}\left\{(1-\tau)\frac{\partial q}{\partial \tau}\right\}\right]$ being the contribution of the profit squeeze to this benefit. Figure H.1(a) further decomposes the profit squeeze contribution. The overall value of this contribution at different marginal tax rates τ (holding other



(a) Decomposition of profit squeeze redistribution term

(b) Decomposition of job destruction term

Figure H.1

elements of the tax system fixed at their optimal values) is depicted by the red line in the figure (with units: 2017 per capita dollars). The profit squeeze redistributes amongst employed workers within and across talent markets and between the employed and the unemployed. The blue line illustrates the value of profit squeeze induced redistribution amongst employed workers *within* talent markets. Consistent with theory, this is positive: maximal job prices within talent markets decline in response to higher tax rates. This decline dissipates as workers climb the job ladder, implying that low earners within talent markets are the main beneficiaries and social criteria that favor such redistribution are elevated. However, in our quantitative exercise, this benefit is offset by the marginal cost of profit squeeze induced redistribution across talent markets. This is indicated by the green line in the figure. Theory for the model without matching, implies that this term is negative: job price falls are greatest in higher talent markets implying redistribution from poorer to richer workers across talent markets and, under out utilitarian criterion, a redistributive cost. The profit squeeze also transfers resources from the policymaker's budget to the employed. Since these resources are partly spent on providing basic income to the unemployed an implicit redistribution from the unemployed to the employed occurs. This is a further negative redistributive cost of the profit squeeze shown by the orange line in the figure.

Figure H.1(b) displays components of the marginal benefit terms associated with the employment margin and given in the last line of Equation (32). These components are consolidated into the orange line in Figure 3 in the main text (which coincides with the red total line plot in Figure H.1(b)). Recall that when creating vacancies, firms fail to internalize the positive utility benefits to unemployed workers who find jobs and the extra tax revenues created. On the other hand, they also do not internalize the negative congestion and poaching impacts on other firms. A higher income tax rate deters vacancy creation. The social marginal cost stemming from lost utility to job finders and reduced tax revenues is shown in blue. This declines from about -$103 to -$184 2017 per capita US dollars over the tax range displayed. The social marginal benefit from correcting congestion and poaching is stable at around $162 over this tax range.

H.2 Sensitivity

This subsection reports how optimal tax and benefit policy varies with utility parameters σ and γ. The conclusion that frictions are a force for moderately lower optimal tax rate carries over to the case with a higher concavity of the utility function and a lower elasticity of labor supply.

Table H.1: Optimal Affine Tax Policy

	$\sigma = 4$		$\gamma = 2$	
	Frictional	Frictionless	Frictional	Frictionless
τ	34.7%	39.3%	45.2%	47.9%
τ_0	6612	3168	6324	2436
b	9960	10104	12720	11520

τ_0, b: annual 2017 US $ amounts.

I Optimal non-linear taxation

This appendix extends optimal affine tax equations for frictional economies to the optimal nonlinear taxation case. It does so for the general Burdett-Mortensen model with endogenous matching. We slightly generalize the model in the main text by allowing vacancy costs to have the form: $\kappa(v, \theta) = \overline{\kappa}(\theta)\frac{v^{1+\chi}}{1+\chi}$. This implies that firms earn rents on infra-marginal vacancies (that are taxed at 100%). Formulas for other cases, e.g. the bBM model and the frictionless model can be derived as special or limiting cases. In the nonlinear tax setting, the policymaker has access to policies $\mathcal{P} = (b, T)$, with $T : \mathbb{R}_+ \to \mathbb{R}$ a twice differentiable non-linear tax function. The equilibrium model is identical to that for the affine case up to the modification of the worker's optimized value function and effort choice problem which becomes:

$$V(q, \theta; \mathcal{P}) = \max_z U(z - q - T[z - q], z, \theta). \tag{I.1}$$

Maximal job prices $\bar{q}$ continue to satisfy: $V(\bar{q}, \theta; \mathcal{P}) = U(b, 0, \theta)$, with V modified as in (I.1). Formulas (25) to (29) continue to hold. The policymaker's problem is now:

$$\begin{aligned}\max_{\mathcal{P}} \int_{\underline{\theta}}^{\overline{\theta}} \Gamma(\theta) U(b, 0, \theta) k(\theta) d\theta & \qquad \text{(I.2)} \\ + \int_{\tilde{\theta}(p)}^{\overline{\theta}} \mu(\theta; \mathcal{P}) \int_0^1 \Gamma(\theta) \left\{ V(q(i, \theta; \mathcal{P}), \theta; \mathcal{P}) - U(b, 0, \theta) \right\} di \; k(\theta) d\theta, &\end{aligned}$$

subject to the resource constraint:

$$\begin{aligned}-b + \chi \int_{\tilde{\theta}(p)}^{\overline{\theta}} \mu(\theta; \mathcal{P}) \int_{[0,1]} q(i, \theta; \mathcal{P}) di \; k(\theta) d\theta \\ + \int_{\tilde{\theta}(p)}^{\overline{\theta}} \mu(\theta; \mathcal{P}) \left\{ \int_0^1 T[z(i, \theta; \mathcal{P}) - q(i, \theta; \mathcal{P})] di \right\} k(\theta) d\theta \geq G.\end{aligned}$$

Let $\Omega : \mathbb{R}_+ \to \mathbb{R}$ denote a smooth function and consider a perturbation to the optimal tax code in the direction Ω. The resulting first order condition formatted in terms of incomes x is:

$$\begin{aligned}&- \int_{\mathbb{R}_+} E\left[\left\{ \frac{M}{\Lambda} + \frac{T'[x]}{1 - T'[x]} \tilde{\eta} - 1 \right\} \left\{ \Omega[x] + (1 - T'[x]) \partial_\Omega q \right\} \middle| x \right] h(x) dx \\ &\quad = \int_{\mathbb{R}_+} \left\{ \Omega'[x] - T''[x] \partial_\Omega q \right\} \frac{x T'[x]}{1 - T'[x]} E[\tilde{\mathcal{E}}^c | x] h(x) dx \\ &\quad\quad + \int_{\mathbb{R}_+} \frac{1}{1 - T'[x]} E\left[\left\{ \frac{\Delta U}{\Lambda} + T[x] + b + \chi q \right\} \mathcal{E}_{\Omega,\mu} + (1 - \chi)(1 - T'[x]) \partial_\Omega q \middle| x \right] h(x) dx,\end{aligned} \tag{I.3}$$

where ∂q_Ω is understood as the differential $\left.\frac{\partial q(T + \varepsilon \Omega)}{\partial \varepsilon}\right|_{\varepsilon = 0}$ and where: $\tilde{\mathcal{E}}^c = \frac{\mathcal{E}^c}{1 + \frac{T''[x] x}{1 - T'[x]} \mathcal{E}^c}$ and $\tilde{\eta} = \frac{\eta}{1 + \frac{T''[x] x}{1 - T'[x]} \mathcal{E}^c}$ are the non-linear tax adjusted compensated effective labor supply elasticity and income effect. Finally, $\mathcal{E}_{\Omega,\mu}$ gives elasticities of μ with respect to the tax perturbation. Expression (I.3) preserves the structure of the optimal affine tax formulas and its terms admit the same interpretation. It is, however, parameterized by the particular perturbation Ω. We next seek a factorization of the formula that renders it multiplicative in Ω. First, we deploy the standard integration by parts technique to reformulate the Ω' term in (I.3) in terms of Ω:

$$\int_0^\infty \frac{T'[x]x}{1-T'[x]} E[\tilde{\mathcal{E}}^c|x]\Omega'(x)h(x)dx = \frac{T'[x]x}{1-T'[x]} E[\tilde{\mathcal{E}}^c|x]\Omega(x)h(x)\Big]_0^\infty - \int_0^\infty \frac{d}{dx}\left\{\frac{T'[x]x}{1-T'[x]} E[\tilde{\mathcal{E}}^c|x]h(x)\right\}\Omega(x)dx. \tag{I.4}$$

Assuming that the first right hand side term in (I.4) is zero, this permits the gathering of terms in (I.3) into a standard Mirrleesian budget perturbation:

$$\int_{\mathbb{R}_+} \partial\mathcal{B}^M(x)\cdot\Omega[x]h(x)dx := \int_0^\infty \Bigg[1 - \frac{T'[x]}{1-T'[x]}E[\tilde{\eta}|x] + \left(\frac{x}{\frac{T'[x]}{1-T'[x]}E[\tilde{\mathcal{E}}^c|x]}\frac{\partial\left(\frac{T'[x]}{1-T'[x]}E[\tilde{\mathcal{E}}^c|x]\right)}{\partial x} + 1 + \frac{xh'(x)}{h(x)}\right)\frac{T'[x]}{1-T'[x]}E[\tilde{\mathcal{E}}^c|x]\Bigg]\Omega(x)h(x)dx.$$

Substituting this into the policymaker first order condition gives:

$$\begin{aligned}
-\int_{\mathbb{R}_+} E\left[\frac{M}{\Lambda}\{\Omega[x] + (1-T'[x])\partial_\Omega q\} + \left\{\frac{T'[x]}{1-T'[x]}\tilde{\eta} - 1\right\}(1-T'[x])\partial_\Omega q \,\Big|\, x\right] h(x)dx \\
= -\int_{\mathbb{R}_+}\{T''[x]\partial_\Omega q\}\frac{T'[x]x}{1-T'[x]}E[\tilde{\mathcal{E}}^c|x]h(x)dx - \int_{\mathbb{R}_+}\partial\mathcal{B}^M(x)\cdot\Omega[x]h(x)dx \\
+ \int_{\mathbb{R}_+}\frac{1}{1-T'[x]}E\left[\left\{\frac{\Delta U}{\Lambda} + T[x] + b + \chi q\right\}\mathcal{E}_\mu + (1-\chi)(1-T'[x])\partial_\Omega q \,\Big|\, x\right] h(x)dx.
\end{aligned} \tag{I.5}$$

In (I.5), $E[\partial_\Omega q|x]$ gives the expected perturbation to job prices of those earning x. To factorize the expression, we now calculate the expected perturbation to job prices induced by the tax function perturbation at x. Note that such a tax perturbation impacts job prices of those talents for whom x is the lowest income earned. As a first step evaluating $\partial_\Omega q$ at talent-extractiveness index pair (i, θ)

gives:

$$\partial q_{\Omega}(i,\theta) = -\left\{1 - 2\left(\frac{\hat{\lambda}(\theta)/\hat{\delta}(\theta)\}(1-i)}{1+\{\hat{\lambda}(\theta)/\hat{\delta}(\theta)\}(1-i)}\right)\mathcal{E}_{\lambda,\overline{q}}(\theta)\right\}\frac{q(i,\theta)}{\overline{q}(\theta)}\frac{\Omega[\underline{x}(\theta)]}{1-T'[\underline{x}(\theta)]}$$
$$= -\mathcal{M}(i,\theta)\frac{\Omega[\underline{x}(\theta)]}{1-T'[\underline{x}(\theta)]}, \tag{I.6}$$

where $\underline{x}(\theta)$ is the minimal income in market θ and where, in the second line of (I.6), $\mathcal{M}(i,\theta) = \left\{1 - 2\left(\frac{\{\hat{\lambda}(\theta)/\hat{\delta}(\theta)\}(1-i)}{1+\{\hat{\lambda}(\theta)/\hat{\delta}(\theta)\}(1-i)}\right)\mathcal{E}_{\lambda,\overline{q}}(\theta)\right\}\frac{q(i,\theta)}{\overline{q}(\theta)}$. To compute the total job price response induced by an income tax perturbation at a given x, define the maximal talent earning income x, $\overline{\theta}(x)$, according to: $x = \underline{x}(\overline{\theta}(x))$. Changing variables from θ to x, we have that if taxes are perturbed by Ω and, in particular, taxes are increased at income x by $\Omega(x)$, agents $(i,\overline{\theta}(x))$ experience a job price reduction of:

$$\partial_{\Omega} q(i,\overline{\theta}(x)) = -\mathcal{M}(i,\overline{\theta}(x))\frac{\Omega(x)}{1-T'[x]}.$$

Next evaluating the sensitivity of $\mu(\theta) = \frac{\lambda(\theta)}{\delta+\lambda(\theta)}$ gives:

$$\frac{\partial_{\Omega}\mu(\theta)}{\mu} = \frac{1}{1+\lambda(\theta)/\delta}\frac{\partial\lambda(\theta)}{\lambda} = \mathcal{E}_{\lambda,\overline{q}}(\theta)\frac{1}{1+\lambda(\theta)/\delta}\frac{\partial\overline{q}(\theta)}{\overline{q}} = -\mathcal{E}_{\lambda,\overline{q}}(\theta)\frac{1}{1+\lambda(\theta)/\delta}\frac{1}{\overline{q}(\theta)}\frac{\Omega(\underline{x}(\theta))}{1-T'[\underline{x}(\theta)]},$$

where again $\underline{x}(\theta)$ is the minimal income in talent market θ. Changing variables as before to obtain the impact of a tax perturbation at x gives:

$$\frac{\partial_{\Omega}\mu(\overline{\theta}(x))}{\mu} = -\mathcal{E}_{\overline{q},\lambda}(\overline{\theta}(x))\frac{1}{1+\lambda(\overline{\theta}(x))/\delta}\frac{1}{\overline{q}(\overline{\theta}(x))}\frac{\Omega(x)}{1-T'[x]},$$

Collecting terms the overall impact of a tax perturbation at x via adjustment in job prices and in the employment rate is:

$$\mathcal{D}(x)\cdot\frac{\Omega(x)}{1-T'[x]} = -\int_0^1(\chi - T'[x(\overline{\theta}(x),i)]\phi(\overline{\theta}(x),i))\mathcal{M}(\overline{\theta}(x),i)di\cdot\frac{k(\overline{\theta}(x))}{h(x)}\cdot\frac{\Omega(x)}{1-T'[x]}$$
$$-\int_0^1\left\{T[x(\overline{\theta}(x),i)] + b + \chi q(\overline{\theta}(x),i)\right\}di$$
$$\cdot\frac{\lambda(\overline{\theta}(x))}{\delta+\lambda(\overline{\theta}(x))}\frac{\delta}{\delta+\lambda(\overline{\theta}(x))}\frac{\mathcal{E}_{\lambda,\overline{q}}(\overline{\theta}(x))}{\overline{q}(\overline{\theta}(x))}\frac{k(\overline{\theta}(x))}{h(x)}\cdot\frac{\Omega(x)}{1-T'[x]},$$

where $\phi := \frac{1+\eta}{1+\frac{T''[x]x}{1-T'[x]}\mathcal{E}^c}$. This term gives the impact on the policymaker's budget of tax-induced adjustment in job prices and the employment rate. It nets out the

the marginal social cost of adjustments in vacancy posting implied by externalities in the posting process (and that translate into and are absorbed by reduced profit tax revenues). Substituting, the overall impact on the policymaker's budget from the reform is:

$$\partial\mathcal{B}(\Omega) = \int_0^\infty \left[\frac{\mathcal{D}(x)}{1-T'[x]} + \partial\mathcal{B}^M(x)\right]\Omega(x)h(x)dx.$$

The direct effect on social welfare of the perturbation is:

$$\int_{\mathbb{R}_+} \left\{-E\left[\frac{M}{\Lambda}\middle| x\right] + \frac{\mathcal{F}(x)}{1-T'[x]}\right\}\Omega[x]h(x)dx,$$

where:

$$\mathcal{F}(x)\cdot\frac{\Omega(x)}{1-T'[x]} = \int_0^1 \frac{M(\overline{\theta}(x),i)}{\Lambda}(1-T'[x(\overline{\theta}(x),i)])\mathcal{M}(\overline{\theta}(x),i)di\cdot\frac{k(\overline{\theta}(x))}{h(x)}\cdot\frac{\Omega(x)}{1-T'[x]}$$
$$+\int_0^1 \frac{\Delta U}{\Lambda}(x(\overline{\theta}(x),i)di\cdot\frac{\lambda(\overline{\theta}(x))}{\delta+\lambda(\overline{\theta}(x))}\frac{\delta}{\delta+\lambda(\overline{\theta}(x))}\frac{\mathcal{E}_{\lambda,\overline{q}}(\overline{\theta}(x))}{\overline{q}(\overline{\theta}(x))}\frac{k(\overline{\theta}(x))}{h(x)}\cdot\frac{\Omega(x)}{1-T'[x]},$$

Assembling the pieces gives:

$$0 = \int_{\mathbb{R}_+} \left\{-E\left[\frac{M}{\Lambda}\middle| x\right] + \frac{\mathcal{F}(x)}{1-T'[x]} + \frac{\mathcal{D}(x)}{1-T'[x]} + \partial\mathcal{B}^M(x)\right\}\Omega[x]h(x)dx. \tag{I.7}$$

Since the preceding must hold for an arbitrary perturbation at the optimum, for almost all x:

$$0 = -E\left[\frac{M}{\Lambda}\middle| x\right] + \frac{\mathcal{F}(x)}{1-T'[x]} + \partial\mathcal{B}^M(x) + \frac{\mathcal{D}(x)}{1-T'[x]},$$

where the first two terms give the direct welfare impact of the tax perturbation inclusive of job price adjustment on the job ladder $\overline{\theta}(x)$ and impact on the numbers of involuntarily unemployed agents in this labor market and the final two terms give budgetary implications inclusive of job price and employment effects in market $\overline{\theta}(x)$. Of course, (I.7) reduces to the affine optimal tax formula for affine translations Ω of an optimal affine tax code T and, more generally, the terms in (I.7) can be related back to their optimal affine counterparts.

ONLINE 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.