Federal Reserve Bank of Chicago

Redistribution with Performance Pay

Pawel Doligalski, Abdoulaye Ndiaye, and Nicolas Werquin

April 22, 2022

WP 2022-12

https://doi.org/10.21033/wp-2022-12

Redistribution with Performance Pay

Paweł Doligalski **Abdoulaye Ndiaye** **Nicolas Werquin** *

April 22, 2022

Abstract

Half of the jobs in the U.S. feature pay-for-performance. We derive novel incidence and optimum formulas for the overall rate of tax progressivity and the top tax rates on total earnings and bonuses, when such labor contracts arise from moral hazard frictions within firms. Optimal taxes account for the fiscal externalities and welfare consequences of two distinct forces: a direct crowding-out of private insurance and a countervailing crowding-in due to endogenous labor effort responses. These imply that the amount of pre-tax earnings risk to which the worker is exposed is roughly invariant to tax progressivity, whereas the (adverse) welfare consequences of the crowd-out outweigh those of the crowd-in. Quantitatively, the optimal tax policy with performance-pay contracts is close to that prescribed by standard models that treat pre-tax earnings risk as exogenous. Finally, we uncover an efficiency-based argument for taxing bonuses at strictly lower rates than base earnings.

*Doligalski: University of Bristol. Ndiaye: New York University, Stern School of Business and CEPR. Werquin: Federal Reserve Bank of Chicago, Toulouse School of Economics, and CEPR. Opinions expressed in this article are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Chicago or the Federal Reserve System. Nicolas Werquin acknowledges support from ANR under grant ANR-17-EURE-0010 (Investissements d'Avenir program). We thank Árpád Ábrahám, Joydeep Bhattacharya, Katherine Carey, Ashley Craig, Alex Edmans, Emmanuel Farhi, Antoine Ferey, Xavier Gabaix, Daniel Garrett, Mike Golosov, Andreas Haufler, Louis Kaplow, Paul Kindsgrab, Helen Koshy, Etienne Lehmann, Iacopo Morchio, Sergio Feijoo Moreira, Yena Park, Laura Pilossoph, Morten Ravn, Kjetil Storesletten, Florian Scheuer, Karl Schulz, Ali Shourideh, Stefanie Stantcheva, Aleh Tsyvinski, Hélène Turon, Venky Venkateswaran, Philipp Wangner, and Larry White, for helpful comments and suggestions. Emmanuel Farhi passed away a few weeks after we completed the first version of this paper. While the feedback he gave us about this paper in particular was very helpful, our whole generation of public finance economists owes him a much larger debt for his fundamental contributions to the field, his kindness and generosity, and his unwavering support. He continues to be a source of inspiration. This paper is dedicated to his memory.

Introduction

The dramatic increase in income inequality observed since the 1980s is in large part due to the explosion of performance-based forms of remuneration at the top of the income distribution, such as the rise in bankers' bonuses or CEOs' stock options (see, e.g., Bell and Van Reenen 2013, 2014; Lemieux, MacLeod, and Parent 2009; Piketty and Saez 2003). While performance-pay contracts are particularly prevalent for high earners, they are common throughout the income distribution and across occupations, from agricultural workers paid a piece rate to real estate brokers or retail workers who earn a commission on their sales (Bandiera, Barankay, and Rasul 2005; Levitt and Syverson 2008; Shearer 2004). Performance incentives can also be provided over time via promotions or salary raises.[1] Lemieux et al. (2009) estimate that in the U.S., almost half of all jobs, and three quarters within the top percentile of the income distribution, involve performance-based compensation. In the UK, Bell and Van Reenen (2014) find that bonuses are a feature of more than 80 percent of the contracts of the top 1 percent earners and can account for up to 35 percent of their compensation.

Given the importance of performance-based earnings and their contribution to growing income inequality, it is natural to ask how governments should tax them. What is the incidence of tax policy on performance-based contracts? Should the overall progressivity of the income tax schedule be modified to account for the existence of these jobs? How should income taxes on top earners be designed when their earnings are partly composed of bonuses? Should bonuses be treated separately from base earnings by the tax system, and if so, at what rates should they be taxed?

This paper provides answers to these questions. To do so, we set up a model of moral hazard in the labor market that gives rise to performance pay in equilibrium. In our model, income disparities arise from two distinct sources: namely, ex-ante ability differences and ex-post performance (or output) shocks. While the former cannot be insured by private firms, the impact of the latter on inequality is very much shaped in the labor market. In particular, if firms can observe workers' ability (say, their education level) and their realized output but not their labor effort, they design a contract that provides only partial insurance against output risk in order to give

[1] Abrahám, Alvarez-Parra, and Forstner (2016) study the implications of dynamic performance-pay contracts for wage inequality and wage dynamics.

incentives to workers to exert the optimal labor effort.[2] Importantly, the amount of earnings risk—the size of the bonus, or the reliance on stock options—is endogenous to tax policy. In our model, the government uses income taxes to redistribute between workers with different ex-ante abilities, taking into account their effect on insurance within firms.

We first take a positive standpoint and evaluate the incidence of nonlinear taxes on the design of performance pay. In our baseline model, based on Rogerson (1985), the output of each worker is stochastic and binary: either high or low. Workers can increase the probability of high output by exerting more (unobservable and costly) effort. Firms offer a bonus when realized output is high in order to motivate workers to exert the right amount of labor effort. Since workers are risk-averse, the optimal contract strikes a balance between such incentives for effort and insurance against output risk—providing better insurance is a less costly way for the firm to give workers their reservation utility. Workers' earnings, including the bonus, are subject to an income tax. Both the size of the bonus and the frequency of receiving a bonus are endogenous to the tax system.

We first show in a parametric version of the model that an increase in the rate of tax progressivity causes two offsetting effects on the labor contract. First, it leads to a standard crowding-out of the private insurance provided by the firm, that is, a one-for-one spread of the pre-tax earnings distribution. In response to a change in social insurance, firms adjust earnings endogenously so that the workers' incentives for effort and participation remain unchanged despite the reform.[3] Second, it also generates a crowding-in of the earnings distribution. This is due to the fact that, as in standard models of taxation—e.g., Mirrlees (1971)—higher marginal tax rates reduce optimal labor effort. In our setting, this is because higher tax rates on high earnings make high-powered incentives more costly to provide for the firm.[4] In turn, to elicit

[2]There is strong field-experimental evidence of moral hazard frictions in the workplace (see, e.g., the review by Lazear 2018). Performance pay can also be micro-founded by models of adverse selection, in which incentive pay is offered to attract workers with higher unobserved ability. Lazear (2000), Leaver, Ozier, Serneels, and Zeitlin (2021) and Brown and Andrabi (2020) find empirical support for both moral hazard and adverse selection, the former accounting for at least a half of the overall effect of performance pay on productivity.

[3]Theoretically, the crowd-out of private insurance has been shown to severely limit the ability of governments to provide social insurance—see Attanasio and Rıos-Rull (2000), Golosov and Tsyvinski (2007) and Krueger and Perri (2011). Empirically, crowd-out has been observed in health insurance (Cutler and Gruber 1996a, 1996b; Schoeni 2002) and unemployment insurance (Cullen and Gruber 2000).

[4]This argument echoes the wage-cum-labor demand effect of Lehmann, Parmentier, and Van der

this lower effort level in the presence of moral hazard, the firm lowers the sensitivity of pre-tax earnings to performance—that is, it compresses the earnings distribution. This crowd-in mechanism counteracts the usual crowd-out.

Our first main insight is that, regardless of the size of the labor effort elasticity, these two effects are of the same order of magnitude. Taken separately these effects are both significant, but summing them implies that taxes barely affect the sensitivity of (pre-tax) pay to performance. This finding may help explain why empirical studies of the impact of income taxes on the structure of performance-pay contracts often fail to find significant crowding-out; see Rose and Wolfram (2002) and Frydman and Molloy (2011). This insight remains robust to alternative forms of performance pay: We derive the same result when optimal contracts are a linear function of performance such as piece rates or commissions (Holmstrom and Milgrom 1987), convex such as stock options (Edmans and Gabaix 2011), or when incentives are provided over time (Edmans, Gabaix, Sadzik, and Sannikov 2012; Sannikov 2008). The general principle that makes our result robust to different forms of performance pay is that the local incentive constraints in all of these models of moral hazard share a common structure: namely, they imply that the sensitivity of earnings to performance is proportional to the marginal disutility of effort. As a result, in order to elicit a given increase in labor effort, the firm must raise the pass-through of output risk proportionately to the *inverse* of the (Frisch) labor effort elasticity. Therefore, the crowding-in of private insurance, equal to the product of the elasticity of labor effort and the change in performance-sensitivity required to elicit the desired effort adjustment—the inverse Frisch elasticity—is large and robust to the size of the labor effort elasticity.

The second part of our theoretical analysis takes a normative standpoint and characterizes optimal taxes in both our parametric and general environments. Even though taxes have little effect on the contract, because of offsetting crowd-in and crowd-out effects, the endogeneity of private insurance can potentially have large effects on welfare. Indeed, recall that the crowd-in channel operates via optimal labor effort decisions. By an envelope theorem, the resulting effect on the worker's utility is at most second-order. By contrast, the crowding-out of private insurance has first-order welfare consequences. Consequently, although the amount of within-firm insurance appears insensitive to tax policy, the worker's utility reacts *as if* private

Linden (2011), whereby tax progressivity tends to make high pre-tax wages less attractive (which, in their setting, leads to lower unemployment).

insurance was fully crowded out.

We derive three sets of results. First, we characterize the optimal rate of progressivity of the overall tax schedule. The endogeneity of performance-pay contracts affects optimal redistribution mainly via the welfare consequences of crowd-out. Greater progressivity triggers a crowd-out of within-firm insurance which offsets any gains from social insurance. This eliminates the role of the income tax as an insurance device. In addition, the crowding-out generates an additional adverse effect on welfare by making tax cuts less accurately targeted towards high-marginal-utility agents than in standard models. To see this, suppose that the tax burden decreases uniformly by a constant amount. The consumption gain, however, will not be uniform across income levels, as firms need to adjust gross earnings to preserve effort incentives. We show that the effective consumption gains are distributed regressively: High-performers of a given ability type—whose marginal utility of consumption is relatively low—receive a larger share of the tax cut than do low-performers. This dampens the welfare benefits of redistribution via progressive taxation. These two negative welfare effects of tax progressivity due to the crowd-out of private insurance are accompanied by new fiscal externalities: To the extent that the crowd-out and crowd-in do not offset each other perfectly, mean-preserving changes in earnings risk affect tax revenue when the income tax is non-linear. Accounting for all of these effects, we show analytically that the optimal tax progressivity in the presence of performance pay is strictly lower than in the model with exogenous earnings risk.

Second, we characterize the optimal tax rate on top income earners in the general model with arbitrarily non-linear taxes. We start with the case where tax rates apply to total (fixed plus variable) earnings. Our formula generalizes that of Saez (2001) by accounting for the fiscal and welfare effects of crowd-out and crowd-in, in addition to standard Pareto coefficients and labor effort elasticities. In particular, a higher top tax rate creates a fiscal spillover typically ignored by standard models: It affects the base earnings of agents who would have been in the top tax bracket had they earned a bonus. In contrast to Chetty and Saez (2010), whose approach is based purely on reduced-form sufficient statistics, we provide analytical expressions for all of the relevant parameters in our formulas—including the crowd-in and crowd-out of private insurance and the social marginal welfare weights.

Third, we ask whether there would be gains from taxing bonuses and base earnings separately. We first show that it is optimal to tax bonuses strictly less than base pay.

In particular, starting from a joint tax on total earnings, we construct a reform that (i) lowers the tax rate on bonuses and raises the tax rate on base pay, (ii) does not affect the expected utility of any agent, and (iii) generates more tax revenue—a Pareto improvement—by fostering labor effort. To understand this result, recall that the bonus is a constrained efficient instrument used by firms to overcome agency frictions in the labor market. Taxing bonuses erodes the firm's ability to provide incentives to workers and is thus highly distortionary. Consequently, it is optimal to shift the tax burden from bonuses to base pay. Finally, we establish an analytical formula for the optimal top tax rate on bonuses, for a given tax schedule on base earnings. As in the case of the joint taxation of overall earnings, the optimal bonus tax rate depends on the fiscal externalities and welfare effects of the crowd-out and crowd-in.

To complement this theoretical analysis, we calibrate the model to U.S. data. Our model features both performance-pay and fixed-pay jobs and accurately matches the earnings distribution, including the incidence of performance-pay jobs by income quartiles. We first document that, in line with our theoretical findings, changes in tax progressivity have very little impact on the pay-performance sensitivity of labor contracts. Quantitatively, the crowd-in effect offsets more than 90 percent of the crowd-out in terms of the bonus-to-base-pay ratio, and more than 100 percent in terms of the variance of log earnings. Thus, while higher progressivity does lead to lower labor effort and lower mean earnings—with implied elasticities that match empirical evidence—it leaves earnings risk largely unaffected.

We then proceed to computing the optimal tax policy. Our robust finding is that ignoring the endogeneity of earnings risk due to performance pay when setting tax policy has a small welfare cost. In other words, applying standard tax formulas allows the policymaker to reap the vast majority of the welfare gains from the optimal tax reform. This result first holds when we optimize over the overall rate of tax progressivity, in which case the welfare loss from applying the standard—rather than optimal—tax formula is equivalent to a modest 0.29 percent fall in consumption. The result also holds when we focus on the top tax rates: Those implied by the benchmark formula of Saez (2001) fall within a 1 percentage point margin of the optimal ones.

Intuitively, since earnings risk is virtually insensitive to tax policy, the policymaker is not making a large mistake when calculating the fiscal effects of tax reforms using the standard formulas. In addition, the welfare effects of crowd-out, which are not accounted for by the standard formulas, are of limited magnitude. In the case of reforming the overall tax progressivity, this is because only half of the jobs in the U.S. feature performance pay—if all workers had performance-pay contracts, the welfare cost of ignoring endogenous earnings risk would increase more than fourfold. The welfare cost of crowd-out is also small in the case of the top tax rate, since top earners are typically associated with small marginal welfare weights in the social objective.

Finally, we calculate the optimal top bonus tax rate for a given tax schedule on base pay. Consistent with our theoretical findings, the optimal bonus tax is lower than the tax on base pay, with a difference of 20 percentage points or more for a utilitarian or Rawlsian government.

Literature Review. Several papers study optimal taxation with labor markets constrained by agency frictions. Golosov and Tsyvinski (2007) analyze a model in which firms employ ex ante identical workers who are subject to private productivity shocks and can engage in hidden asset trades. They show that tax reforms generate a large crowd-out of within-firm insurance, which in turn reduces the benefits of social insurance. The government optimally refrains from providing any insurance and, instead, uses a capital tax to correct the externality generated by hidden trades. Stantcheva (2014) focuses on the adverse selection model of the labor market of Miyazaki (1977) and shows that the upward hours distortions (rat race) due to employer screening allow the government to redistribute more than in the standard Mirrlees model. Scheuer (2013) studies the role of profit taxes in correcting the choice between payroll employment and entrepreneurship, which is distorted by adverse selection in the credit market. In contrast to these papers, we study labor markets that are constrained by moral hazard frictions.

Ferey, Haufler, and Perroni (2022) also focus on moral hazard and show that globalization in the presence of performance pay can reconcile the joint evolution of rising earnings inequality and falling optimal tax progressivity. The main difference between our framework and theirs lies in the labor effort adjustment margin: While we allow for a continuous effort choice, Ferey et al. (2022) assume that effort is binary, which limits individual effort responses and the crowd-in effect within firms that is at the core of our analysis.[5] Chetty and Saez (2010) derive a reduced-form sufficient

[5] In their framework, tax reforms still have a distortionary impact on effort in the aggregate by affecting the measure of firms that choose to incentivize effort via performance-pay contracts—an extensive margin response. By contrast, in our framework, taxes distort individual labor effort and within-firm insurance in already existing performance-pay jobs—an intensive margin response.

statistics formula for the optimal linear tax in the presence of linear private insurance contracts that can be subject to agency frictions. By contrast, we provide an explicit and tractable structural microfoundation for the equilibrium labor contracts. This allows us to characterize analytically the effects of nonlinear government policy on private insurance contracts via crowd-out and crowd-in responses, and derive explicit theoretical formulas for the tax incidence and optimal taxes in terms of underlying structural parameters. Kaplow (1991) studies optimal linear government policy in a different (disaster relief) setting that has endogenous private insurance constrained by moral hazard. He finds that the optimal level of social insurance is zero. This result holds in our setting as well, but in our model workers are also ex-ante heterogeneous, thus giving the government a redistributive role.

We contribute to the growing literature on the optimal taxation and regulation of bonuses. Besley and Ghatak (2013) study the role of bonus taxes in correcting the excessive risk taking in the financial sector, while Thanassoulis (2012) and Bénabou and Tirole (2016) show that bonus caps can restrict inefficiencies stemming from the competition for talent. In contrast to these papers, we focus on the basic model of moral hazard in which performance pay is constrained efficient, and uncover a force towards a *lower* tax rate on bonuses than on base earnings: This policy alleviates the labor effort distortions implied by income redistribution. Two related papers, Gietl and Haufler (2018) and Haufler and Nishimura (2022), study bonus tax competition in settings where top managers are internationally mobile, a feature that we ignore in this paper.

An important strand of papers studies income taxation in the presence of endogenous consumption insurance, which can take form of private insurance markets (Cremer and Pestieau 1996; Netzer and Scheuer 2007), asset trades (Ábrahám, Koehne, and Pavoni 2016; Chang and Park 2017; Park 2014) or informal exchanges in family networks (Attanasio and Rıos-Rull 2000; Heathcote, Storesletten, and Violante 2017; Krueger and Perri 2011; Raj 2019). In contrast to these papers, it is *pre-tax earnings* risk—rather than *consumption* risk—that is endogenous to policy in our model. This distinction matters since changes in earnings risk have a direct impact on tax revenue when the income tax is non-linear (as a result of Jensen's inequality), while changes in consumption risk with exogenous wages do not have a direct fiscal impact.[6]

[6]Naturally, there can be indirect effects from consumption insurance to tax revenue through precautionary labor supply, as in Netzer and Scheuer (2007).

Other papers endogenize earnings risk by focusing on human capital accumulation (Craig 2019; Findeisen and Sachs 2016; Kapička and Neira 2019; Makris and Pavan 2021; Stantcheva 2017), job search (Sleet and Yazici 2017), or wage randomization in response to excessive tax regressivity (Doligalski 2019).

Finally, our paper relates to the literature on redistributive taxation in environments with earnings uncertainty and moral hazard; see, e.g., Varian (1980) and, for models that also allow for ability differences, Eaton and Rosen (1980), Shourideh (2014), and Boadway and Sato (2015) (who also offer a more complete synthesis of the literature). In these papers, however, there is no layer of endogenous private insurance between workers and the government: Earnings risk is thus exogenous, and the government is the sole provider of insurance.

Outline of the Paper. The paper is organized as follows. In Section 1, we study a simple version of our model and characterize the optimal rate of tax progressivity. In Section 2, we study our general environment and characterize the optimal top tax rates on total earnings and on bonuses when they are taxed separately. The proofs of our results and various extensions are in the Appendix. In particular, we study alternative forms of performance pay in Appendix C.

1 Performance Pay and Tax Progressivity

In this section, we set up a simple version of our general model to derive some of the main insights most transparently. We study the incidence of tax progressivity on individual earnings and welfare in Section 1.2, characterize the optimal rate of progressivity in Section 1.3, and evaluate our findings quantitatively in Section 1.4. Finally, in Section 1.5, we show that our results apply to several alternative models of performance pay that are aimed at capturing different forms of variable compensation: bonuses, piece rates or commissions, stock options, and dynamic incentives.

1.1 Environment

There is a continuum of mass one of agents indexed by their exogenous ability $\theta \in \mathbb{R}_+$ that is distributed according to the c.d.f. $F(\theta)$ and density $f(\theta)$. Preferences over consumption c and labor effort ℓ are represented by the utility function $u(c)-h(\ell)$, where u and h are twice continuously differentiable, u is strictly concave, and h is strictly convex.

A worker with ability θ who provides effort ℓ can produce two levels of output y:

$$y = \begin{cases} \theta & \text{with probability } \pi(\ell) \\ 0 & \text{with probability } 1 - \pi(\ell) \end{cases}$$

where $\pi : \mathbb{R}_+ \to [0,1]$ is continuously differentiable and concave. Without loss of generality, and unless otherwise stated, we normalize units of effort so that $\pi(\ell) = \ell \in [0,1]$.[7]

Firms observe both the agent's ability θ and realized output y, but not their effort ℓ. As a consequence, the labor contract specifies earnings z as a function of observed performance. If output is low, the worker earns a base salary $\underline{z}$. If output is high, earnings take a larger value $\bar{z} \geq \underline{z}$. We define the bonus as the difference between the high-level pay and the base pay, $b = \bar{z} - \underline{z}$. We also introduce the bonus rate β, or the pass-through of output risk to log-earnings, defined such that $\bar{z} = e^{\beta}\underline{z}$.[8]

The government levies non-linear income taxes. A worker with earnings z consumes their after-tax earnings $c = z - T(z) \equiv R(z)$, where $T : \mathbb{R} \to \mathbb{R}$ is the tax schedule and R is the retention function. Throughout the paper, we denote the utility over pre-tax earnings by $v(z) \equiv u(R(z))$. In this section, we restrict preferences and taxes to the following functional forms.

Assumption 1 *The utility of consumption is logarithmic, $u(c) = \log c$. The tax schedule has a constant rate of progressivity (CRP):*[9] *There exist $\tau \in \mathbb{R}$ and $p \in (-\infty, 1)$ such that $R(z) = \frac{1-\tau}{1-p} z^{1-p}$.*

A firm that hires a worker with ability θ takes the tax schedule and the worker's reservation value $U(\theta)$ as given. It chooses the earnings contract $\{\underline{z}(\theta), \bar{z}(\theta)\}$ to

[7] The disutility of effort must then be re-normalized as $\tilde{h} \equiv h \circ \pi^{-1}$. For clarity, we keep the notation $h(\ell)$, except in our calibration exercises that require positing functional forms for $h(\cdot)$ and $\pi(\cdot)$.

[8] Intuitively, for a small bonus b, β coincides with the ratio of bonus to base pay $b/\underline{z}$.

[9] The CRP tax code is a good approximation of the U.S. tax system, see for instance, Heathcote et al. (2017). The rate of progressivity p is equal to (minus) the elasticity of the retention rate $R'(z)$ with respect to income z. Alternatively, $1-p$ is equal to the ratio of marginal retained income $R'(z)$ to average retained income $R(z)/z$.

maximize its expected profit:

$$\Pi(\theta) \;=\; \max_{\underline{z},\bar{z}} \; \ell\theta - \mathbb{E}[z], \tag{1}$$

where $\mathbb{E}[z] = (1-\ell)\underline{z} + \ell\bar{z}$,[10] subject to the following constraints.

First, the incentive constraint requires that the worker's effort level maximizes their expected utility over all possible effort choices, taking the labor contract as given:

$$\ell \;=\; \arg\max_{l\in[0,1]} \; (1-l)v(\underline{z}) + lv(\bar{z}) - h(l). \tag{2}$$

Second, the participation constraint requires that the worker's expected utility is at least as large as their reservation value:

$$\mathbb{E}[v(z)] - h(\ell) \;\geq\; U(\theta), \tag{3}$$

where $\mathbb{E}[v(z)] = (1-\ell)v(\underline{z}) + \ell v(\bar{z})$.

To pin down the workers' reservation value $U(\theta)$, we assume that there is free-entry of firms in the labor market for workers of type θ. Thus, in equilibrium, expected profits are equal to zero:

$$\Pi(\theta) \;=\; 0. \tag{4}$$

Moral Hazard and the Bonus Rate. The incentive constraint (2) can be simplified by taking the first-order condition of the worker's maximization problem:

$$h'(\ell) \;=\; v(\bar{z}) - v(\underline{z}). \tag{5}$$

We can show that this first-order condition is necessary and sufficient to ensure global incentive compatibility as long as the equilibrium level of effort is interior.[11] Using Assumption 1, we can rewrite this expression as follows:

$$\beta \;=\; \frac{h'(\ell)}{1-p}. \tag{6}$$

[10]Throughout the paper, unless explicitly stated the operator $\mathbb{E}$ denotes the expectation of a random variable conditional on a given worker θ, as opposed to a population average. We also let $\mathbb{V}(\cdot \mid \theta)$ denote the variance conditional on productivity θ.

[11]See Lemma 4, which proves this result in the general setting of Section 2.

This equation plays an important role in our analysis. Intuitively, inducing a worker to provide a given level of costly effort ℓ requires a larger reward for performance (i.e., a higher bonus rate β) if the marginal disutility $h'(\ell)$ is higher. Since h is convex, this implies that the sensitivity of earnings to performance β is strictly increasing in labor effort ℓ. This captures the key moral hazard insight that eliciting higher effort from the worker requires raising their exposure to output risk.

Optimal Contract. The following proposition characterizes the equilibrium labor contract between a firm and a worker with ability θ.

Proposition 1 *Suppose that Assumption 1 holds. Effort ℓ is then independent of θ, so that the bonus rate β is constant and given by (6). Equilibrium earnings are given by*

$$\underline{z}(\theta) \;=\; \frac{1}{1+\ell(e^{\beta}-1)}\ell\theta \quad \text{and} \quad \bar{z}(\theta) \;=\; \frac{e^{\beta}}{1+\ell(e^{\beta}-1)}\ell\theta. \tag{7}$$

Optimal labor effort satisfies

$$\theta \;=\; b + b\,\ell(1-\ell)\frac{h''(\ell)}{1-p}. \tag{8}$$

Expected utility is given by

$$U(\theta) \;=\; v(\ell\theta) - h(\ell) - (1-p)\left\{u(\mathbb{E}z) - \mathbb{E}[u(z)]\right\}, \tag{9}$$

with $u(\mathbb{E}z) - \mathbb{E}[u(z)] = \log(1+\ell(e^{\beta}-1)) - \beta\ell$.

Full-Insurance Benchmark. To interpret Proposition 1, consider first the benchmark setting where the firm can perfectly monitor the worker's effort, so that the incentive constraint can be ignored. In this case, the firm provides full insurance against output risk: Workers with ability θ who provide effort ℓ earn their expected output regardless of their performance, $\underline{z} = \bar{z} = \theta\ell$. Their utility is then equal to $\mathcal{U}(\theta) \equiv v(\theta\ell) - h(\ell)$. Note that this setting is equivalent to the standard Mirrlees (1971) model, where the relationship $z = \theta\ell$ between labor effort and income also holds.

Equilibrium Earnings. With moral hazard frictions, the previous relationship continues to hold on average: The free-entry condition (4) imposes that workers' expected earnings $\mathbb{E}[z]$ remain equal to their expected output $\ell\theta$. However, providing effort incentives implies that realized earnings must now be dispersed around their mean. The firm provides only partial insurance against output risk: we have $0 < \underline{z} < \theta\ell < \bar{z} < \theta$. At the heart of our paper lies the observation that the optimal degree of within-firm insurance—captured by the parameter β in equation (6)—is *endogenous* to the tax system. We analyze this dependence in the next section.

Labor Effort. To interpret the optimality condition for effort (8), suppose that the firm aims to elicit marginally higher effort from the worker. The expected output gain, on the left hand side, is θ. Keeping the earnings structure $\underline{z}, \bar{z}, b$ fixed, inducing higher effort costs b to the firm, since the worker receives the bonus more frequently.[12] This is the first term on the right hand side of (8). In addition, the firm needs to raise the earnings spread in order to incentivize the worker to actually exert this extra effort. By equation (6), the bonus rate must increase proportionally to the rise in the marginal disutility of effort, $h''(\ell)$. This creates an additional cost for the firm—the *marginal cost of incentives* (MCI)—given by the second term on the right hand side of (8): Intuitively, since workers are risk averse, exposing them to more earnings risk while keeping the participation constraint satisfied is costly because it requires increasing mean earnings.

Expected Utility. Equation (9) decomposes the worker's expected utility into three components. The first is the utility they would attain under full insurance, $\mathcal{U}(\theta) = v(\theta\ell) - h(\ell)$. Second, the incompleteness of private insurance makes the risk-averse worker worse off: The utility loss associated with a given earnings lottery z is equal to the utility difference between expected earnings $\mathbb{E}[z]$ and the certainty equivalent.[13] This difference is the term in curly brackets in expression (9). Third, this utility loss is weighted by $(1-p)$: All else being equal, a higher level of tax progressivity reduces the variance of disposable income that the consumer faces, which dampens the welfare cost of earnings uncertainty. Thus, keeping earnings risk and the level of effort fixed, higher social insurance raises welfare.

[12]Note that, by equation (5), this is just sufficient to compensate the worker for the higher effort level and keep the participation constraint satisfied.

[13]Recall that the certainty equivalent z^{CE} is defined by $u(z^{CE}) = \mathbb{E}[u(z)]$.

1.2 Tax Incidence

In this section, we explore how the characteristics of the equilibrium contract: labor effort ℓ, mean earnings $\mathbb{E}z$, earnings risk β, and expected utility $U(\theta)$, respond to a change in tax progressivity p.

Effect on Average Earnings and Labor Effort. It is immediate that the effect of taxes on mean earnings $\mathbb{E}z = \ell\theta$ is given by:

$$\frac{\partial \log \mathbb{E}z}{\partial \log(1-p)} = \frac{\partial \log \ell}{\partial \log(1-p)} \equiv \varepsilon_{\ell,1-p},$$

where $\varepsilon_{\ell,1-p}$ denotes the elasticity of labor effort (or of the frequency of receiving a bonus) with respect to (one minus) the rate of tax progressivity. In the full-insurance setting, we have $\varepsilon_{\ell,1-p} = \frac{\varepsilon_\ell^F}{1+\varepsilon_\ell^F}$ where $\varepsilon_\ell^F \equiv \frac{h'(\ell)}{\ell h''(\ell)}$ is the Frisch elasticity of labor supply.[14] The following lemma characterizes $\varepsilon_{\ell,1-p}$ with endogenously incomplete insurance.

Lemma 1 *Higher tax progressivity reduces optimal labor effort, i.e., $\varepsilon_{\ell,1-p} > 0$. Suppose that the disutility of effort is isoelastic, so that the Frisch elasticity ε_ℓ^F is constant. Then $\varepsilon_{\ell,1-p} > \frac{\varepsilon_\ell^F}{1+\varepsilon_\ell^F}$.*

Lemma 1 shows that, as in standard models of taxation, tax progressivity disincentivizes labor effort. This behavioral response plays a key role in our analysis. Intuitively, greater progressivity makes high-powered incentives more costly to provide. Indeed, when high levels of income are taxed away more heavily, eliciting marginally higher effort requires a larger increase in the dispersion of pre-tax earnings and, therefore, a larger cost for the firm. Formally, both the bonus b and the MCI term in equation (8) are increasing in p.[15] Furthermore, Lemma 1 shows that raising tax progressivity causes a stronger negative response of effort than in the benchmark setting without exposure to output risk. Intuitively, this is because the MCI term is absent when effort is observable.

[14] This elasticity is lower than the Frisch elasticity because of the income effect of taxes on labor supply.

[15] Note that it would not be possible to deliver incentives by lowering base earnings $\underline{z}$ rather than raising the bonus, since this would violate the worker's participation constraint.

Effect on Earnings Risk. We infer from the local incentive constraint (6) that taxes affect earnings risk β through two channels that work in opposite directions. First, β is inversely proportional to $1-p$, so that, *ceteris paribus*, a higher rate of progressivity leads to a steeper pre-tax earnings schedule. This direct effect of taxes is a standard *crowding-out* of private insurance by social insurance. Intuitively, by raising tax progressivity, the government compresses the disposable income distribution and, therefore, reduces the worker's exposure to output risk. The firm responds by spreading out pre-tax earnings (i.e., raising β) in order to preserve the worker's incentives for effort. The elasticity of output risk to tax progressivity is given by

$$\varepsilon_{\beta,1-p} \;\equiv\; \frac{\partial \log \beta}{\partial \log(1-p)} \;=\; -1,$$

which means that, absent effort responses, the firm adjusts the contract so as to keep consumption insurance—the variance of log-consumption—fixed.

Second, β is proportional to $h'(\ell)$. As a result, tax progressivity affects earnings risk indirectly via the endogenous choice of labor effort. Recall that a higher rate of progressivity reduces the optimal labor effort. The strength of this effect is captured by the elasticity $\varepsilon_{\ell,1-p} > 0$. Now, the firm implements this lower level of effort by reducing the worker's exposure to risk β: that is, by compressing the pre-tax earnings distribution. This effect is a *crowding-in* of private insurance by social insurance. The elasticity of earnings risk to desired labor effort is given by

$$\varepsilon_{\beta,\ell} \;\equiv\; \frac{\partial \log \beta}{\partial \log \ell} \;=\; \frac{\ell h''(\ell)}{h'(\ell)} = \frac{1}{\varepsilon_\ell^F},$$

where $\varepsilon_\ell^F > 0$ is the Frisch elasticity.

Lemma 2 *The total impact of tax progressivity on earnings risk is given by*

$$\frac{d \log \beta}{d \log(1-p)} \;=\; \varepsilon_{\beta,1-p} + \varepsilon_{\beta,\ell} \cdot \varepsilon_{\ell,1-p} \;=\; -1 + \frac{\varepsilon_{\ell,1-p}}{\varepsilon_\ell^F}.$$

It is negative (net crowding-out) if $\varepsilon_{\ell,1-p} < \varepsilon_\ell^F$ and positive (net crowding-in) otherwise. By Lemma 1, a lower bound for the net effect is given by $\frac{d\log\beta}{d\log(1-p)} > -\frac{\varepsilon_\ell^F}{1+\varepsilon_\ell^F}$.

Crowd-Out and Crowd-In Approximately Offset Each Other. We now argue that these two counteracting forces—crowd-out and crowd-in—are of the same order of magnitude, so that tax progressivity has only a modest impact on the amount of within-firm insurance. The key insight is that the strength of the moral hazard friction is inversely proportional to the Frisch elasticity of labor supply, $\varepsilon_{\beta,\ell} = 1/\varepsilon_\ell^F$; This is an immediate consequence of equation (6). If the elasticity of labor effort $\varepsilon_{\ell,1-p}$ is roughly equal to the Frisch elasticity, then the crowd-in effect is equal to $\frac{1}{\varepsilon_\ell^F}\varepsilon_{\ell,1-p} \approx 1$: that is, about the same size (in absolute value) as the direct crowd-out $\varepsilon_{\beta,1-p} = -1$.

Importantly, this insight is robust to small values of the labor effort elasticity. To see this, suppose that effort hardly diminishes in response to an increase in tax progressivity, which happens when $\varepsilon_\ell^F \to 0$. Lemma 2 shows that the crowd-in effect is bounded from below by $1/(1+\varepsilon_\ell^F)$, which converges to 1 as $\varepsilon_\ell^F \to 0$. Thus, in this case, the crowding-in of private insurance offsets (at least as much as) the entire crowd-out. Intuitively, it is precisely because of the inelastic behavior that the firm must dramatically reduce the worker's exposure to output risk in order to implement the desired (tiny) reduction in labor effort: formally, $\varepsilon_{\beta,\ell} = 1/\varepsilon_\ell^F \to \infty$. As a result, the product of elasticities $\varepsilon_{\beta,\ell}\,\varepsilon_{\ell,1-p}$ does not vanish in the limit: The crowd-in effect remains significant even when effort is almost inelastic.

This discussion is correct as long as the labor effort elasticity is indeed approximately equal to the Frisch elasticity. In practice, when the Frisch elasticity is bounded away from zero, this need not be exactly the case. Suppose that the disutility of (unnormalized) labor effort is isoelastic on $\mathbb{R}_+$, $h(\ell) = \ell^{1+1/e}$, with an empirically realistic value of $e = 0.5$. The relevant Frisch elasticity ε_ℓ^F in Lemma 2, however, is lower than e, since it is that of the function $h \circ \pi^{-1}$, where π is concave (see footnote 7). If $\pi(\ell) = \sqrt{\ell}$, for instance, we get $\varepsilon_\ell^F = e/(2+e) = 0.2$. In this case, we find a lower bound for the crowd-in effect $\varepsilon_{\beta,\ell}\,\varepsilon_{\ell,1-p} > \frac{1}{1+0.2} = 0.83$. That is, the earnings risk adjustment due to labor effort responses offsets at least 83 percent of the crowd-out of private insurance caused by tax progressivity. Even when we raise the Frisch elasticity to $e = 1$, the high end of empirical estimates, the crowd-in effect offsets at least 75 percent of the crowd-out.

Finally, the above results are based on the assumption that $u(c) = \log(c)$, which implies a particular strength of the income effect on labor effort and a particular degree of risk aversion. Nevertheless, our prediction—that the crowd-in offsets most of the crowd-out—continues to hold when there is no income effect and workers are arbitrarily risk-averse. In Appendix C.1, we study the moral-hazard framework of

Holmstrom and Milgrom (1987) with preferences $u(c, \ell) = -\frac{1}{\gamma} e^{-\gamma(c-h(\ell))}$, which imply no income effect and an arbitrary absolute risk aversion controlled by the coefficient $\gamma \geq 0$. Assuming that the Frisch elasticity is constant, we show that the rate at which the crowd-in offsets the crowd-out is decreasing in risk aversion γ. When workers are risk-neutral, the crowd-in offsets the crowd-out exactly. Even in the limit where the risk aversion coefficient goes to infinity, the two effects remain of comparable magnitude. For instance, assuming a Frisch elasticity of 0.5, the crowd-in always offsets at least two-thirds of the crowd-out.

Effect on Expected Utility. In the full-insurance benchmark, the envelope theorem implies that expected utility $\mathcal{U}(\theta) \equiv v(\theta\ell) - h(\ell) = \log \frac{1-\tau}{1-p} + (1-p)\log(\ell\theta) - h(\ell)$ changes in response to an increase in progressivity by $\frac{d\mathcal{U}(\theta)}{dp} = \frac{1}{1-p} - \log(\mathbb{E}z)$. In our setting, we obtain the following result.

Lemma 3 *The impact of tax progressivity on expected utility is given by*

$$\frac{dU(\theta)}{dp} = \frac{d\mathcal{U}(\theta)}{dp} + [\log(\mathbb{E}z) - \mathbb{E}(\log z)] + \frac{b}{\beta \mathbb{E}z} \mathbb{V}(\log z \mid \theta)\ \varepsilon_{\beta,1-p}, \tag{10}$$

where $\mathbb{V}(\log z \mid \theta) = \beta^2 \ell(1-\ell)$ is the variance of log-earnings conditional on θ. Note in particular that, to a first order, the worker's expected utility is unaffected by changes in labor effort (envelope theorem).

Higher progressivity p raises expected utility when earnings are uncertain by reducing the consumption spread that workers face. This is captured by the second term (in square brackets) on the right-hand side of (10). In addition, higher progressivity raises the dispersion of pre-tax earnings via the crowd-out elasticity $\varepsilon_{\beta,1-p} = -1$. This causes a loss in expected utility proportional to the variance of log-earnings $\mathbb{V}(\log z \mid \theta)$.

Importantly, note that the crowding-*in* of private insurance, $\varepsilon_{\beta,\ell}\varepsilon_{\ell,1-p}$, does not appear in formula (10). This is because this effect operates via optimal labor effort choices. The envelope theorem implies that its impact on welfare is only of second-order.[16] Therefore, while the crowd-out and crowd-in have offsetting effects on the

[16]More precisely: Consider an (equivalent) dual formulation of the firm's problem, which consists of maximizing the worker's expected utility subject to making non-negative profits. The envelope theorem applied to this problem implies that changes in labor effort do not have first-order effects on expected utility.

structure of pre-tax earnings, only the former affects the worker's utility. In other words, observing that the amount of pre-tax insurance embedded in the labor contract is insensitive to taxes does *not* imply that the endogeneity of private insurance is negligible for welfare.

1.3 Optimal Taxation

The government chooses the tax schedule, or the retention function R, to maximize a weighted utilitarian social welfare function:

$$\int_0^\infty \alpha(\theta)U(\theta)f(\theta)d\theta \tag{11}$$

subject to a budget constraint:

$$\int_0^\infty \mathbb{E}[z(\theta) - R(z(\theta))]f(\theta)d\theta \ \geq \ G, \tag{12}$$

where the Pareto weights $\alpha(\theta) \geq 0$ satisfy $\int_0^\infty \alpha(\theta)f(\theta)d\theta = 1$, and where $G \geq 0$ is an exogenous expenditure requirement. We denote by g the ratio of public spending to aggregate output.

Optimal Tax Progressivity. Recall that, in this section, we restrict the tax schedule to the CRP class. Theorem 1 characterizes the optimal rate of tax progressivity p under the additional assumptions that the distribution of ability types is lognormal, and that the social welfare objective is utilitarian.[17]

Theorem 1 *Suppose that* $\log\theta \sim \mathcal{N}(\mu_\theta, \sigma_\theta^2)$ *and that* $\alpha(\theta) = 1$ *for all* θ*. The optimal rate of progressivity satisfies*

$$\frac{p}{(1-p)^2} = \frac{\sigma_\theta^2 + \kappa_1(1+\varepsilon_{\beta,1-p})\mathbb{V}(\log z \mid \theta)}{\left[1 + \frac{g}{(1-g)p} + \ell\kappa_3\right]\varepsilon_{\ell,1-p} + \kappa_2\,\varepsilon_{\beta,\ell}\,\varepsilon_{\ell,1-p}\mathbb{V}(\log z \mid \theta)}, \tag{13}$$

with $\kappa_1 = \frac{1}{\beta(1-p)}\frac{\bar{c}-\underline{c}}{\mathbb{E}c}$, $\kappa_2 = \frac{1-p}{\beta p}\left(\frac{b}{\mathbb{E}z} - \frac{\bar{c}-\underline{c}}{\mathbb{E}c}\right)$, *and* $\kappa_3 = \frac{1-p}{p}\left(\frac{b}{\mathbb{E}z} - \frac{1}{1-p}\frac{\bar{c}-\underline{c}}{\mathbb{E}c}\right)$, *where* $\underline{c}, \bar{c}$ *denote after-tax earnings in the low- and high-performance states. We have* $\kappa_1 > 0$, $\kappa_2 > 0$, *and* $\kappa_3 = 0$ *if* $p = 0$.

[17] In the Appendix, we generalize formula (13) to social welfare weights given by $\alpha(\theta) \propto \theta^{-a}$, where $a \geq 0$.

To understand Theorem 1, it is helpful to decompose it by considering first the full-insurance benchmark, then the case of exogenously incomplete private insurance, and finally the general setting with endogenously incomplete private insurance.

A. Full-Insurance Benchmark

Suppose first that firms provide complete insurance against output risk, or equivalently that there is no output risk as in Mirrlees (1971). That is, $\underline{z} = \bar{z} = \ell\theta$. Formula (13) then reduces to

$$\frac{p}{(1-p)^2} = \frac{\sigma_\theta^2}{\left[1 + \frac{g}{(1-g)p}\right]\varepsilon_{\ell,1-p}}.$$

The optimal rate of progressivity is increasing in inequality, measured by the variance of the log-ability distribution σ_θ^2, and decreasing in the elasticity of labor effort $\varepsilon_{\ell,1-p}$, which captures the efficiency cost of distortionary taxation. Moreover, it is decreasing in the share of government expenditures in GDP g: This is because a marginal tax increase induces a larger deadweight loss if the tax burden is already large due to high spending needs.

B. Exogenously Incomplete Insurance

Suppose next that the firm provides incomplete insurance against output shocks, so that $\bar{z} > \underline{z}$, but that earnings risk $\beta > 0$ is exogenous. Thus, the elasticities of crowd-out and moral hazard $\varepsilon_{\beta,1-p}$ and $\varepsilon_{\beta,\ell}$ are both set equal to zero. Formula (13) then reads

$$\frac{p}{(1-p)^2} = \frac{\sigma_\theta^2 + \kappa_1 \mathbb{V}(\log z \mid \theta)}{\left[1 + \frac{g}{(1-g)p} + \ell\kappa_3\right]\varepsilon_{\ell,1-p}}. \tag{14}$$

There are two differences to the full-insurance benchmark. First, the dispersion of earnings in the population is now mechanically larger than that of ability types θ. Tax progressivity thus plays two roles: redistribution across ex-ante ability differences (measured by the variance σ_θ^2) and social insurance against ex-post earnings risk (measured by the conditional variance of pre-tax earnings $\mathbb{V}(\log z \mid \theta)$). Up to a second-order as $\beta \to 0$, the numerator of (14) equals the total variance of log-earnings in the population, $\sigma_\theta^2 + \mathbb{V}(\log z \mid \theta)$. That is, optimal progressivity is an increasing function of overall earnings inequality, regardless of whether it is driven by innate ability differences or idiosyncratic performance shocks.

Second, the labor effort distortions in the denominator of (14) are also modified. In contrast to the full-insurance benchmark, a change in taxes affects not only the income levels $\underline{z}, \bar{z}$ on the intensive margin, but also triggers a response on the frequency (or extensive) margin by altering the probability ℓ with which the high income level $\bar{z}$, and hence the high tax payment, occurs. The former behavioral responses affect government revenue proportionally to the income-weighted marginal tax rates $\mathbb{E}[T'(z)z]\varepsilon_{\ell,1-p}$, while the latter affects revenue proportionally to the gap in total tax payments between the high- and low-performance states $\ell(T(\bar{z}) - T(\underline{z}))\varepsilon_{\ell,1-p}$. Accounting for these frequency responses gives rise to the additional term $\ell\kappa_3$ in the optimum tax formula.[18]

C. Endogenously Incomplete Insurance

We finally analyze the general case where earnings risk β is endogenous to taxes. The optimal progressivity formula (13) is modified by two novel terms. These account for fiscal externalities (in the denominator) and welfare effects (in the numerator) that we describe in turn.

Fiscal Externalities from Crowd-Out and Crowd-In. If the tax schedule is progressive, a spread of the pre-tax earnings distribution caused by the crowding-out analyzed in Section 1.2 generates a positive fiscal externality, i.e., a first-order gain in government revenue. Conversely, an earnings compression due to the crowding-in of private insurance induces a negative fiscal externality. These are consequences of Jensen's inequality: A progressive tax code generates more revenue for the government if earnings are more volatile, keeping their mean constant.

Formally, the crowd-out $\varepsilon_{\beta,1-p} < 0$ and crowd-in $\varepsilon_{\beta,\ell}\,\varepsilon_{\ell,1-p} > 0$ affect government revenue by $-\frac{p}{(1-p)^2}\kappa_2(\varepsilon_{\beta,1-p} + \varepsilon_{\beta,\ell}\,\varepsilon_{\ell,1-p})\mathbb{V}(\log z \mid \theta)$. If the tax schedule is linear ($p = 0$), these fiscal externalities are equal to zero. If the tax schedule is progressive ($p > 0$), the crowd-out (resp., crowd-in) raises (resp., lowers) government revenue and, hence, the optimal level of taxes. Note finally that the denominator of formula (13) features only the negative externality from crowd-*in*. This is because, as we

[18] When $p = 0$, we have $\kappa_3 = 0$. Recall that the fiscal impact of the intensive margin response depends on the marginal tax rate, while that of the frequency margin response depends on the average tax rate. Since the marginal and average tax rates are equal when $p = 0$, there is no need to correct the efficiency cost of taxes in this case.

argue next, the positive fiscal externality caused by the crowd-out is (more than) compensated for by its negative welfare impact.

Welfare Effects of Crowd-Out and Envelope Theorem. The crowd-out of private insurance affects the social welfare objective (11) by $\left[\kappa_1 + \frac{p}{(1-p)^2}\kappa_2\right]\varepsilon_{\beta,1-p}\,\mathbb{V}(\log z \mid \theta)$. Thus, since $\varepsilon_{\beta,1-p} = -1$, it exactly offsets both the insurance benefits of tax progressivity and the positive fiscal externality discussed in the previous paragraphs.

The fact that the crowd-out of private insurance offsets the welfare benefits of social insurance ($\kappa_1\mathbb{V}(\log z \mid \theta)$) is straightforward. Any attempt by the government to compress the distribution of disposable income leads the firm to raise the dispersion of pre-tax earnings one-for-one in order to preserve the worker's effort incentives.[19] Thus, in contrast to the case of exogenous private insurance, the government should not provide any social insurance against performance shocks: The numerator of (13) reduces to the benefits of insuring exogenous ability disparities θ, measured by the variance σ_θ^2.

The additional negative welfare effect of the crowd-out is more subtle. It is due to the fact that tax cuts lead to endogenous changes in the pay structure that render them less precisely targeted than in the standard framework—we elaborate on this point in more detail in Section 2. To build intuition, suppose that the tax liabilities at the base pay $T(\underline{z})$ and at the high-level pay $T(\bar{z})$ are lowered by the same amount. In Section 2.2 we show that such a reform increases the worker's reservation value $U(\theta)$ by (say) $\Delta U > 0$. The incentive constraint (5) implies that, to maintain effort ℓ, the ex-post utility of both low- and high-performers must increase by the same amount ΔU.[20] Absent changes in gross earnings, however, the (uniform) tax cut would not by itself lead to a uniform rise in ex-post utility, since the marginal utility of consumption is strictly decreasing. Consequently, the firm must raise the bonus and, to ensure that profits remain non-negative, lower the base pay. This implies that high-performers capture a disproportionately large share of the tax cut. This regressive distribution of rents—away from individuals whose marginal utility of consumption is the highest—further reduces the welfare benefits of redistribution via progressive taxes.

[19]More precisely, recall that the firm does not actually keep labor effort unchanged. But by the envelope theorem, the welfare consequences of the corresponding crowding-in are second-order.

[20]This is a standard consequence of the separability of the utility function; see, e.g., Golosov, Kocherlakota, and Tsyvinski (2003). We derive this property formally in equations (16) and (17) below.

Finally, note that these welfare losses due to the crowding-out are not counteracted by corresponding gains from the crowding-*in* responses. As explained in the discussion following Lemma 3, this is a consequence of the envelope theorem: The crowding-in of private insurance in response to an increase in tax progressivity operates via adjustments in labor effort that is chosen optimally, leading to an (at most) second-order impact on welfare. Hence, the observation that earnings risk is almost invariant to taxes (Section 1.2) does *not* imply that optimal tax design can safely ignore the endogeneity of private insurance.

Taking Stock: Endogenous Insurance Reduces Optimal Progressivity. In sum, there are two channels through which endogenous earnings risk matters for tax progressivity: (i) the negative fiscal externality due to crowd-in; and (ii) the negative welfare effect of crowd-out, which exactly offsets the benefits of social insurance.[21] As a result, the optimal rate of progressivity (13) is *strictly lower* than in a setting with exogenous earnings risk.

1.4 Quantitative Analysis

To evaluate our results quantitatively, we extend our baseline model with the following elements. A share s_{pp} of workers have a *performance-pay job*, and the remaining share $s_{fp} = 1 - s_{pp}$ have a *fixed-pay job*. Performance-pay jobs are subject to the moral hazard friction described above: output is stochastic, equal to the worker's ability θ with probability $\pi(\ell_{pp})$ and 0 otherwise, where ℓ_{pp} is their effort level. The optimal contract specifies earnings as a function of the output realization according to the bonus rate $\beta = h'(\ell_{pp})/(1-p)$. Fixed-pay jobs, by contrast, are not subject to agency frictions and guarantee a risk-free earnings level $\theta\ell_{fp}$, where ℓ_{fp} is the worker's effort level. In equilibrium, all fixed-pay workers exert the same effort.

We treat the job type of a worker as exogenous. In the data, the share of performance-pay jobs increases with earnings (see Lemieux et al. 2009, Gittleman and Pierce 2013, Grigsby, Hurst, and Yildirmaz 2019). To account for this fact in the model, we allow for a positive correlation between job type and ability. Specifically, we assume that ability is drawn from the job-type-specific Pareto-lognormal

[21] In the general environment of Section 2, the positive fiscal externality and the additional negative welfare effect of crowd-out no longer exactly cancel out, and both appear explicitly in the optimal tax formulas.

distribution (Colombi 1990). Thus, conditional on the job type $j \in \{fp, pp\}$, log-ability is the sum of independently drawn Gaussian and exponential random variables: $\log(\theta) = x_N + x_E$, where $x_N \sim N(\mu_{\theta,j}, \sigma^2_{\theta,j})$ and $x_E = Exp(\lambda_{\theta,j})$.

Calibration. We calibrate to model to match empirical evidence on performance-pay jobs, earnings elasticities, and the overall earnings distribution in the U.S. The chosen parameter values are summarized in Table 1.

Lemieux et al. (2009) (LMP) use the Panel Study of Income Dynamics to show that the fraction of performance-pay jobs s_{pp} was 0.45 in 1998, the most recent year included in their analysis. We replicated their analysis and found that mean earnings were 58 percent higher in performance-pay jobs than in fixed-pay jobs in 1998. This value pins down $\mu_{\theta,pp} - \mu_{\theta,fp}$, the difference in mean log-abilities between the two types of jobs. We postulate that the probability of a high output realization is given by $\pi(\ell) = \bar{\pi}\,\ell^{\rho}$, with $\rho \in (0, 1]$. In the data, the average probability of receiving a bonus conditional on having a performance-pay job is 23%, which pins down $\bar{\pi}$.[22] The exponent ρ affects the magnitude of earnings risk due to performance pay. LMP report that the variance of log-earnings in performance-pay jobs is 42 percent higher than in fixed-pay jobs. This excess variance can be explained either by the additional earnings risk generated by stochastic bonuses (controlled by ρ) or by a greater dispersion in the ex ante abilities of performance-pay workers ($\sigma^2_{\theta,pp} > \sigma^2_{\theta,fp}$). In our baseline calibration, we assume that the entire excess variance arises from the former channel; that is, we assume that log-abilities in the two types of jobs have the same dispersion ($\sigma^2_{\theta,pp} = \sigma^2_{\theta,fp} \equiv \sigma^2_{\theta}$). This makes performance pay as powerful in affecting earnings dispersion as the data allows: in our baseline calibration, performance pay explains 30 percent of the cross-sectional variance of log-earnings of performance-pay workers. We show later that our main conclusions are robust to this assumption.

To calibrate the overall mean and variance of log-earnings in the economy, we turn to the Survey of Consumer Finances (SCF), which uses data from the IRS Statistics of Income program to accurately represent the distribution of high-income households. Based on the SCF, Heathcote and Tsujiyama (2021) report a mean household labor

[22] Based on Table II and footnote 15 in LMP, we calculate that (i) in the 1990s, 82 percent of performance-pay jobs used bonuses, while the rest used piece rates or commissions; and (ii) the average frequency of receiving performance pay, conditional on having a performance-pay job, was 37 percent. Given that piece rates and commissions are paid out with certainty, we calculate that the probability of receiving a bonus, conditional on having a job with bonuses, is 23 percent.

Table 1: Calibrated parameters

parameter	value	description	source or target
s_{pp}	0.45	share of performance-pay jobs	Lemieux et al. (2009)
$\mu_{\theta,fp}$	3.43	mean log-ability at fixed-pay jobs	mean earnings in the economy
$\mu_{\theta,pp}$	5.29	mean log-ability at perf.-pay jobs	diff. in mean earnings between job types
σ_θ	0.29	normal variance of log-ability	variance of log-earnings in the economy
λ_θ	2.2	tail parameter of log-ability	Heathcote and Tsujiyama (2021)
$\bar{\pi}$	45	level parameter of $\pi(\cdot)$ function	mean frequency of bonus payments
ρ	0.82	curvature of $\pi(\cdot)$ function	diff. in var. of log-earnings between job types
ε_ℓ^F	0.5	Frisch elasticity of labor effort	Chetty et al. (2011),Keane (2011)
p	0.181	tax progressivity	Heathcote et al. (2017)
G/Y	0.188	share of government spending in GDP	Heathcote and Tsujiyama (2021)

Note: All the target moments are matched exactly.

income of \$77,325 and an overall variance of log-labor income of 0.618 in 2007. They also estimate that the tail parameter of the log wage distribution is equal to 2.2. We assume that the ability distributions in both types of jobs have a common tail parameter $\lambda_{\theta,pp} = \lambda_{\theta,fp} = 2.2$ and choose $\sigma_\theta^2 = 0.29$ to match the overall variance of log-earnings.

We model the disutility of labor effort as isoelastic: $h(\ell) = \frac{\ell^{1+1/\varepsilon_\ell^F}}{1+1/\varepsilon_\ell^F}$. A Frisch elasticity $\varepsilon_\ell^F = 0.5$ implies a compensated elasticity at fixed-pay jobs of approximately 0.3. Both values are consistent with empirical evidence (Chetty, Guren, Manoli, and Weber 2011; Keane 2011). Regarding government policy, Heathcote et al. (2017) estimate a value of 0.181 for the U.S. rate of tax progressivity, and Heathcote and Tsujiyama (2021) report a ratio of government purchases to output of 18.8 percent.

The implied distribution of earnings and job types is depicted in Figure 1. Specifically, in panel (b), we compare the (untargeted) shares of performance-pay jobs by earnings quartiles in the data and in the model.[23] The calibrated model successfully matches the empirical prevalence of performance-pay jobs: Both in the data and in the model, the share of performance-pay jobs is approximately 40 percent in the bottom three quartiles and it rises to more than 60 percent in the top quartile.

Incidence of a Large Tax Reform: From the Status Quo to the Optimum. We extend the optimal progressivity formula in the Appendix to account for fixed-pay jobs and a Pareto tail of earnings (see the proof of Theorem 1). We find that the utilitarian optimum rate of progressivity is equal to 0.376. This is more than twice as

[23]The data shares are computed for the year 1998 from PSID using the methodology of LMP. We are grateful to Karl Schulz for computing these values.

Figure 1: Joint distribution of earnings and job types

(a) Densities of log earnings (b) Share of perf.-pay jobs by earnings

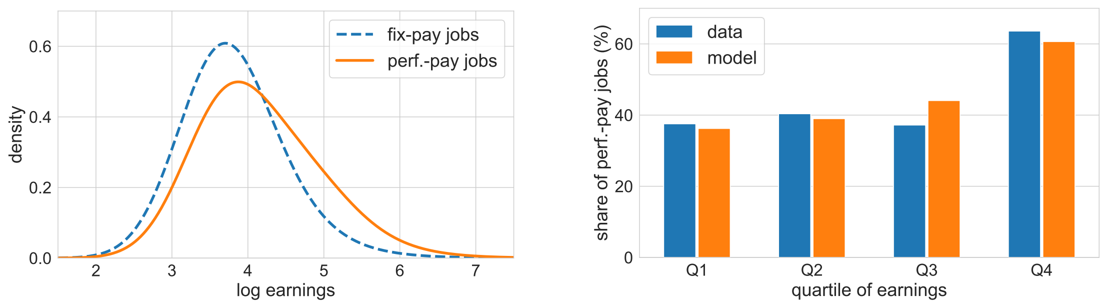
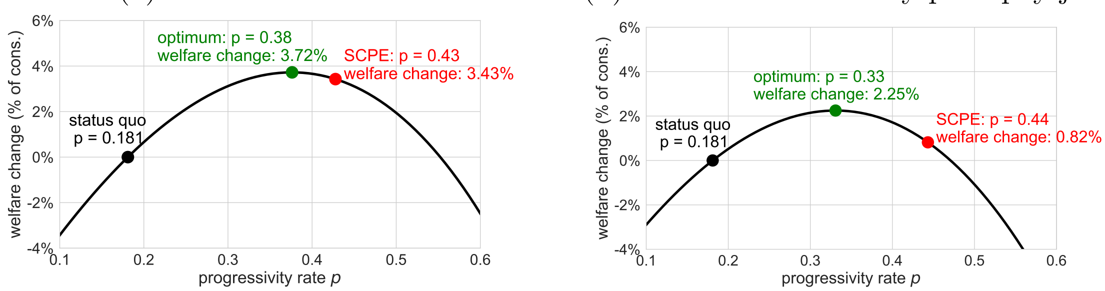


Figure 2: Optimal progressivity and Self-Confirming Policy Equilibrium

(a) Calibrated model (b) Counterfactual: only perf.-pay jobs

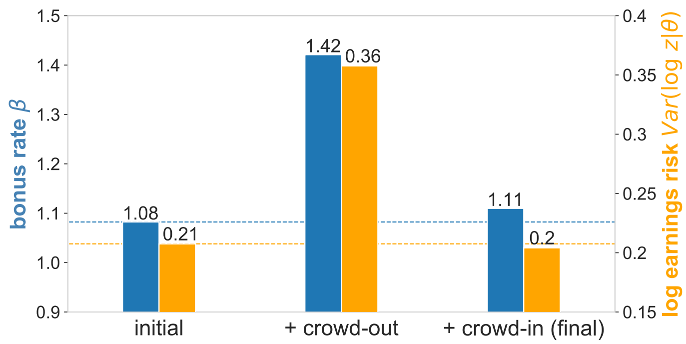


high as the current rate of tax progressivity in the U.S., and the implied increase in social welfare is equivalent to a 3.7 percent increase in consumption (see panel (a) of Figure 2). In this paragraph, we analyze the impact of a large reform that implements the optimal rate of progressivity, while keeping the overall tax revenue unchanged.

The impact of the reform on earnings risk in performance-pay jobs is depicted in Figure 3. Following a large increase in progressivity, the bonus rate β increases modestly from 1.08 to 1.11. As a result, the variance of log-earnings conditional on ability, equal to $\beta^2\pi(\ell_{pp})(1-\pi(\ell_{pp}))$, actually slightly decreases, as the impact of a higher bonus rate β is dominated by the impact of a lower effort ℓ_{pp}, distorted downwards by higher tax progressivity.

Underlying the weak response of earnings risk are two countervailing forces: the crowding-out and the crowding-in of private insurance. If firms attempted to motivate workers to maintain their original level of effort, better social insurance via tax progressivity would crowd-out private insurance one-for-one. For that to happen, the bonus rate β would need to increase from 1.08 to 1.42, raising the log-earnings risk

Figure 3: Optimal progressivity reform and earnings risk



Note: Tax progressivity rate is increased from 0.181 (status quo) to 0.376 (utilitarian optimum).

of performance-pay workers by 72 percent. However, in equilibrium firms choose to elicit a 9 percent lower effort level (which implies the same fall in mean earnings) and reduce the power of incentive-pay accordingly. This crowding-in effect counteracts the crowd-out and brings the bonus rate back to the vicinity of its original level. As a result, workers end up much better insured: The variance of log-consumption falls by 43 percent.

Importance of Performance Pay for Optimal Tax Progressivity. How important is it to account for endogenous performance pay when setting tax policy? To answer this question, we compare the optimal rate of progressivity with a rate chosen by the government that erroneously assumes that the entire earnings risk is exogenous. The latter is found by applying the formula for the optimal rate of progressivity from the model with exogenous partial insurance to our calibrated model economy—in which private insurance is actually endogenous. Since the rate of progressivity affects the perceived earnings dispersion, we iterate on the tax formula from the exogenous-insurance model until convergence to a fixed point. Following C. Rothschild and Scheuer (2016), we call the resulting allocation a Self-Confirming Policy Equilibrium (SCPE). The results are depicted in panel (a) of Figure 2.

A policymaker who ignores the endogeneity of earnings risk chooses a rate of progressivity (0.43) that is higher than the optimum (0.38). While we allow such a planner to estimate the elasticity of earnings precisely, the rate of progressivity is too high, since the negative effects of crowd-out on welfare and crowd-in on tax revenue are not internalized. Nevertheless, the welfare cost of the policy mistake is relatively small, equal to approximately 0.3 percent of consumption—this is significantly smaller than the gains from raising taxes from the status quo to the optimum. Recall that in the calibration we assumed that it is performance pay, rather than a greater dispersion of abilities, that explains the higher variance of earnings among performance-pay jobs. Allowing for more dispersed abilities at performance-pay jobs would reduce the magnitude of endogenous earnings risk due to stochastic bonuses. Thus, it would further reduce the already small welfare cost of ignoring the endogeneity of earnings risk. Hence, our baseline results provide an upper bound for the impact of endogenous insurance on social welfare.

We also simulate a counterfactual economy, assuming that all the jobs feature performance pay; see panel (b) of Figure 2. In this economy, the difference between the rate of progressivity at the optimum and in the SCPE doubles (from 0.055 to 0.11), while the welfare loss from the policy mistake more than quadruples (from 0.3 percent to 1.4 percent). Thus, the finding that the cost of ignoring performance pay for the design of tax policy is modest is not a theoretical necessity. It is rather a quantitative result that is due to the fact that only half of the jobs in the U.S. feature performance pay. If the share of performance-pay jobs continues to increase in the future, accounting for the welfare effects of endogenous earnings risk may become important for tax policy.

1.5 Commissions, Stock Options, Dynamic Incentives

Our baseline model, in which earnings are binary, is well suited to analyzing contracts that consist of a base salary and a bonus. We argued above that these represent more than 80 percent of the performance-based earnings contracts. Nevertheless, it is important to evaluate whether our main insights carry over to other types of compensation. In Appendix C, we set up several alternative frameworks. The first, based on Holmstrom and Milgrom (1987), provides conditions under which linear contracts are optimal: It is thus a natural setting to study piece rates or commissions. Stock options can be represented by contracts that are convex in performance: To analyze the impact and optimality of tax policy with such contracts in a tractable way, we build on Edmans and Gabaix (2011). Finally, high-powered incentives can be provided over time via, e.g., promotions or salary raises. To study such effects, we use the frameworks of Sannikov (2008) or Edmans et al. (2012).

Despite their differences, all of these models share a common structure that is critical for our analysis: When the utility is logarithmic and the tax schedule is CRP, the slope of the earnings contract, which measures the sensitivity of earnings to output shocks, is given by $h'(\ell)/(1-p)$.[24] This is a direct consequence of the local incentive constraint, which is common to all of these models and states that the sensitivity of utility to output shocks must be equal to the marginal disutility of effort $h'(\ell)$. In turn, this general principle implies that our discussion of the incidence of tax progressivity continues to hold: A tax change creates crowd-out and crowd-in effects on the pre-tax earnings distribution captured by the elasticities $\varepsilon_{\beta,1-p} = -1$ and $\varepsilon_{\beta,\ell} = 1/\varepsilon_{\ell}^{F}$. To the extent that the impact of the tax change on effort $\varepsilon_{\ell,1-p}$ is close to the Frisch elasticity of labor supply, the contract remains insensitive to tax policy.[25] By contrast, taxes have a large impact on welfare since the crowd-in affects utility only to a second-order. As in Theorem 1, these effects are reflected in the optimal tax formulas in these respective environments (see, e.g., Theorems 4 and 5 in Appendix C).

2 General Analysis of Non-Linear Taxation

The model we studied in Section 1 relies on several strong assumptions: The utility of consumption is logarithmic, the tax schedule is a function of total earnings and has a constant rate of progressivity, and the distribution of earnings is lognormal. We now relax these restrictions. In particular, we allow the government to tax base earnings and bonuses either jointly or separately in an arbitrarily nonlinear way. We characterize the incidence of tax reforms in Section 2.2 and derive analytical formulas for the optimal tax rate on top earners—levied either on total earnings or on bonuses separately from the base pay—in Sections 2.3 and 2.4. We evaluate these results quantitatively in Section 2.5. Additional results are derived in Appendix B.

[24] In the case of Holmstrom and Milgrom (1987), the theoretical restrictions on agents' preferences allow us to study only CARA utility function with affine taxes. In this case, the relevant notion of pay-performance sensitivity becomes $\theta^{-1}h'(\ell)/(1-\tau)$, where τ is the tax rate. Clearly, our discussion continues to hold once we focus on reforms of the tax rate τ.

[25] This discussion suggests that these results hold more generally in any moral-hazard model with a continuous choice of effort in which the first-order approach (or the one-shot deviation principle in the continuous-time setting of Sannikov 2008) is valid.

2.1 Environment

The environment is the same as in Section 1, except that the utility of consumption u is now a twice continuously differentiable and concave function, and the retention schedule $R : \mathbb{R}_+^2 \to \mathbb{R}$ is an a.e. twice continuously differentiable function of the agent's base pay and realized bonus. Thus, a worker with high output realization consumes $c = R(\underline{z}, b)$ and gets utility $v(\underline{z}, b) \equiv u(R(\underline{z}, b))$, while a worker with low output realization consumes $c = R(\underline{z}, 0)$ and gets utility $v(\underline{z}, 0) \equiv u(R(\underline{z}, 0))$. This specification allows us to consider both the joint taxation of overall earnings, and the separate taxation of fixed and variable pay. We impose the following assumption, which ensures that the tax schedule is not "too regressive:"[26]

Assumption 2 *The utility of pre-tax earnings v is concave on $\mathbb{R}_+^2$.*

The firm maximizes its expected profit (1) subject to, *mutatis mutandis*, the incentive constraint (2) and the participation constraint (3). The equilibrium reservation value is pinned down by the free-entry condition (4). The following lemma shows formally that we can replace the global incentive constraint with the corresponding first-order condition for effort (5).[27]

Lemma 4 *Suppose that the equilibrium effort level is interior. The firm's problem is equivalent to maximizing (1) subject to (3) and the local incentive constraint:*

$$h'(\ell) \;=\; v(\underline{z}, b) - v(\underline{z}, 0). \tag{15}$$

As in the previous section, equation (15) along with the convexity of h implies that eliciting a higher effort level from workers requires raising their exposure to output risk. The following characterization of the optimal contract generalizes Proposition 1 and has an analogous interpretation.

[26] See Appendix for details. In particular, this condition always holds under Assumption 1 regardless of the value of p. It is a natural restriction: Doligalski (2019) shows that when this condition is violated, firms have incentives to offer stochastic earnings even in the absence of moral hazard frictions. Furthermore, a tax schedule that encourages such earnings randomization is Pareto inefficient.

[27] Rogerson (1985) credits an unpublished paper by Holmstrom (1984) for the first proof of validity of the first-order approach in such a setting.

Proposition 2 *The base pay $\underline{z}(\theta)$ and high-level pay $\bar{z}(\theta)$ satisfy*

$$v(\underline{z}, 0) - h(\ell) \;=\; U(\theta) - \ell h'(\ell) \tag{16}$$
$$v(\underline{z}, b) - h(\ell) \;=\; U(\theta) + (1-\ell)h'(\ell). \tag{17}$$

The effort level $\ell(\theta)$ exerted by the worker satisfies

$$\theta \;=\; b + \left[\frac{1}{v_2(\underline{z}, b)} - \frac{1 - \frac{R_1(\underline{z},b)}{R_2(\underline{z},b)}\ell}{1-\ell}\frac{1}{v_1(\underline{z}, 0)}\right]\ell(1-\ell)h''(\ell). \tag{18}$$

Expected utility $U(\theta)$ satisfies $\underline{z} + \ell b = \theta\ell$.

2.2 Tax Incidence

In this section, we evaluate the impact of tax reforms on workers' labor contracts and welfare. For conciseness, we focus on the taxation of overall earnings and relegate the analysis of the separate taxation of bonuses to the Appendix (Lemma 9). Consider a given baseline retention schedule $R : \mathbb{R}_+ \to \mathbb{R}$ and another function ("tax reform") $\hat{R} : \mathbb{R}_+ \to \mathbb{R}$. The first-order change in a functional $\Psi(R)$ in response to the reform $\hat{R}$ is given by the Gateaux derivative $\hat{\Psi}(R, \hat{R}) \equiv \lim_{\delta\to 0} \left(\Psi(R + \delta\hat{R}) - \Psi(R)\right)/\delta$.

Earnings. We start by studying the effect of tax reforms on the worker's earnings. We obtain the following characterization:

Lemma 5 *The first-order impact of a tax reform $\hat{R}$ on earnings $\underline{z}, \bar{z}$ is given by*

$$\hat{\underline{z}} \;=\; \left[-\frac{\hat{R}(\underline{z})}{R'(\underline{z})} + \frac{\hat{U}}{v'(\underline{z})}\right] - \frac{\ell h''(\ell)}{v'(\underline{z})}\hat{\ell} \tag{19}$$
$$\hat{\bar{z}} \;=\; \left[-\frac{\hat{R}(\bar{z})}{R'(\bar{z})} + \frac{\hat{U}}{v'(\bar{z})}\right] + \frac{(1-\ell)h''(\ell)}{v'(\bar{z})}\hat{\ell}, \tag{20}$$

where $\hat{U}$ and $\hat{\ell}$ denote the first-order changes in the reservation utility and labor effort due to the reform. The terms in square brackets constitute the crowd-out, while the terms multiplied by $\hat{\ell}$ constitute the crowd-in.

Lemma 5 shows that the tax reform affects earnings $z \in \{\underline{z}, \bar{z}\}$ via three channels, the first two of which jointly constitute the crowd-out, while the last is the crowd-in.

The first term in equations (19) and (20), $-\hat{R}(z)/R'(z)$, implies that, *ceteris paribus*, the agent's consumption $c = R(z)$ remains unchanged despite the tax change. Indeed, this term implies

$$\hat{c} \;\equiv\; \hat{R}(z) + R'(z)\hat{z} \;=\; \hat{R}(z) - R'(z)\frac{\hat{R}(z)}{R'(z)} \;=\; 0.$$

Thus, absent any change in the reservation utility and optimal labor effort, the firm would simply adjust pre-tax earnings so as to keep the worker's disposable income levels $\underline{c}$ and $\bar{c}$ fixed. In other words, any attempt by the government to affect consumption insurance would be fully offset by the firm in order to preserve incentives.

Second, the tax reform affects the earnings contract via its impact on the equilibrium reservation utility. The increase in income z resulting from an increase in the reservation value $\hat{U} > 0$ is inversely proportional to the marginal utility $v'(z)$. Thus, the earnings of the high-performers increase by a larger amount than those of the low-performers with the same ability. This ensures that the utility gain $\hat{U}$ is distributed uniformly across agents regardless of their performance, thus preserving incentive compatibility (15).

Third, the tax reform modifies the desired effort level. Recall that, by equation (15), eliciting higher effort $\hat{\ell} > 0$ requires widening the gap between the utility of high- and low-performers by $\Delta h'(\ell) = h''(\ell)\hat{\ell}$. The implied change in earnings—the crowd-in effect—is given by the third term in equations (19) and (20).

Expected Utility. Next, we evaluate the effect of tax reforms on the worker's expected utility.

Lemma 6 *The first-order impact of a tax reform $\hat{R}$ on expected utility $U(\theta)$ is given by*

$$\hat{U} \;=\; \mathbb{E}\left[\mu(z)u'(R(z))\hat{R}(z)\right] \;\; \text{where} \;\; \mu(z) = \frac{\frac{1}{v'(z)}}{\mathbb{E}\left[\frac{1}{v'}\right]} \;\text{for}\; z \in \{\underline{z}, \bar{z}\}, \tag{21}$$

Note in particular that, to a first order, the worker's expected utility is unaffected by changes in labor effort (envelope theorem).

In the standard (full-insurance) model, a tax cut $\hat{R}(z) > 0$ affects the worker's utility in proportion to their marginal utility of consumption $u'(R(z))$, and the envelope theorem ensures that the endogenous behavioral responses have no first-order impact on utility. This is no longer true in the model with performance pay. While the envelope theorem still applies to the endogenous effort (i.e., crowd-in) responses, the earnings adjustments caused by the crowd-out of private insurance have a first-order impact on welfare. The additional factor μ present in equation (21) accounts for these welfare effects.[28]

The intuition is analogous to that laid out in Section 1: A tax cut initially generates a rent to the firm—the first term in equations (19) and (20)—which is then transferred to the worker via free entry and the resulting adjustment of the reservation value. This transfer needs to keep the original effort level incentive-compatible and, hence, the utility difference $v(\bar{z}) - v(\underline{z})$ unchanged. Thus, earnings z must change in proportion to $\frac{1}{v'(z)}$, leading to the factor $\mu(z)$ in Lemma 6. Since the marginal utility v' is decreasing, this means that the high-level pay rises more than the base pay, resulting in a regressive distribution of rents.

To measure the welfare impact of tax reforms, the standard approach is to use marginal social welfare weights, defined as the welfare impact of marginally increasing the consumption of a given agent (Saez & Stantcheva, 2016). In the standard Mirrlees model, the marginal welfare weight of a worker with ability θ and earnings z is given by

$$g(z \mid \theta) = \frac{1}{\lambda}\alpha(\theta)u'(R(z)), \tag{22}$$

where $\alpha(\theta)$ is the Pareto weight of type θ in the social objective, and λ is the marginal value of public funds. The social marginal welfare weights are convenient in the standard model because, by the envelope theorem, all endogenous responses to tax changes have at most a second-order impact on the agent's utility.

By contrast, recall that in our model the envelope theorem does not apply to all endogenous responses: The crowd-out of private insurance has a first-order impact on welfare. To account for this effect, we define the *modified* marginal social welfare weights $\tilde{g}(z \mid \theta)$ as the welfare impact of marginally reducing the tax liability of an agent with type θ and realized income z.

Corollary 1 *The modified marginal social welfare weights satisfy*

$$\frac{\tilde{g}(\underline{z} \mid \theta)}{g(\underline{z} \mid \theta)} = \mu(\underline{z}) < 1 \quad \textit{and} \quad \frac{\tilde{g}(\bar{z} \mid \theta)}{g(\bar{z} \mid \theta)} = \mu(\bar{z}) > 1. \tag{23}$$

[28] Chang and Park (2017) show a similar partial applicability of the envelope theorem in an Alvarez and Jermann (2000) economy with asset trades and endogenous borrowing limits.

Relative to the standard model, the contribution to social welfare of agents with type θ is adjusted upwards when they receive a bonus, and downwards otherwise. These regressive adjustments to the marginal social welfare weights reflect the fact that, as explained above, tax cuts are passed through within the firm primarily to the high-performing workers. It is therefore more difficult for the government to target transfers at unlucky workers, i.e., to directly raise the consumption of those whose marginal utility is relatively large.

Labor Effort. Next, we characterize the incidence of tax reforms on labor effort. The general analytical expressions for the elasticities of effort with respect to marginal and average tax rates are given in Lemma 8 in the Appendix. For conciseness, the following Lemma only focuses on their signs.

Lemma 7 *The first-order impact of a tax reform $\hat{R}$ on labor effort ℓ can be expressed as*

$$\frac{\hat{\ell}}{\ell} \quad = \quad \varepsilon_{\ell,R'(\underline{z})}\frac{\hat{R}'(\underline{z})}{R'(\underline{z})} \; + \; \varepsilon_{\ell,R'(\bar{z})}\frac{\hat{R}'(\bar{z})}{R'(\bar{z})} \; + \; \varepsilon_{\ell,R(\underline{z})}\frac{\hat{R}(\underline{z})}{R(\underline{z})} \; + \; \varepsilon_{\ell,R(\bar{z})}\frac{\hat{R}(\bar{z})}{R(\bar{z})}. \tag{24}$$

Suppose that the base pay $\underline{z}$ and high-level pay $\bar{z}$ are both located in brackets where the marginal tax rate is locally constant. A higher marginal tax rate on base (resp. high-level) pay raises (resp., reduces) labor effort: $\varepsilon_{\ell,R'(\underline{z})} < 0 < \varepsilon_{\ell,R'(\bar{z})}$. Suppose moreover that the utility function is logarithmic. A higher average tax rate on base pay raises labor effort: $\varepsilon_{\ell,R(\underline{z})} < 0$. A higher average tax rate on high-level pay reduces labor effort, so that $\varepsilon_{\ell,R(\bar{z})} > 0$, if and only if $R'(\underline{z}) < R(\underline{z})/\underline{z}$.

In the standard model, effort responds negatively to the marginal tax rate (MTR) due to the substitution effect and positively to the average tax rate (ATR) due to the income effect. Our model is more complex, as there are two relevant levels of pay (base and high-level) and, thus, four tax rates that impact behavior. Focusing first on the average rates, we find that a higher ATR levied on the base pay increases effort, but a higher ATR levied on the high-level pay reduces effort (when the tax schedule is progressive). Intuitively, adjusting effort in this way allows the worker to escape some of the increased tax burden by reducing the probability of receiving the income that is taxed more heavily.

Moving on to marginal tax rates, we find that a higher MTR levied on high-level pay reduces effort, but a higher MTR on base pay, perhaps surprisingly, *increases* effort. Note that a change in marginal rates alone does not modify workers' incentives for effort, since the incentive constraint (15) is unaffected. What are modified, however, are the firm's incentives to offer different pay structures. Intuitively, a higher MTR gives firms incentives to reduce gross pay and save on payroll, since the implied reduction in after-tax income—which matters for workers—is now smaller. A higher MTR at the high-level pay $\bar{z}$ then leads to a lower value of $\bar{z}$ and, via the incentive constraint, to a lower effort level. Conversely, a higher MTR on base pay $\underline{z}$ leads to a lower value of $\underline{z}$ and, hence, a higher effort level.

Summing up, we find that labor effort responds very differently to taxes on base pay and on high-level pay. In particular, increasing the marginal tax rate on base pay actually encourages effort. These findings underlie our later results on the efficiency gains from taxing bonuses at a lower rate than base pay.

Taking Stock: Crowd-Out and Crowd-In. We can now gather our previous results to obtain a full characterization of the incidence of tax reforms on the earnings contract. The following result, obtained by substituting expression (21) into (19) and (20), generalizes the crowding-in and crowding-out forces (highlighted in Section 1) to arbitrary tax systems and reforms.

Proposition 3 *The first-order impact of a tax reform $\hat{R}$ on earnings $\underline{z}, \bar{z}$ is given by*

$$(1-\ell)\hat{\underline{z}} \;=\; -\,(1-\varepsilon_{out})(1-\ell)\frac{\hat{R}(\underline{z})}{R'(\underline{z})} \;+\; \varepsilon_{out}\,\ell\frac{\hat{R}(\bar{z})}{R'(\bar{z})} \;-\; \ell\bar{z}\,\varepsilon_{in}\,\frac{\hat{\ell}}{\ell} \tag{25}$$

$$\ell\hat{\bar{z}} \;=\; (1-\varepsilon_{out})(1-\ell)\frac{\hat{R}(\underline{z})}{R'(\underline{z})} \;-\; \varepsilon_{out}\,\ell\frac{\hat{R}(\bar{z})}{R'(\bar{z})} \;+\; [\ell\bar{z}\varepsilon_{in}+\underline{z}]\,\frac{\hat{\ell}}{\ell}, \tag{26}$$

where the crowd-out parameter $\varepsilon_{out} \in (0,1)$ and the crowd-in parameter $\varepsilon_{in} > 0$ are respectively given by

$$\varepsilon_{out} \;=\; \frac{1}{\mathbb{E}\left[\frac{1}{v'(z)}\right]}\,\frac{1-\ell}{v'(\underline{z})} \qquad \textit{and} \qquad \varepsilon_{in} \;=\; \frac{\ell(1-\ell)h''(\ell)}{\bar{z}\,v'(\underline{z})},$$

and where the labor supply response $\hat{\ell}/\ell$ is given by Lemma 7 (and Lemma 8 in the Appendix).

Proposition 3 shows that earnings $\underline{z}, \bar{z}$ respond to changes in tax payments at both levels of compensation. A tax hike on high-level earnings $\hat{R}(\bar{z}) < 0$, for instance, leads the firm to lower the base pay $\underline{z}$ and raise the high-level pay $\bar{z}$. The magnitude of this crowding-out of private insurance is determined by the parameter $\varepsilon_{out} \in (0,1)$, for which our model gives a simple analytic expression.[29] Moreover, a tax reform that leads to a lower optimal effort level, $\hat{\ell} < 0$, triggers an increase in the base pay and a decrease in the high-level pay. The magnitude of this crowding-in of private insurance is determined by the product of ε_{in} and the percentage change in labor effort $\hat{\ell}/\ell$.

2.3 Optimal Top Tax Rate on Total Earnings

We now proceed to characterizing the optimal non-linear taxes in our setting. In the main body of the paper, we focus on the problem of finding the optimal tax rate in the highest bracket, and we relegate the analysis of the full optimal taxation problem to Appendix B.2. Determining the optimal top tax rate is an especially salient policy question: A large share of the rise in inequality since the 1980s has been driven by the explosion of performance-based forms of compensation for the highest income earners, such as bankers' bonuses and CEOs' stock options (see, e.g., Bell and Van Reenen 2013; Lemieux et al. 2009; Piketty and Saez 2003).[30]

We consider the following policy experiment. Fixing the threshold z^* of the top tax bracket, we optimally set the marginal tax rate τ that applies to this bracket. Before stating our main result, we introduce some notation.

Top and Intermediate Earners. There are two kinds of workers who react to changes in the top-bracket tax rate. Some (the most productive) of them receive an income in the top tax bracket regardless of their performance: We refer to them as "top earners." Others earn base pay that is below the top bracket threshold z^*, but high-level pay that is above that threshold. Because they have a positive probability of landing in the top bracket, their entire contract—including their base salary—adjusts in response to the tax change. We refer to this group as "intermediate earners." We

[29] For instance, if the utility of consumption is CRRA with risk aversion σ and the tax schedule is CRP, we have $\varepsilon_{out} = (1-\ell)/[1-\ell+\ell e^{(p+(1-p)\sigma)\beta}]$, where β is the bonus rate defined in Section 1.

[30] Bell and Van Reenen (2014) show that in the UK, 83 percent of workers in the top percentile received a bonus in 2008, and that bonuses represented 35 percent of their total compensation (44 percent in the financial sector).

denote by σ^I the fraction of intermediate workers among all (top and intermediate) workers, and by s^I the income share of intermediate earners in the top bracket.

Income distribution. Since earnings of top earners always fall into the top bracket, it is sufficient to keep track of their mean earnings, which we denote by Z^T. By contrast, for intermediate earners, we need to keep track of the level of their base pay $\underline{z}$, which falls short of z^*, as well as the endogenous frequency with which it (rather than the high-level pay $\underline{z} + b$) is realized. To do so, it suffices to define the mean frequency-adjusted base pay $(1 - \ell)\underline{z}$ among intermediate earners, which we denote by $\underline{Z}^I$. Responses of $\underline{Z}^I$ to changes in the top tax rate account for both the level responses $(1 - \ell)\hat{\underline{z}}$ and the frequency responses $-\hat{\ell}\underline{z}$. Analogously, we denote the mean frequency-adjusted high-level pay $\ell\bar{z}$ of intermediate earners by $\bar{Z}^I$, and their mean labor effort ℓ by L^I. Aggregating over intermediate and top earners in the top bracket yields $Z^* \equiv (1 - \sigma^I)Z^T + \sigma^I \bar{Z}^I$. Finally, we introduce the Pareto coefficient of the income distribution in the top bracket $\rho = \frac{E[z|z \geq z^*]}{E[z|z \geq z^*] - z^*}$, where E denotes the expectation with respect to the realized (i.e., observed) earnings distribution.

Elasticities. Consider the tax reform that raises the marginal tax rate τ by $\hat{\tau} > 0$ above the income level z^*. To keep track of the impact of this tax reform on various aggregates, we define the elasticity of any variable x (e.g., L^I, $\underline{Z}^I$, Z^*) with respect to the top net-of-tax rate $1 - \tau$ as $e_x \equiv -\frac{1-\tau}{\hat{\tau}}\frac{\hat{x}}{x}$. This shorthand notation measures the total percentage change in x caused by the reform $\hat{\tau}$.[31] These elasticities can be easily expressed in terms of primitives using Lemma 7 and Proposition 5.

Welfare Weights. We defined the modified marginal welfare weights $\tilde{g}$ in Corollary 1. We denote the income-weighted average modified welfare weight in the top bracket by $\tilde{\mathcal{G}}$.

Optimum Top Tax Rate. We now proceed to characterizing the optimal tax rate on total earnings in the top bracket τ. Suppose for simplicity that the second highest tax bracket is large enough so that all the intermediate workers pay the same marginal tax rate t on their base pay.

[31] The variable e_x is thus defined as a "policy elasticity," a concept proposed by Hendren (2016).

Theorem 2 *The optimal tax rate on total earnings in the top bracket is given by*

$$\frac{\tau}{1-\tau} = \frac{1 - \tilde{\mathcal{G}} - s^I \left(\frac{t}{1-\tau}\kappa_1 e_{\underline{Z}^I} + \frac{t-\tau}{1-\tau}\kappa_2 e_{L^I}\right)}{\rho\, e_{Z^*}}, \tag{27}$$

where $\kappa_1 = (\underline{Z}^I/\bar{Z}^I)\rho^I > 0$, $\kappa_2 = \rho^I - 1 > 0$, and $\rho^I = \frac{\bar{Z}^I}{\bar{Z}^I - L^I z^*}$. The elasticity $e_{\underline{Z}^I}$ can in turn be decomposed as

$$\kappa_1 e_{\underline{Z}^I} = \overline{\varepsilon_{out}}^I - \rho^I \overline{\varepsilon_{in} e_\ell}^I + \kappa_1 \overline{e_{1-\ell}}^I, \tag{28}$$

where $\overline{\varepsilon_{out}}^I$, $\overline{\varepsilon_{in} e_\ell}^I$, $\overline{e_{1-\ell}}^I$ are income-weighted averages (formally defined in the Appendix) of the crowd-out, crowd-in and effort elasticities among intermediate earners, respectively.

The remainder of this section is devoted to analyzing this formula.

Full-Insurance Benchmark. Consider first the setting where firms provide full insurance against output risk. The optimal top tax rate is then determined by the following formula (Saez, 2001):

$$\frac{\tau}{1-\tau} = \frac{1-\mathcal{G}}{\rho\, e_{Z^*}}, \tag{29}$$

where $\mathcal{G}$ is the aggregate marginal social welfare weights (see equation (22)) in the top bracket. This expression shows that the optimal tax rate is an increasing function of income inequality measured by the inverse Pareto coefficient $1/\rho$, a decreasing function of the average elasticity of top earnings e_{Z^*} that captures the efficiency cost of raising the tax rate, and a decreasing function of the average social welfare weight $\mathcal{G}$ that the social objective assigns to agents with earnings $z \geq z^*$.

Welfare Effect of Crowd-Out. The first novelty introduced by the endogeneity of private insurance is the adjustment to the marginal social welfare weights, which accounts for the first-order welfare effects of crowd-out. As shown in Corollary 1, these responses are captured by placing lower weights on the (unlucky) workers who do not receive a bonus and higher weights on the (lucky) workers who do. Suppose for simplicity that the top bracket is only composed of the lucky intermediate earners. In this case, the average social marginal welfare weight is unambiguously higher than in the standard model, i.e., $\tilde{\mathcal{G}} > \mathcal{G}$.[32] Thus, the crowd-out reduces the welfare benefits of raising taxes on top incomes and contributes to a strictly lower optimal top tax rate. In the Appendix, we show that if the utility function is logarithmic, this result also holds in the presence of top earners who receive both their base pay and their high-level pay in the top bracket.[33]

Fiscal Externality: Base-Pay of Intermediate Workers. Second, applying the benchmark full-insurance formula (29) in our setting would mistakenly ignore that perturbing the top tax rate spills over to workers who could have, but haven't, made it to the top. Indeed, some agents—the unlucky intermediate earners—react to the top tax rate change even though their income is lower than the threshold z^*. This fiscal externality is captured by the elasticity $e_{\underline{Z}^I}$ in formula (27). This elasticity can in turn be decomposed into the average crowd-out parameter $\overline{\varepsilon_{out}}^I$, the average product of the crowd-in parameter and the individual labor effort elasticity $\overline{\varepsilon_{in} e_\ell}^I$, and the average elasticity of the base pay frequency $\overline{e_{1-\ell}}^I$, as shown in expression (28). Intuitively, the crowd-out and crowd-in forces determine the level response of base pay, while the effort elasticity captures the change in the frequency of base pay realization. If the resulting elasticity $e_{\underline{Z}^I}$ is negative, say, then a higher top tax rate leads to a higher average base pay for intermediate workers, which increases the tax revenue collected from the lower tax bracket. In that case, this fiscal externality contributes to a higher optimal top tax rate.

Fiscal Externality: Frequency Margin with a Nonlinear Tax Schedule. Finally, we need to correct the calculation of the fiscal impact of frequency responses of intermediate earners—i.e., changes in their probability of receiving a bonus due to a change in the top tax rate—when the tax schedule is nonlinear at the top (i.e., $\tau \neq t$). The elasticities e_{Z^*} and $e_{\underline{Z}^I}$ already incorporate the frequency responses, but the corresponding fiscal externality terms in formula 27 implicitly assume that both income levels are taxed at the same rate (τ or t, respectively). However, the fiscal impact of these responses is proportional to the difference in tax liabilities between the two states, which is equal to neither τ nor t but to a convex combination of the

[32]This argument takes as a given the marginal value of public funds λ.

[33]More generally, this result holds as long as the utility losses caused by raising the top tax rate, $(z - z^*)u'(R(z))$, are increasing with income.

two.[34] The last term in the numerator of (27) makes this necessary correction; This is the same adjustment that led to the term κ_3 in formula (13) in Section 1.3.

Taking stock. To sum up, the endogeneity of private insurance modifies the standard formula (29) in three ways: The marginal social welfare weights are adjusted to account for the welfare consequences of crowd-out; Two new fiscal externalities arise from the endogenous responses of base pay in the lower bracket and from the frequency responses when the tax schedule is nonlinear at the top. We explore the importance of these effects quantitatively in Section 2.5.

2.4 Separate Taxation of Fixed and Variable Pay

So far, we have focused on tax schedules that are functions of total earnings, i.e., the sum of the base pay and, if any, the bonus. This corresponds to the tax treatment of bonuses in most countries. It is conceivable, however, that the tax system could treat the two types of pay differently. In the U.S., for instance, CEO compensation in excess of $1 million used to be deductible from the company's corporate income tax only if it was performance-based (Hall & Liebman, 2000).[35] Thus, top performance-based earnings used to have a tax advantage.[36]

In this section we investigate whether there are gains from separating the tax treatment of base earnings and bonuses. We show that it is indeed optimal to tax bonuses at a lower rate than base pay. This provides an explanation for the past tax practice in the U.S. We then characterize analytically the optimal top tax rate on bonuses.

Local Gains From Separating Taxes on Fixed and Variable Pay. Suppose that base earnings and bonuses are initially taxed jointly, as in the previous section. For clarity of exposition, we assume that the utility function is logarithmic and the tax schedule is initially linear with rate $\tau > 0$, but the main insight holds more

[34] To be precise, for intermediate workers we have $\hat{\ell}(T(\bar{z}) - T(\underline{z})) = \hat{\ell}(\tau(\bar{z} - z^*) + t(z^* - \underline{z}))$.

[35] Preferential treatment of performance-based compensation ended with the 2017 Tax Cuts and Jobs Act.

[36] One may be concerned that taxing bonuses at a lower effective rate will lead to an effort to disguise base pay as bonuses. This is precisely the reason why performance pay of top CEOs needed to be qualified (i.e., based on objective targets and administered by an external committee) to remain deductible (Rose & Wolfram, 2002). Such a qualification is costly and, arguably, not practical beyond top earners.

generally. Consider the following policy experiment. Raise the marginal tax rate on base pay from τ to $\tau + \delta_{\underline{z}}$ and simultaneously lower the marginal tax rate on bonuses from τ to $\tau - \delta_b$, with $\delta_{\underline{z}}, \delta_b > 0$, thus effectively decoupling their tax treatment. We engineer this perturbation in such a way that every worker's expected utility remains constant, which is the case when $\delta_{\underline{z}} = \frac{\ell b}{\underline{z}}\delta_b$ for all agents.[37]

Proposition 4 *Suppose that the utility is logarithmic and the tax schedule is initially a linear function of total earnings with a positive tax rate. Lowering the marginal tax rate on bonuses by* $-\delta_b < 0$ *and raising the marginal tax rate on base pay by* $\delta_{\underline{z}} = \frac{\ell b}{\underline{z}}\delta_b > 0$ *yields a strict improvement in social welfare.*

To understand this result, note first that by Lemma 9 in the Appendix, the tax reform keeps expected utility unchanged, $\hat{U}(\theta) = 0$ for all θ, so that the social objective (11) remains constant. As a result, the tax reform improves social welfare if and only if it yields a strict gain in government revenue.[38] We then show that the tax reform indeed generates higher tax revenue, by raising the effort level of all agents. As we show in Lemma 7, both the reduction in the bonus tax rate and the increase in the base pay tax rate boost labor effort. Therefore, the planner achieves a Pareto improvement by inducing individuals to work harder while keeping their expected utility unchanged, thereby freeing up resources available for redistribution.

The intuition behind this result is as follows: Compensating workers with bonuses for high performance is a *constrained efficient* way of dealing with agency frictions in the labor market. Taxing bonuses removes the only instrument at the firm's disposal to maintain workers' incentives for effort. It is, thus, highly distortive. On the other hand, taxing base pay does not have such a large efficiency cost—on the margin it even encourages effort. It is therefore optimal to tax bonuses at a lower rate to minimize distortions, and to tax base pay at a higher rate to provide redistribution.

Optimal Top Tax Rate on Bonuses. Our last theoretical result is the characterization of the optimal top tax rate on bonuses for a given tax schedule on base pay.[39]

[37] Recall from our analysis of Section 1 that with log utility and a linear tax schedule, $\ell b/\underline{z}$ is independent of θ.

[38] This tax revenue can, in turn, be rebated to workers to provide a higher level of redistribution.

[39] We refrain from optimizing simultaneously over the top tax rate on bonuses and the top tax rate on base pay for the following reason: Suppose for simplicity that neither bonuses nor base pay

Let $[b^*, \infty)$ denote the highest bonus tax bracket, and τ_b be the corresponding marginal tax rate. Let $[\underline{z}^*, \infty)$ denote the interval of base earnings of workers whose bonus falls in the top bracket ("top earners"). We suppose for simplicity that $\underline{z}^*$ is itself in the top bracket of the base pay tax schedule and denote by τ_z the corresponding marginal tax rate. Our policy experiment consists of choosing optimally the top bonus tax rate τ_b, while keeping the bonus tax schedule $T_b(\cdot)$ below b^* and the base pay tax schedule fixed.

Denote the mean frequency-adjusted bonus ℓb among top earners by B. Thus, tracking changes in B accounts for responses to top tax rate changes along two margins: the level of bonuses $(\ell \hat{b})$ and their frequency $(\hat{\ell} b)$. Second, denote by L the average probability with which top earners receive a bonus and by $\underline{\mathcal{Z}}$ their average base pay. Finally, the Pareto coefficient of the bonus distribution ρ_b is given by $\rho_b = \frac{E[b|b \geq b^*]}{E[b|b \geq b^*] - b^*} = \frac{B/L}{B/L - b^*}$, where E denotes the expectation over the (observed) distribution of realized bonuses.

Theorem 3 *The optimal top tax rate on bonuses is given by*

$$\frac{\tau_b}{1-\tau_b} = \frac{1 - \tilde{\mathcal{G}} - \frac{\tau_z}{1-\tau_b}\kappa_1 e_{\underline{\mathcal{Z}}} - \frac{t_b - \tau_b}{1-\tau_b}\kappa_2 e_L}{\rho_b \, e_B}, \tag{30}$$

where $t_b = T_b(b^*)/b^*$ is the average bonus tax rate at b^*, $\kappa_1 = (\underline{\mathcal{Z}}/B)\rho_b$ and $\kappa_2 = 1 - \rho_b$, and the modified marginal social welfare weight $\tilde{\mathcal{G}}$ is derived formally in the Appendix. The elasticity of mean base earnings $e_{\underline{\mathcal{Z}}}$ can in turn be decomposed as

$$\kappa_1 e_{\underline{\mathcal{Z}}} = \overline{\varepsilon_{out}} - \rho_b \overline{\varepsilon_{in} e_\ell}, \tag{31}$$

where $\overline{\varepsilon_{out}}$, $\overline{\varepsilon_{in} e_\ell}$ are income-weighted averages of the crowd-in and crowd-out param-

are taxed below the respective thresholds b^* and $\underline{z}^*$, but above these thresholds both forms of pay are taxed with a rate close to 100 percent. In the standard model, taxing top incomes close to 100 percent is never optimal, as it severely distorts the effort of high-ability workers. In the moral-hazard model, however, the effort incentives of high-ability types are driven by $u(R(\underline{z}, b)) - u(R(\underline{z}, 0)) \approx u(\underline{z}^* + b^*) - u(\underline{z}^*)$. Thus, effort at the top does not converge to zero, but rather remains at the constant level implied by $\underline{z}^*$ and b^*, and the government essentially extracts all the surplus from high-ability workers with almost no distortions. This is reminiscent of Stantcheva (2014), where agency frictions in the labor market (in her case, adverse selection) also allowed the government to redistribute more than in the standard model. We believe that our result, while technically correct, is not policy-relevant, since such heavy taxation of performance-pay contracts would motivate firms to find different ways of incentivizing workers' effort (e.g., via monitoring), which we do not model explicitly. Fixing the tax schedule on base pay, on the other hand, yields a well-defined and non-corner optimum top tax rate on bonuses, as shown in Theorem 3.

eters, and are defined formally in the Appendix.

It is instructive to compare the optimal bonus tax rate to the benchmark Saez (2001) formula applied to top bonuses:

$$\frac{\tau_b}{1-\tau_b} = \frac{1-\mathcal{G}}{\rho_b\, e_B}. \tag{32}$$

The optimum formula (30) differs from this benchmark in three ways, which are analogous to those we found when studying the optimal joint tax rate on top earnings in Theorem 2.

First, the average marginal social welfare weight at the top is adjusted upwards due to the first-order welfare consequences of crowd-out: We show in the Appendix that, as long as t is not too high relative to τ_b, $\tilde{\mathcal{G}} > \mathcal{G}$. Intuitively, in response to an increase in the bonus tax rate, a share of the extra tax burden is endogenously transmitted to the unlucky workers who have not received a bonus. This lowers the welfare benefits of this tax increase.

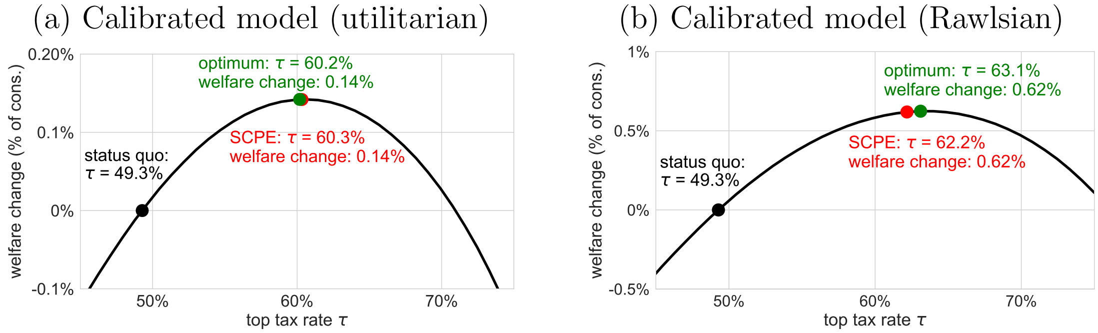
Second, a higher tax on bonuses affects the level of base pay according to the elasticity $e_{\underline{\mathcal{Z}}}$. This response can in turn be expressed as the difference between average crowd-out and crowd-in parameters, as shown in equation (31). If, for instance, the crowd-out dominates the crowd-in on average, then $e_{\underline{\mathcal{Z}}} > 0$ and a higher tax on bonuses results in lower base pay. In this case, the spillover to the base earnings tax base erodes the tax revenue and reduces the optimal bonus tax rate.

Third, the last term in the numerator of formula (30) is a corrective adjustment to the efficiency cost of taxation due to the frequency margin responses. A reduction in aggregate labor effort by $e_L > 0$, and hence in the frequency with which the worker receives a bonus, does not reduce government revenue as much if, rather than being taxed uniformly at rate τ_b, bonuses are taxed at a lower average rate up to b^*. Thus, if $t_b < \tau_b$, this correction tends to raise the optimal top tax rate.

2.5 Quantitative Analysis

We now adapt the quantitative framework with performance-pay and fixed-pay jobs of Section 1.4 to study the optimal taxation of top incomes and bonuses. We proceed in two steps. First, we evaluate the optimal top tax rate on total earnings, characterized theoretically in Section 2.3. Second, we evaluate the optimal top tax

Figure 4: Optimal Top Tax Rate on Total Income and SCPE



rate on bonuses for a given tax schedule on base earnings, as analyzed in Section 2.4.

Optimal Top Tax Rate on Total Earnings. We approximate the empirical income tax schedule with a piecewise linear function with two brackets and a lump-sum transfer. The top bracket starts at $z^* = \$500,000$ and applies to approximately the top 1 percent of earnings. The top tax rate in the status quo is equal to the marginal tax rate implied by the calibrated CRP tax function at z^*. We choose the bottom tax rate and the lump-sum transfer by matching the aggregate tax revenue and the average tax rate at z^* implied by the calibrated CRP tax. We obtain a top tax rate of 49 percent, a bottom tax rate of 41 percent, and a lump-sum transfer of \$16,460. This gives us a well-defined notion of top tax rate that we will choose optimally.

We compute the optimal top tax rate under two social objectives: Rawlsian and utilitarian. The former maximizes the extraction of revenue from top earners, i.e., it sets the average marginal welfare weight at the top $\tilde{\mathcal{G}}$ to zero in (27). We compare the optimal top rates with the rates obtained if the policymaker ignored the endogeneity of earnings risk (Self-Confirming Policy Equilibrium, SCPE). More specifically, a policymaker in the SCPE ignores the impact of changing the top tax rate on earnings in the lower bracket and calculates the welfare impact of reforms without accounting for the crowd-out of private insurance.[40] The results are depicted in Figure 4.

[40]Here we ignore the third discrepancy between the standard and the optimal tax formula, which comes from the fact that the frequency responses are taxed at a different average tax rate than the level responses (see Theorem 2). Such a policy mistake would arise if the policymaker in the SCPE tracked only the aggregate earnings in the top bracket. Instead, we are implicitly assuming that the policymaker can observe individual earnings and, hence, correctly calculate individual tax burdens.

We find that the top tax rates chosen in the SCPE and in the optimum are almost identical: They differ by less than 1 percentage point. The welfare cost of such a policy mistake is minuscule. Furthermore, and unlike the case of the optimal rate of progressivity analyzed in Section 1.4, we verify that the welfare cost of ignoring endogenous earnings risk would not be much different if all workers had performance-pay contracts. This is because performance-pay contracts are already widely adopted at the top: They account for 76 percent of earnings in the top bracket.

The right panel of Figure 4 shows that the planner who attempts to maximize tax revenue (Rawlsian), but does not account for the impact of the top tax rate on earnings in the lower bracket (SCPE), chooses a top rate that is slightly too low. That is because a higher top rate leads to higher earnings (and, hence, higher tax revenue) in the lower bracket. To understand why this is the case, focus on the intermediate earners whose base pay is in the lower bracket and whose high-level pay is in the top bracket; this group accounts for 90 percent of the performance-pay workers who are affected by the top tax rate. By Proposition 3, a higher top tax rate affects their base pay via two channels: the crowd-out, which reduces it, and the crowd-in, which works in the opposite direction. We quantify these effects in Table 2. While the crowd-out is sizable, and on its own would imply a 4 percent fall in base pay, the crowd-in has an almost equal magnitude and offsets most of this change. As a result, the base pay level falls only by 0.5 percent on average.[41] In addition, the mean frequency at which solely the base pay is paid out increases by 1.7 percent (second column of Table 2). Taking into account these level and frequency changes, we obtain that the overall base earnings of intermediate earners increase, leading to higher tax revenue from spillovers to the lower bracket.

In contrast to the Rawlsian case, the utilitarian policymaker who ignores the endogeneity of private insurance (SCPE) would choose a top tax rate that is marginally too high. Effectively, this planner makes two mistakes: It underestimates the tax revenue from the lower tax bracket(similarly to the Rawlsian planner), and it also underestimates the negative welfare effect on high-income individuals—this second mistake is new. Focus again on the intermediate earners. A naive planner would believe that the tax hike affects only the lucky intermediate workers who receive a

[41] There is a third effect at play: The tax reform also affects the relative probabilities with which different intermediate earners receive their base pay; this in turn modifies how the average base pay among these workers is computed in the first column of Table 2. This effect, however, is small relative to the crowd-out and crowd-in (0.1 percent).

Table 2: Impact of top tax rate on base pay of intermediate earners

	level $\underline{z}$	frequency $1-\pi(\ell)$	overall $\underline{z}\cdot(1-\pi(\ell))$
Status quo (initial)	\$264,360	0.76	\$202,210
+ crowd-out	-4.0%		
+ crowd-in	$+3.4\%$		
+ change of relative freq.	$+0.1\%$		
Rawlsian optimum (final)	\$263,040 (-0.5%)	0.78 ($+1.7\%$)	\$204,630 ($+1.2\%$)

Note: Top tax rate is increased from 49.3% (status quo) to 63.1% (Rawlsian optimum). All statistics are averages over intermediate earners.

Table 3: Impact of top tax rate on earnings risk

	intermediate earners		**top earners**	
	bonus rate β	$Var(\log z \mid \theta)$	bonus rate β	$Var(\log z \mid \theta)$
Status quo (initial)	1.05	0.2	1.1	0.22
+ crowd-out	$+9\%$	$+18\%$	$+16\%$	$+35\%$
+ crowd-in	-18%	-38%	-10%	-26%
Rawlsian optimum (final)	0.95 (-9%)	0.16 (-20%)	1.17 ($+6\%$)	0.24 ($+9\%$)

Note: Top tax rate is increased from 49.3% (status quo) to 63.1% (Rawlsian optimum). All statistics are averages over either intermediate or top earners.

bonus. However, the resulting crowd-out of private insurance contributes to a lower level of base pay and, thus, also negatively affects the unlucky intermediate workers. Although this change in base earnings is mostly offset by the crowd-in effect, its welfare impact remains significant due to the envelope theorem. The negative welfare impact on top earners, whose base pay is in the top bracket, is similarly underestimated. In our quantitative model, the mistake in the welfare effect dominates that in the revenue from the lower bracket, leading to a tax rate that is too high relative to the optimum.

Finally, Table 3 shows how increasing the top tax rate influences the earnings risk of affected workers. Starting from top earners, who earn both their base pay and their high-level pay in the top bracket, we find that the crowd-in offsets most (63–75 percent) of the crowd-out when measured either by the mean bonus rate $\beta = \log(\bar{z}/\underline{z})$ or by the mean variance of log earnings conditional on ability. As a result, a strong increase in the top tax rate raises the earnings risk of top earners, but much less than the crowd-out alone would suggest.

The picture is quite different for intermediate workers: for them the crowd-in is twice as large as the crowd-out. As a result, earnings risk actually falls after an increase in the top tax rate. To understand this, recall that the crowd-in effect is driven by the endogenous labor effort adjustment. Lemma 7 implies that intermediate earners adjust their labor effort more than do top earners. This is because intermediate earners experience an increase in the marginal and the average tax rate on their high-level pay, both of which reduce effort. In addition, top earners face an increase in the marginal and average tax rates on their base pay, which mitigates their labor effort responses. Quantitatively, the mean earnings of intermediate workers fall on average by 5.6 percent, while those of top workers fall by 3.6 percent, leading to a much more powerful crowd-in for the former group.

These results highlight that (i) the crowd-in is an important force that we need to take into account; and (ii) the relative size of the crowd-out and crowd-in depends on the specifics of the tax reform. In particular, strongly progressive reforms (such as the one experienced by intermediate earners) lead to a stronger effort response and magnify the crowd-in effect.

Separate Taxation of Top Bonuses. In Section 2.4 we showed that there are efficiency gains from taxing bonuses at a lower rate than base pay and characterized the optimal top bonus tax. In this section we quantify these results. In particular, we compare our economy in which all earnings are taxed jointly with a single tax schedule to the economy in which top bonuses are taxed with a separate rate that is chosen optimally, taking as given the tax on base pay. For simplicity, we assume that the original joint tax schedule is affine, i.e., it consists of a lump-sum transfer and a constant marginal tax rate.[42] We then suppose that a policymaker can adjust the tax rate on the top 5 percent of bonuses, taking the rest of the tax schedule as given. The results are shown in Table 4.

We find that the optimal bonus tax rate can be substantially lower than the tax on base earnings, and that such a separation can bring sizable welfare benefits, particularly when the initial joint tax rate is high. For instance, consider an initial joint tax rate that is chosen optimally according to a utilitarian or Rawlsian social welfare criterion. In the case of utilitarian social preferences, the optimal tax rate

[42]The assumption that the initial joint tax schedule is affine allows us to represent the tax schedule, both before and after the reform, in the additively separable form: $T(\underline{z}, b) = T_{\underline{z}}(\underline{z}) + T_b(b)$, which significantly simplifies the analysis.

Table 4: Optimal Tax on Top Bonuses for a Given Base Pay Tax

	Utilitarian		**Rawlsian**	
joint tax rate (initial)	bonus tax rate	welfare gain	bonus tax rate	welfare gain
40%	36.8%	0.0%	38.9%	0.0%
60%	42.1%	0.3%	45.1%	0.4%
62%*	42.8%	0.4%	45.8%	0.5%
80%	51.3%	1.6%	54.5%	1.8%
84%**	54.5%	2.2%	57.5%	2.3%

Note: Welfare gain from the optimal separation of the tax on top 5% of bonuses (taking as given the tax on base pay and on bonuses below the top) is expressed as an equivalent percentage change of consumption. * Optimal Utilitarian joint tax rate. ** Optimal Rawlsian joint tax rate.

on top bonuses would then be 19 percentage points lower than that on the base pay (42.9 percent vs. 62 percent), with a welfare gain from separating the tax treatment of the two forms of pay equal to 0.4 percent of consumption. In the case of Rawlsian social preferences, it is optimal to cut the top bonus tax rate relative to the base pay tax rate by 27 percentage points (57.5 percent vs. 84 percent), with a welfare gain of 2.4 percent of consumption. Note that the welfare gains from separating the tax rates increase with the size of the initial tax rate. This is because, for a (sub-optimally) low initial joint tax rate, the efficiency gains from lowering the tax on bonuses are countervailed by the redistributive losses from smaller taxes on the bonuses of the most productive agents.

Conclusion

We have set up and analyzed a tractable moral-hazard environment in which firms design labor contracts that trade off effort incentives with insurance against performance shocks. The government uses the tax-and-transfer system to redistribute income across workers who differ in uninsurable ex ante ability. The key feature of our model is that earnings risk is endogenous and has a productive role as it motivates labor effort. We found that standard models that ignore the endogeneity of pre-tax earnings risk come very close to accurately evaluating the impact of taxes on labor contracts. We derived optimal tax formulas for the overall rate of tax progressivity, the top tax rate on total earnings, and the top tax rate on bonuses. Perhaps surprisingly, we showed that it is optimal to tax bonuses at a strictly lower rate than base earnings.

It would be interesting to extend our analysis in several directions. First, we only considered the impact of taxes on compensation for already existing performance-pay jobs. One could also model the incentives of firms to create such performance-pay jobs (rather than monitored jobs) in the first place. Second, in our model, private markets are perfectly competitive and constrained efficient. In other words, we gave private markets their best chance of making government policy redundant. Introducing market power and frictions such as adverse selection in private markets, whereby firms cannot perfectly observe workers' ability are natural next steps. Third, our theoretical analysis delivers predictions regarding the impact of various types of tax reforms on the structure of incentive-based compensation. Testing these predictions empirically should be particularly fruitful.

A Proofs of Section 1

Proof of Proposition 1. Defining β by $\bar{z} = e^{\beta}\underline{z}$, the free-entry condition (4) can be expressed as

$$(1-\ell)\underline{z} + \ell e^{\beta}\underline{z} = \theta\ell.$$

This immediately leads to the solution (7) for the equilibrium values of the base pay $\underline{z}$ and high-performance pay $\bar{z}$. From a firm's viewpoint, the participation constraint (3) determines the base pay $\underline{z}$ given ℓ and the reservation value U. With log utility and a CRP tax schedule, it reads:

$$\log\frac{1-\tau}{1-p} + (1-p)[(1-\ell)\log(\underline{z}) + \ell\log(\bar{z})] - h(\ell) = U.$$

The bonus rate $\beta \equiv \log(\bar{z}) - \log(\underline{z})$ is given as a function of the desired effort level ℓ by the incentive constraint (6). Substituting $\beta = h'(\ell)/(1-p)$ into the previous equation, we get:

$$\log\frac{1-\tau}{1-p} + (1-p)\left[\log(\underline{z}) + \ell\frac{h'(\ell)}{1-p}\right] - h(\ell) = U.$$

Solving for $\underline{z}$ leads to:

$$\underline{z} = \left(\frac{1-\tau}{1-p}\right)^{-1/(1-p)} e^{\frac{1}{1-p}U} e^{\frac{1}{1-p}[h(\ell)-\ell h'(\ell)]} \tag{33}$$

and hence

$$\bar{z} = \left(\frac{1-\tau}{1-p}\right)^{-1/(1-p)} e^{\frac{1}{1-p}U} e^{\frac{1}{1-p}[h(\ell)+(1-\ell)h'(\ell)]}.$$

Substituting these expressions into the free-entry condition determines the reservation value U as a function of labor effort:

$$e^{\frac{1}{1-p}U} = \left(\frac{1-\tau}{1-p}\right)^{1/(1-p)} e^{-\frac{1}{1-p}[h(\ell)-\ell h'(\ell)]} \frac{\theta\ell}{1-\ell+\ell e^{\frac{1}{1-p}h'(\ell)}},$$

i.e.,

$$U = \log\left(\frac{1-\tau}{1-p}\right) + (1-p)\log(\theta\ell) - h(\ell) + (1-p)[\ell\beta - \log(1-\ell+\ell e^{\beta})].$$

Noting that $u(\mathbb{E}z) = \log(\theta\ell)$ and $\mathbb{E}[u(z)] = (1-\ell)\log \underline{z} + \ell \log(e^{\beta}\underline{z}) = \log \underline{z} + \ell\beta$, and using expression (7), we obtain:

$$U = v(\theta\ell) - h(\ell) + (1-p)\{\mathbb{E}[u(z)] - u(\mathbb{E}z)\}.$$

Finally, the first-order condition for effort is obtained by differentiating the firm's expected profit $\theta\ell - [\underline{z} + \ell b]$ and equating it to zero:

$$\theta = b + \frac{\partial \underline{z}}{\partial \ell} + \ell \frac{\partial b}{\partial \ell} = b + \frac{1}{1-p}\ell h''(\ell)\underline{z}e^{\frac{1}{1-p}h'(\ell)} + [1 - \ell + \ell e^{\frac{1}{1-p}h'(\ell)}]\frac{\partial \underline{z}}{\partial \ell},$$

where the second equality follows from the fact that $b = (e^{\beta} - 1)\underline{z}$ by definition, with $\beta = \frac{1}{1-p}h'(\ell)$ since the contract must respect the incentive constraint (6). Since the firm takes as given the worker's reservation utility U, we differentiate equation (33) to obtain:

$$\frac{\partial \underline{z}}{\partial \ell} = -\underline{z}\frac{1}{1-p}\ell h''(\ell).$$

Substituting into the previous expression gives

$$\theta = b + [e^{\frac{1}{1-p}h'(\ell)} - 1]\underline{z}\frac{1}{1-p}\ell(1-\ell)h''(\ell),$$

which leads to (8). Note that in equilibrium, we can use (7) to rewrite this equation as

$$1 = \frac{\ell(e^{\beta} - 1)}{1 + \ell(e^{\beta} - 1)}\left[1 + \frac{1}{1-p}\ell(1-\ell)h''(\ell)\right],$$

and hence

$$\ell(1-\ell) = \frac{1}{\beta(e^{\beta} - 1)}\frac{h'(\ell)}{\ell h''(\ell)}.$$

This expression shows that the optimal effort level ℓ is independent of the worker's productivity θ. ■

Proof of Lemma 1. The first order condition (8) for labor effort can be rewritten as $1 - p = \ell^2(1-\ell)h''(\ell)[e^{h'(\ell)/(1-p)} - 1]$. Apply the implicit function theorem to get:

$$\varepsilon_{\ell,1-p} \equiv \frac{1-p}{\ell}\frac{\partial \ell}{\partial(1-p)} = \frac{1 + \frac{\beta e^{\beta}}{e^{\beta}-1}}{\frac{2-3\ell}{1-\ell} + \frac{\beta e^{\beta}}{e^{\beta}-1}\frac{\ell h''(\ell)}{h'(\ell)} + \frac{\ell h'''(\ell)}{h''(\ell)}}.$$

Recall that

$$\frac{\partial \Pi}{\partial \ell} = \theta - b\Big[1 + \frac{\ell(1-\ell)h''(\ell)}{1-p}\Big].$$

Differentiating this expression using $\frac{\partial \underline{z}}{\partial \ell} = -\ell \underline{z}\frac{h''(\ell)}{1-p}$ and $\frac{\partial b}{\partial \ell} = [\underline{z} + (1-\ell)b]\frac{h''(\ell)}{1-p}$ leads to

$$\frac{\partial^2 \Pi}{\partial \ell^2} = -[\underline{z} + (2-3\ell)b]\frac{h''(\ell)}{1-p} - \ell(1-\ell)[\underline{z} + (1-\ell)b]\Big(\frac{h''(\ell)}{1-p}\Big)^2 - b\ell(1-\ell)\frac{h'''(\ell)}{1-p}.$$

The second-order condition for optimal labor effort, $\frac{\partial^2 \Pi}{\partial \ell^2} \leq 0$, can therefore be expressed as

$$\Big[\ell\beta\frac{\ell h''(\ell)}{h'(\ell)} - \frac{1}{1-\ell}\frac{1}{e^\beta - 1}\Big] - \Big[\frac{2-3\ell}{1-\ell} + \frac{\beta e^\beta}{e^\beta - 1}\frac{\ell h''(\ell)}{h'(\ell)} + \frac{\ell h'''(\ell)}{h''(\ell)}\Big] \leq 0,$$

where we used the fact that $\underline{z}/b = 1/(e^\beta - 1)$. The first-order condition for labor effort implies that the first square bracket is equal to zero. Therefore, we obtain $\varepsilon_{\ell,1-p} > 0$. Now, suppose that the disutility of effort is isoelastic, $h(\ell) = \frac{\ell^{1+1/\varepsilon_\ell^F}}{1+1/\varepsilon_\ell^F}$ with ε_ℓ^F constant. We can then rewrite the labor effort elasticity as

$$\varepsilon_{\ell,1-p} = \frac{\varepsilon_\ell^F}{1 + \frac{1-\ell/(1-\ell)}{1+\beta e^\beta/(e^\beta - 1)}\varepsilon_\ell^F}.$$

This expression implies that $\varepsilon_{\ell,1-p} > \frac{\varepsilon_\ell^F}{1+\varepsilon_\ell^F}$, and that $\varepsilon_{\ell,1-p} > \varepsilon_\ell^F$ if and only if $1 - \frac{\ell}{1-\ell} < 0$, i.e., $\ell > \frac{1}{2}$. ■

Proof of Lemma 2. This result follows immediately from equation (6) and Lemma 1. ■

Proof of Lemma 3. Differentiating equation (9) gives

$$\frac{\partial U(\theta)}{\partial(1-p)} = \Big[-\frac{1}{1-p} + \log(\ell\theta) + \Big(\frac{1-p}{\ell} - h'(\ell)\Big)\frac{\partial \ell}{\partial(1-p)}\Big] + (\log(\mathbb{E}z) - \mathbb{E}[\log z])$$
$$- (1-p)\Big\{\Big[\frac{e^\beta - 1}{1+\ell(e^\beta - 1)} - \beta\Big]\frac{\partial \ell}{\partial(1-p)} + \frac{(e^\beta - 1)\ell(1-\ell)}{1+\ell(e^\beta - 1)}\frac{\partial \beta}{\partial(1-p)}\Big\}.$$

But recall that

$$\frac{d\mathcal{U}(\theta)}{d(1-p)} = -\frac{1}{1-p} + \log(\ell\theta)$$

and that

$$\frac{\partial\beta}{\partial(1-p)} = \frac{\beta}{1-p}[\varepsilon_{\beta,\ell}\varepsilon_{\ell,1-p} + \varepsilon_{\beta,1-p}] = \frac{\partial\beta}{\partial\ell}\frac{\partial\ell}{\partial(1-p)} + \frac{\beta}{1-p}\varepsilon_{\beta,1-p}.$$

Substituting into the previous equation and using $\frac{\partial\beta}{\partial\ell} = \frac{\beta}{\ell}\frac{\ell h''(\ell)}{h'(\ell)}$ and $\frac{e^{\beta}-1}{1+\ell(e^{\beta}-1)} = \frac{b}{\mathbb{E}z}$ leads to

$$\begin{aligned}\frac{\partial U(\theta)}{\partial(1-p)} &= \frac{d\mathcal{U}(\theta)}{d(1-p)} + (\log(\mathbb{E}z) - \mathbb{E}[\log z]) - \frac{b}{\mathbb{E}z}\beta\ell(1-\ell)\varepsilon_{\beta,1-p} \\ &+ \frac{1-p}{\ell}\left[1 - \frac{\ell h'(\ell)}{1-p} - \frac{\ell(e^{\beta}-1)}{1+\ell(e^{\beta}-1)} + \beta\ell - \frac{\beta(e^{\beta}-1)\ell(1-\ell)}{1+\ell(e^{\beta}-1)}\frac{\ell h''(\ell)}{h'(\ell)}\right]\frac{\partial\ell}{\partial(1-p)}.\end{aligned}$$

Using the first-order condition for labor effort $\beta(e^{\beta}-1)\ell(1-\ell)\frac{\ell h''(\ell)}{h'(\ell)} = 1$ derived in the proof of Proposition 1, we obtain that the term in square brackets that multiplies $\frac{\partial\ell}{\partial(1-p)}$ is equal to zero. This is a manifestation of the envelope theorem in our setting. This easily yields expression (10). ■

Proof of Theorem 1. The proof proceeds in several steps. We first derive the effect of a change in progressivity on the social objective. Second, we evaluate its impact on government revenue by decomposing it into a statutory effect, a standard behavioral effect with exogenous private insurance, and fiscal externalities due to crowd-out and crowd-in. Third, we compute the marginal value of public funds. Finally, we equate the sum of these effects to zero to obtain our characterization of optimal tax progressivity.

Social Welfare Effect. Denote the change in the social welfare objective resulting from a change in tax progressivity by

$$WE = \int \alpha(\theta)\frac{\partial U(\theta)}{\partial(1-p)}dF(\theta).$$

By equation (10), we have

$$WE = -\frac{1}{1-p} + \int \alpha(\theta)\log(\ell\theta)dF(\theta) + \log(\mathbb{E}z) - \mathbb{E}[\log z] - \frac{b}{\mathbb{E}z}\beta\ell(1-\ell)\varepsilon_{\beta,1-p},$$

where $\log(\mathbb{E}z) - \mathbb{E}[\log z] = \log(1+\ell(e^{\beta}-1)) - \beta\ell$. Now suppose that $\log\theta \sim \mathcal{N}(\mu_\theta, \sigma_\theta^2)$ and $\alpha(\theta) \propto e^{-a\log\theta}$. Note that, if a random variable x is normally distributed with mean μ and variance σ^2, we have $\mathbb{E}[e^{-ax}] = e^{-a\mu+\frac{1}{2}a^2\sigma^2}$. Moreover, letting φ denote the pdf of x, we have $\varphi'(x) = -\frac{x-\mu}{\sigma^2}\varphi(x)$, so that

$$\mathbb{E}[(x-\mu)e^{-ax}] = \int (x-\mu)e^{-ax}\varphi(x)dx = -\sigma^2\int e^{-ax}\varphi'(x)dx.$$

An integration by parts implies that this expression is equal to $-a\sigma^2\int e^{-ax}\varphi(x)dx = -a\sigma^2 e^{-a\mu+\frac{1}{2}a^2\sigma^2}$. Therefore, we obtain $\mathbb{E}[xe^{-ax}] = (\mu - a\sigma^2)e^{-a\mu+\frac{1}{2}a^2\sigma^2}$. Hence,

$$\int \alpha(\theta)\log(\ell\theta)dF(\theta) = \log\ell + \frac{\int e^{-a\log\theta}\log\theta f(\theta)d\theta}{\int e^{-a\log\theta}f(\theta)d\theta} = \log\ell + \mu_\theta - a\sigma_\theta^2.$$

Statutory Revenue Effect. Government revenue is equal to

$$\int \mathbb{E}[T(z)]dF(\theta) = \int [(1-\ell)T(\underline{z}) + \ell T(\bar{z})]dF(\theta) = Z - C,$$

where $Z \equiv \int \mathbb{E}[z]dF(\theta)$ is aggregate income, and $C \equiv \int \mathbb{E}[R(z)]dF(\theta)$ is aggregate consumption. Under the assumptions of Theorem 1, we have

$$C = \frac{1-\tau}{1-p}\frac{1+\ell(e^{(1-p)\beta}-1)}{[1+\ell(e^{\beta}-1)]^{1-p}}\int(\theta\ell)^{1-p}dF(\theta), \tag{34}$$

with

$$\int(\theta\ell)^{1-p}dF(\theta) = \left[(1-p)\mu_\theta + \frac{1}{2}(1-p)^2\sigma_\theta^2\right]\ell^{1-p}.$$

The statutory (or mechanical) effect is obtained by evaluating the change in government revenue following a change in progressivity keeping the contract $(\ell, \underline{z}, \bar{z})$ and hence β fixed, that is,

$$ME = \int \left.\frac{\partial\mathbb{E}[T(z)]}{\partial(1-p)}\right|_{\ell,\underline{z},\bar{z}} dF(\theta)$$

We obtain:

$$ME = \left[\frac{1}{1-p} - \frac{\beta \ell e^{(1-p)\beta}}{1+\ell(e^{(1-p)\beta}-1)} + \log[1+\ell(e^{\beta}-1)] - \frac{\partial \log \int (\theta\ell)^{1-p} dF(\theta)}{\partial(1-p)}\right] C,$$

with

$$\begin{aligned}\frac{\partial}{\partial(1-p)} \log \int (\theta\ell)^{1-p} dF(\theta) &= \log \ell + \frac{\partial}{\partial(1-p)} \log \int \theta^{1-p} dF(\theta) \\ &= \log \ell + \mu_\theta + (1-p)\sigma_\theta^2.\end{aligned}$$

Behavioral Effect with Exogenous Private Insurance. By equation (7), in response to a change in progressivity, the income levels change (in percentage terms) by

$$\frac{\partial \log \underline{z}}{\partial \log(1-p)} = \frac{\underline{z}}{\mathbb{E}z}\varepsilon_{\ell,1-p} - \beta\ell\frac{\bar{z}}{\mathbb{E}z}(\varepsilon_{\beta,1-p} + \varepsilon_{\beta,\ell}\varepsilon_{\ell,1-p}),$$

where we used the fact that $1 - \frac{\ell b}{\mathbb{E}z} = \frac{\underline{z}}{\mathbb{E}z}$, and

$$\frac{\partial \log \bar{z}}{\partial \log(1-p)} = \frac{\underline{z}}{\mathbb{E}z}\varepsilon_{\ell,1-p} + \beta(1-\ell)\frac{\underline{z}}{\mathbb{E}z}(\varepsilon_{\beta,1-p} + \varepsilon_{\beta,\ell}\varepsilon_{\ell,1-p}).$$

The standard behavioral effect of an increase in $1-p$ is equal to the change in government revenue triggered by labor effort responses ℓ only – that is, keeping the bonus rate β fixed. We get[43]

$$\begin{aligned}BE =& \frac{1}{1-p}\int \left(\mathbb{E}\left[T'(z)z\ \frac{\partial \log \underline{z}}{\partial \log(1-p)}\bigg|_{\beta}\right] + (T(\bar{z}) - T(\underline{z}))\ell\varepsilon_{\ell,1-p}\right) dF(\theta) \\ =& \frac{1}{1-p}\int \mathbb{E}[T'(z)z]\frac{\underline{z}}{\mathbb{E}z}\varepsilon_{\ell,1-p} dF(\theta) + \int \ell(T(\bar{z}) - T(\underline{z}))\varepsilon_{\ell,1-p} dF(\theta).\end{aligned}$$

Since $\varepsilon_{\ell,1-p}$ and $\frac{\underline{z}}{\mathbb{E}z}$ are constant (independent of θ), this expression can be rewritten

[43]Note that, in a model with only intensive-margin responses to taxes, i.e., with an exogenous probability π of earning the bonus, the free-entry condition $\mathbb{E}z \equiv (1-\pi)\underline{z} + \pi\bar{z} = \ell\theta$ would imply $\frac{\partial \log \underline{z}}{\partial \log(1-p)} = \frac{\partial \log \bar{z}}{\partial \log(1-p)} = \varepsilon_{\ell,1-p}$, and the change in government revenue caused by a change in progressivity would be equal to $\varepsilon_{\ell,1-p} \int \mathbb{E}[T'(z)z] dF(\theta)$. This is the expression we would obtain, for instance, in the full-insurance benchmark.

as:

$$BE = \frac{1}{1-p}\left[\frac{\underline{z}}{\mathbb{E}z}\int \mathbb{E}[T'(z)z]dF(\theta) + \ell\int (T(\bar{z}) - T(\underline{z}))dF(\theta)\right]\varepsilon_{\ell,1-p}.$$

With a CRP tax schedule, we can write

$$\int \mathbb{E}[T'(z)z]dF(\theta) = \int \mathbb{E}[z-(1-\tau)z^{1-p}]dF(\theta) = Z-(1-p)C.$$

The post-tax bonus rate is equal to $\log\frac{\bar{c}}{\underline{c}} = \log\frac{\frac{1-\tau}{1-p}\bar{z}^{1-p}}{\frac{1-\tau}{1-p}\underline{z}^{1-p}} = (1-p)\beta$. Hence, writing $\mathbb{E}c = (1-\ell)\underline{c} + \ell e^{(1-p)\beta}\underline{c}$ leads to $\frac{1}{1+\ell(e^{(1-p)\beta}-1)} = \frac{\underline{c}}{\mathbb{E}c}$ and $\frac{e^{(1-p)\beta}}{1+\ell(e^{(1-p)\beta}-1)} = \frac{\bar{c}}{\mathbb{E}c}$. Therefore, $\frac{b}{\mathbb{E}z}$ and $\frac{\gamma}{\mathbb{E}c}$ are constant, where $\gamma \equiv \bar{c}-\underline{c}$. We can thus write the contribution of extensive margin adjustments to the excess burden of the rise in progressivity as follows:

$$\begin{aligned}\int (T(\bar{z}) - T(\underline{z}))dF(\theta) &= \int \left[\left(\bar{z} - \frac{1-\tau}{1-p}\bar{z}^{1-p}\right) - \left(\underline{z} - \frac{1-\tau}{1-p}\underline{z}^{1-p}\right)\right]dF(\theta)\\ &= \int b dF(\theta) - \int \gamma dF(\theta) = \frac{b}{\mathbb{E}z}\int \mathbb{E}z dF(\theta) - \frac{\gamma}{\mathbb{E}c}\int \mathbb{E}c dF(\theta) = \frac{b}{\mathbb{E}z}Z - \frac{\gamma}{\mathbb{E}c}C.\end{aligned}$$

Collecting terms, and using the fact that $\ell\frac{\gamma}{\mathbb{E}c} = 1 - \frac{\underline{c}}{\mathbb{E}c}$, we obtain

$$\begin{aligned}BE =& \frac{1}{1-p}\left[\frac{\underline{z}}{\mathbb{E}z}Z - (1-p)\frac{\underline{z}}{\mathbb{E}z}C + \ell\frac{b}{\mathbb{E}z}Z - \ell\frac{\gamma}{\mathbb{E}c}C\right]\varepsilon_{\ell,1-p}\\ =& -\frac{1}{1-p}\left[1 - \frac{Z}{C} + (1-p)\frac{\underline{z}}{\mathbb{E}z} - \frac{\underline{c}}{\mathbb{E}c}\right]\varepsilon_{\ell,1-p}C.\end{aligned}$$

Fiscal Externalities from Crowd-Out and Crowd-In. Finally, the change in government revenue due to the endogeneity of the bonus rate β, keeping effort ℓ fixed, is given by

$$\begin{aligned}FE =& \frac{1}{1-p}\int\left[(1-\ell)T'(\underline{z})\underline{z}\left.\frac{\partial\log\underline{z}}{\partial\log(1-p)}\right|_{\ell} + \ell T'(\bar{z})\bar{z}\left.\frac{\partial\log\bar{z}}{\partial\log(1-p)}\right|_{\ell}\right]dF(\theta)\\ =& \frac{1}{1-p}[\varepsilon_{\beta,1-p} + \varepsilon_{\beta,\ell}\varepsilon_{\ell,1-p}]\beta\ell(1-\ell)\left[\frac{\underline{z}}{\mathbb{E}z}\int T'(\bar{z})\bar{z}dF(\theta) - \frac{\bar{z}}{\mathbb{E}z}\int T'(\underline{z})\underline{z}dF(\theta)\right],\end{aligned}$$

where the second equality uses the expressions derived above for the earnings elasticities. The term in square brackets can be rewritten as

$$
\begin{aligned}
&\frac{\underline{z}}{\mathbb{E}z}\int[\bar{z}-(1-\tau)\bar{z}^{1-p}]dF(\theta)-\frac{\bar{z}}{\mathbb{E}z}\int[\underline{z}-(1-\tau)\underline{z}^{1-p}]dF(\theta)\\
=&(1-\tau)\frac{1}{1+\ell(e^{\beta}-1)}\frac{e^{\beta}-e^{(1-p)\beta}}{[1+\ell(e^{\beta}-1)]^{1-p}}\int(\theta\ell)^{1-p}dF(\theta)\\
=&\frac{1}{1-\ell}(1-p)\left[\frac{e^{\beta}}{1+\ell(e^{\beta}-1)}-\frac{e^{(1-p)\beta}}{1+\ell(e^{(1-p)\beta}-1)}\right]C,
\end{aligned}
$$

where the last equality follows from the expression (34) for C derived above. Thus, we obtain

$$FE=\beta\ell\left[\frac{e^{\beta}}{1+\ell(e^{\beta}-1)}-\frac{e^{(1-p)\beta}}{1+\ell(e^{(1-p)\beta}-1)}\right][\varepsilon_{\beta,1-p}+\varepsilon_{\beta,\ell}\varepsilon_{\ell,1-p}]C.$$

Marginal Value of Public Funds. The marginal value of public funds λ, when the tax code is restricted to the CRP class, is defined by the effect on social welfare of an increase the tax parameter τ, normalized to raise government revenue by 1 dollar. At the optimum tax schedule, λ is the Lagrange multiplier of the government budget constraint (12). We have

$$\frac{\partial\int\mathbb{E}[T(z)]dF(\theta)}{\partial\tau}=\frac{\partial Z}{\partial\tau}-\frac{\partial C}{\partial\tau}.$$

The first-order condition for effort (8) implies that $\frac{\partial\ell}{\partial\tau}=0$. Thus, $\frac{\partial Z}{\partial\tau}=0$ and, using expression (34), $\frac{\partial C}{\partial\tau}=-\frac{C}{1-\tau}$. Hence, we obtain

$$\frac{\partial\int\mathbb{E}[T(z)]dF(\theta)}{\partial\tau}=\frac{C}{1-\tau}.$$

The impact on social welfare of the normalized tax change is given by

$$\lambda=\left(\frac{C}{1-\tau}\right)^{-1}\int\alpha(\theta)\frac{\partial U(\theta)}{\partial\tau}dF(\theta)=\left(\frac{C}{1-\tau}\right)^{-1}\int\alpha(\theta)\frac{1}{1-\tau}dF(\theta)=\frac{1}{C}.$$

Optimal Rate of Progressivity. The optimal rate of progressivity is the solution to

$$\begin{aligned} 0 &= \frac{\partial \int \alpha(\theta)U(\theta)dF(\theta)}{\partial(1-p)} + \lambda \frac{\partial \int \mathbb{E}[T(z)]dF(\theta)}{\partial(1-p)} \\ &= WE + \frac{1}{C}[ME + BE + FE]. \end{aligned}$$

That is, the optimal level of p satisfies

$$\begin{aligned} 0 =& -\frac{1}{1-p} + \log \ell + \mu_\theta - a\sigma_\theta^2 - \log(1+\ell(e^\beta - 1)) + \beta\ell - \frac{b}{\mathbb{E}z}\beta\ell(1-\ell)\varepsilon_{\beta,1-p} \\ &+ \frac{1}{1-p} - \frac{\beta\ell e^{(1-p)\beta}}{1+\ell(e^{(1-p)\beta}-1)} + \log[1+\ell(e^\beta - 1)] - (\log \ell + \mu_\theta + (1-p)\sigma_\theta^2) \\ &- \frac{1}{1-p}\left[1 - \frac{Z}{C} + (1-p)\frac{\underline{z}}{\mathbb{E}z} - \frac{\underline{c}}{\mathbb{E}c}\right]\varepsilon_{\ell,1-p} \\ &+ \beta\ell\left[\frac{e^\beta}{1+\ell(e^\beta-1)} - \frac{e^{(1-p)\beta}}{1+\ell(e^{(1-p)\beta}-1)}\right][\varepsilon_{\beta,1-p} + \varepsilon_{\beta,\ell}\varepsilon_{\ell,1-p}]. \end{aligned}$$

Rearranging terms, this formula can be rewritten as

$$\begin{aligned} 0 =& -(1-p+a)\sigma_\theta^2 - \beta\ell(1-\ell)\frac{e^{(1-p)\beta}-1}{1+\ell(e^{(1-p)\beta}-1)} \\ &- \frac{1}{1-p}\left[1 - \frac{Z}{C} + (1-p)\frac{\underline{z}}{\mathbb{E}z} - \frac{\underline{c}}{\mathbb{E}c}\right]\varepsilon_{\ell,1-p} - \beta\ell(1-\ell)\frac{e^{(1-p)\beta}-1}{1+\ell(e^{(1-p)\beta}-1)}\varepsilon_{\beta,1-p} \\ &+ \beta\ell\left[\frac{e^\beta}{1+\ell(e^\beta-1)} - \frac{e^{(1-p)\beta}}{1+\ell(e^{(1-p)\beta}-1)}\right]\varepsilon_{\beta,\ell}\varepsilon_{\ell,1-p}. \end{aligned}$$

We saw that $\frac{1}{1+\ell(e^\beta-1)} = \frac{\underline{z}}{\mathbb{E}z}$, $\frac{e^\beta}{1+\ell(e^\beta-1)} = \frac{\bar{z}}{\mathbb{E}z}$, $\frac{1}{1+\ell(e^{(1-p)\beta}-1)} = \frac{\underline{c}}{\mathbb{E}c}$, and $\frac{e^{(1-p)\beta}}{1+\ell(e^{(1-p)\beta}-1)} = \frac{\bar{c}}{\mathbb{E}c}$. We can therefore rewrite the optimal tax equation as (recall that $\gamma \equiv \bar{c} - \underline{c}$)

$$\begin{aligned} & (1-p+a)\sigma_\theta^2 + \beta\ell(1-\ell)\frac{\gamma}{\mathbb{E}c}(1+\varepsilon_{\beta,1-p}) \\ =& \left[\frac{1}{1-p}\frac{Z}{C} - \left(\frac{\underline{z}}{\mathbb{E}z} + \frac{\ell}{1-p}\frac{\gamma}{\mathbb{E}c}\right)\right]\varepsilon_{\ell,1-p} + \beta\ell\left(\frac{\bar{z}}{\mathbb{E}z} - \frac{\bar{c}}{\mathbb{E}c}\right)\varepsilon_{\beta,\ell}\varepsilon_{\ell,1-p}. \end{aligned}$$

Recall that $\mathbb{V}(\log z) = \beta^2\ell(1-\ell)$, $\frac{\bar{z}}{\mathbb{E}z} - \frac{\bar{c}}{\mathbb{E}c} = (1-\ell)(\frac{b}{\mathbb{E}z} - \frac{\gamma}{\mathbb{E}c})$, and $\frac{Z}{C} = 1 + \frac{G}{Z-G} \equiv 1 + \frac{g}{1-g}$

where g is the ratio of public expenditures G to output $Z = C + G$. We thus get

$$
\begin{aligned}
&(1-p+a)\sigma_\theta^2 + \mathbb{V}(\log z)\frac{1}{\beta}\frac{\gamma}{\mathbb{E}c}(1+\varepsilon_{\beta,1-p}) \\
&= \left[\frac{g/(1-g)+p}{1-p} + \ell\left(\frac{b}{\mathbb{E}z} - \frac{1}{1-p}\frac{\gamma}{\mathbb{E}c}\right)\right]\varepsilon_{\ell,1-p} + \mathbb{V}(\log z)\frac{1}{\beta}\left(\frac{b}{\mathbb{E}z} - \frac{\gamma}{\mathbb{E}c}\right)\varepsilon_{\beta,\ell}\varepsilon_{\ell,1-p}.
\end{aligned}
$$

Dividing through by $(1-p)$ and rearranging terms leads to

$$
\frac{p}{(1-p)^2} = \frac{(1+\frac{a}{1-p})\sigma_\theta^2 + \mathbb{V}(\log z)\frac{1}{(1-p)\beta}\frac{\gamma}{\mathbb{E}c}(1+\varepsilon_{\beta,1-p})}{\varepsilon_{\ell,1-p}\left[(1+\frac{g}{(1-g)p}) + \frac{1-p}{p}\ell(\frac{b}{\mathbb{E}z} - \frac{1}{1-p}\frac{\gamma}{\mathbb{E}c})\right] + \mathbb{V}(\log z)\frac{1-p}{\beta p}(\frac{b}{\mathbb{E}z} - \frac{\gamma}{\mathbb{E}c})\varepsilon_{\beta,\ell}\varepsilon_{\ell,1-p}}
$$

Note that, to a second order as $\beta \to 0$ (keeping ℓ fixed), we get

$$
\begin{aligned}
\kappa_1\mathbb{V}(\log z) &= \frac{\beta\ell(1-\ell)}{1-p}\frac{e^{(1-p)\beta}-1}{1+\ell(e^{(1-p)\beta}-1)} \sim \beta^2\ell(1-\ell) = \mathbb{V}(\log z) \\
\kappa_2\mathbb{V}(\log z) &= \beta\ell(1-\ell)\left[\frac{e^{\beta}-1}{1+\ell(e^{\beta}-1)} - \frac{e^{(1-p)\beta}-1}{1+\ell(e^{(1-p)\beta}-1)}\right] \sim p\mathbb{V}(\log z).
\end{aligned}
$$

Note that $\kappa_2 > 0$ if and only if $\frac{e^\beta-1}{1+\ell(e^\beta-1)} = \frac{e^{(1-p)\beta}-1}{1+\ell(e^{(1-p)\beta}-1)}$, which easily leads to $p > 0$.

Extension to a model with fixed-pay jobs. Let s_{pp} be the fraction of performance-pay (or "moral-hazard") jobs, and s_{fp} the fraction of fixed-pay jobs in the economy. The welfare effect becomes:

$$
\begin{aligned}
WE = &-\frac{1}{1-p} + s_{pp}(\log\ell_{pp} + \mu_{\theta,pp}) + (1-s_{pp})(\log\ell_{fp} + \mu_{\theta,fp}) \\
&+ s_{pp}\left[\log\mathbb{E}z_{pp} - \mathbb{E}\log z_{pp} - \frac{b_{pp}}{\mathbb{E}z_{pp}}\beta\ell_{pp}(1-\ell_{pp})\varepsilon_{\beta,1-p}\right].
\end{aligned}
$$

Aggregate consumption is equal to $C = s_{pp}C_{pp} + (1-s_{pp})C_{fp}$, with

$$
\begin{aligned}
C_{pp} =& \frac{1-\tau}{1-p}\frac{1+\ell_{pp}(e^{(1-p)\beta}-1)}{[1+\ell_{pp}(e^\beta-1)]^{1-p}}\left[(1-p)\mu_{\theta,pp} + \frac{1}{2}(1-p)^2\sigma_{\theta,pp}^2\right]\ell_{pp}^{1-p} \\
C_{fp} =& \frac{1-\tau}{1-p}\left[(1-p)\mu_{\theta,fp} + \frac{1}{2}(1-p)^2\sigma_{\theta,fp}^2\right]\ell_{fp}^{1-p}.
\end{aligned}
$$

The mechanical effect can then be written as

$$
\begin{aligned}
ME = \; & \frac{1}{1-p}C + \left[-\log \ell_{fp} - \mu_{\theta,fp} - (1-p)\sigma^2_{\theta,fp}\right](1-s_{pp})C_{fp} \\
+ & \left[-\frac{\beta \ell_{pp} e^{(1-p)\beta}}{1+\ell_{pp}(e^{(1-p)\beta}-1)} + \log(1+\ell_{pp}(e^{\beta}-1)) - \log \ell_{pp} - \mu_{\theta,pp} - (1-p)\sigma^2_{\theta,pp}\right] s_{pp}C_{pp}.
\end{aligned}
$$

The behavioral effect of the perturbation is equal to

$$
\begin{aligned}
BE = & \frac{1}{1-p}\left[\frac{Z_{fp}}{C_{fp}} - (1-p)\right]\varepsilon_{\ell_{fp},1-p}(1-s_{pp})C_{fp} \\
& - \frac{1}{1-p}\left[1 - \frac{Z_{pp}}{C_{pp}} + (1-p)\frac{\underline{z}_{pp}}{\mathbb{E}z_{pp}} - \frac{\underline{c}_{pp}}{\mathbb{E}c_{pp}}\right]\varepsilon_{\ell_{pp},1-p}s_{pp}C_{pp}
\end{aligned}
$$

where Z_i is the aggregate output of jobs of type i, and $\underline{z}_{pp}/\mathbb{E}z_{pp}$ and $\underline{c}_{pp}/\mathbb{E}c_{pp}$ are constants defined as above. Finally, the fiscal externalities amount to

$$
FE = \beta \ell_{pp}\left[\frac{e^{\beta}}{1+\ell_{pp}(e^{\beta}-1)} - \frac{e^{(1-p)\beta}}{1+\ell_{pp}(e^{(1-p)\beta}-1)}\right][\varepsilon_{\beta,1-p} + \varepsilon_{\beta,\ell}\varepsilon_{\ell,1-p}]s_{pp}C_{pp}.
$$

The optimal rate of progressivity satisfies $0 = WE + \frac{1}{C}[ME + BE + FE]$. Using the previous expressions and rearranging terms following the same steps as above leads to

$$
\frac{p}{(1-p)^2} = \frac{\Sigma^2_{\theta} + \frac{s_{pp}C_{pp}}{C}\kappa_1(1+\varepsilon_{\beta,1-p})\mathbb{V}(\log z_{pp}) - \kappa_4}{E_{\ell,1-p} + \frac{s_{pp}C_{pp}}{C}\ell_{pp}\kappa_3\varepsilon_{\ell_{pp},1-p} + \frac{s_{pp}C_{pp}}{C}\kappa_2\varepsilon_{\beta,\ell_{pp}}\varepsilon_{\ell_{pp},1-p}\mathbb{V}(\log z_{pp})}
$$

where we denote the average variance of abilities by

$$
\Sigma^2_{\theta} = \frac{s_{pp}C_{pp}}{C}\sigma^2_{\theta,pp} + \frac{(1-s_{pp})C_{fp}}{C}\sigma^2_{\theta,fp},
$$

the average labor supply elasticity by

$$
E_{\ell,1-p} = \frac{s_{pp}C_{pp}}{C}\left(1+\frac{1}{p}\left(\frac{Z_{pp}}{C_{pp}}-1\right)\right)\varepsilon_{\ell_{pp},1-p} + \frac{(1-s_{pp})C_{fp}}{C}\left(1+\frac{1}{p}\left(\frac{Z_{fp}}{C_{fp}}-1\right)\right)\varepsilon_{\ell_{fp},1-p},
$$

the constants $\kappa_1, \kappa_2, \kappa_3$ by

$$\kappa_1 = \frac{1}{\beta(1-p)} \left(\frac{\bar{c}_{pp} - \underline{c}_{pp}}{\mathbb{E} c_{pp}} + \frac{1 - \frac{C_{pp}}{C}}{\frac{C_{pp}}{C}} \frac{\bar{z}_{pp} - \underline{z}_{pp}}{\mathbb{E} z_{pp}} \right)$$
$$\kappa_2 = \frac{1-p}{\beta p} \left(\frac{\bar{z}_{pp} - \underline{z}_{pp}}{\mathbb{E} z_{pp}} - \frac{\bar{c}_{pp} - \underline{c}_{pp}}{\mathbb{E} c_{pp}} \right)$$
$$\kappa_3 = \frac{1-p}{p} \left(\frac{\bar{z}_{pp} - \underline{z}_{pp}}{\mathbb{E} z_{pp}} - \frac{1}{1-p} \frac{\bar{c}_{pp} - \underline{c}_{pp}}{\mathbb{E} c_{pp}} \right)$$

and the constant κ_4 is given by

$$\kappa_4 = \frac{1}{1-p}(1 - s_{pp}) \left(1 - \frac{C_{fp}}{C}\right) [\mu_{\theta,fp} + \log \ell_{fp}]$$
$$+ \frac{1}{1-p} s_{pp} \left(1 - \frac{C_{pp}}{C}\right) \left[\mu_{\theta,pp} + \log \ell_{pp} + \log \frac{\underline{z}_{pp}}{\mathbb{E} z_{pp}} + \beta \ell_{pp} \frac{\bar{z}_{pp}}{\mathbb{E} z_{pp}}\right].$$

This concludes the proof. ■

B Proofs of Section 2

Concavity of the Utility of Earnings v. Our analysis requires that the utility of earnings $z \mapsto v(z) \equiv u(R(z))$ is concave. It is easy to show that this is equivalent to $p_1(z)p_2(z) > -\gamma(z)$ where $\gamma(z) \equiv -\frac{R(z)u''(R(z))}{u'(R(z))}$ is the agent's coefficient of relative risk aversion, and $p_1(z) \equiv \frac{1-T(z)/z}{1-T'(z)}$, $p_2(z) \equiv \frac{zT''(z)}{1-T'(z)}$ are two measures of the local rate of progressivity of the tax schedule: the parameter $p_1(z)$ is the ratio of the average and marginal retention rates, and $p_2(z)$ is (minus) the elasticity of the retention rate with respect to income. If the tax schedule has a constant rate of progressivity p (CRP), these variables are respectively equal to $\frac{1}{1-p}$ and p. When we characterize the optimal tax schedule within the CRP class, we assume that $u(c) = \log c$ which implies that $\gamma(z) = -1$. It is easy to verify that in this case the above restriction is always satisfied regardless of the value of p. ■

Proof of Lemma 4. Denote the agent's expected utility of effort ℓ by

$$V(\ell) \equiv (1-\ell)u(R(\underline{z}(\theta))) + \ell u(R(\underline{z}(\theta), b(\theta))) - h(\ell).$$

The first-order condition reads $V'(\ell) = 0$, where

$$V'(\ell) = u(R(\underline{z}(\theta), b(\theta))) - u(R(\underline{z}(\theta))) - h'(\ell).$$

We then have

$$V''(\ell) = -h''(\ell) < 0,$$

where the inequality follows from the convexity of the disutility of effort. Thus, the agent's problem is concave and, as long as the effort choice is interior, the first-order condition is necessary and sufficient. ■

Proof of Proposition 2. The participation constraint reads:

$$(1 - \ell)v(\underline{z}, 0) + \ell v(\underline{z}, b) - h(\ell) = U(\theta),$$

and the local incentive constraint reads:

$$v(\underline{z}, b) - v(\underline{z}, 0) = h'(\ell).$$

Solving this linear system of equations for $v(\underline{z}, 0)$ and $v(\underline{z}, b)$ as functions of ℓ and $U(\theta)$ immediately delivers equations (16) and (17). The optimal effort level $\ell(\theta)$ maximizes the firm's profit $\ell\theta - (\underline{z} + \ell b)$ subject to the participation and incentive constraints, taking the reservation value $U(\theta)$ as given. The first-order condition reads:

$$\theta = b + \frac{\partial \underline{z}}{\partial \ell} + \ell \frac{\partial b}{\partial \ell}.$$

Applying the implicit function theorem to equations (16) and (17) leads to

$$v_1(\underline{z}, 0)\frac{\partial \underline{z}}{\partial \ell} = -\ell h''(\ell)$$

and

$$v_1(\underline{z}, b)\frac{\partial \underline{z}}{\partial \ell} + v_2(\underline{z}, b)\frac{\partial b}{\partial \ell} = (1 - \ell)h''(\ell).$$

Solving for $\frac{\partial \underline{z}}{\partial \ell}, \frac{\partial b}{\partial \ell}$ and substituting into the first-order condition yields

$$\theta = b - \left[1 - \ell\frac{v_1(\underline{z}, b)}{v_2(\underline{z}, b)}\right] \frac{1}{v_1(\underline{z}, 0)}\ell h''(\ell) + \frac{1}{v_2(\underline{z}, b)}\ell(1 - \ell)h''(\ell).$$

Rearranging terms and noting that $\frac{v_1(\underline{z},b)}{v_2(\underline{z},b)} = \frac{R_1(\underline{z},b)}{R_2(\underline{z},b)}$ leads to equation (18). Finally, the zero-profit condition $\underline{z} + \ell b = \ell\theta$ pins down the equilibrium reservation utility $U(\theta)$. ■

B.1 Incidence and Optimal Taxation on Total Earnings

Proof of Lemma 5. Suppose that the tax system is over total earnings, so that $R(\underline{z}, b) \equiv R(\bar{z}+b)$ for all $b \geq 0$. Equations (16) and (17), which characterize the equilibrium base pay and bonus for a given a recommended effort level ℓ and reservation utility $U(\theta)$, can then be rewritten as:

$$u(R(\underline{z})) - h(\ell) = U(\theta) - \ell h'(\ell)$$
$$u(R(\bar{z})) - h(\ell) = U(\theta) + (1-\ell)h'(\ell).$$

Consider a reform $\delta\hat{R} : \mathbb{R}_+ \to \mathbb{R}_+$ of the tax schedule, where $\delta \in \mathbb{R}$. Denote by $\hat{\underline{z}}$ and $\hat{\bar{z}}$ the Gateaux derivatives of base pay and high-performance pay following this reform, and by $\hat{\ell}$ and $\hat{U}$ those of labor effort and reservation utility. To a first order as $\delta \to 0$, the values of $\hat{\underline{z}}$ and $\hat{\bar{z}}$ are the solution to the following system:

$$u[R(\underline{z} + \delta\hat{\underline{z}}) + \delta\hat{R}(\underline{z})] - h(\ell + \delta\hat{\ell}) = U(\theta) + \delta\hat{U} - (\ell + \delta\hat{\ell})h'(\ell + \delta\hat{\ell})$$
$$u[R(\bar{z} + \delta\hat{\bar{z}}) + \delta\hat{R}(\bar{z})] - h(\ell + \delta\hat{\ell}) = U(\theta) + \delta\hat{U} + (1 - \ell - \delta\hat{\ell})h'(\ell + \delta\hat{\ell}).$$

Linearizing this system around the initial equilibrium leads to

$$u'(R(\underline{z}))\hat{R}(\underline{z}) + R'(\underline{z})u'(R(\underline{z}))\hat{\underline{z}} - h'(\ell)\hat{\ell} = \hat{U} - [h'(\ell) + \ell h''(\ell)]\hat{\ell}$$
$$u'(R(\bar{z}))\hat{R}(\bar{z}) + R'(\bar{z})u'(R(\bar{z}))\hat{\bar{z}} - h'(\ell)\hat{\ell} = \hat{U} + [-h'(\ell) + (1-\ell)h''(\ell)]\hat{\ell}.$$

Rearranging terms and noting that $R'u' = v'$ leads to equations (19) and (20). ■

Proof of Lemma 6. The perturbed free-entry condition reads:

$$(\underline{z} + \delta\hat{\underline{z}}) + (\ell + \delta\hat{\ell})(b + \delta\hat{b}) = \theta(\ell + \delta\hat{\ell}).$$

Linearizing this system around the initial equilibrium as $\delta \to 0$ leads to $\hat{\underline{z}} + \ell\hat{b} + b\hat{\ell} = \theta\hat{\ell}$

or, since $\hat{b} = \hat{\bar{z}} - \hat{\underline{z}}$,

$$(1-\ell)\hat{\underline{z}} + \ell\hat{\bar{z}} = (\theta - b)\hat{\ell}.$$

Substituting expressions (19) and (20) into this equation and rearranging terms, we obtain

$$\begin{aligned}\left[\frac{1-\ell}{v'(\underline{z})} + \frac{\ell}{v'(\bar{z})}\right]\hat{U} =& \frac{1-\ell}{R'(\underline{z})}\hat{R}(\underline{z}) + \frac{\ell}{R'(\bar{z})}\hat{R}(\bar{z}) \\ &+ \left[\theta - b - \left(\frac{1}{v'(\bar{z})} - \frac{1}{v'(\underline{z})}\right)\ell(1-\ell)h''(\ell)\right]\hat{\ell}.\end{aligned}$$

But the first-order condition for labor effort (18) when taxes are levied on total earnings can be written as

$$\theta = b + \left[\frac{1}{v'(\bar{z})} - \frac{1}{v'(\underline{z})}\right]\ell(1-\ell)h''(\ell). \tag{35}$$

Thus, the Gateaux derivative of expected utility is given by

$$\hat{U} = \frac{(1-\ell)\frac{\hat{R}(\underline{z})}{R'(\underline{z})} + \ell\frac{\hat{R}(\bar{z})}{R'(\bar{z})}}{(1-\ell)\frac{1}{v'(\underline{z})} + \ell\frac{1}{v'(\bar{z})}},$$

which is equal to expression (21). ■

Lemma 8 *Suppose that the initial tax schedule is piecewise linear. The effect of a tax reform $\hat{R}$ on labor effort ℓ is given by:*

$$\frac{\hat{\ell}}{\ell} \;=\; \varepsilon_{\ell,R'(\underline{z})}\frac{\hat{R}'(\underline{z})}{R'(\underline{z})} + \varepsilon_{\ell,R'(\bar{z})}\frac{\hat{R}'(\bar{z})}{R'(\bar{z})} + \varepsilon_{\ell,R(\underline{z})}\frac{\hat{R}(\underline{z})}{R'(\underline{z})\underline{z}} + \varepsilon_{\ell,R(\bar{z})}\frac{\hat{R}(\bar{z})}{R'(\bar{z})\bar{z}} \tag{36}$$

where the elasticities of labor effort with respect to the marginal tax rates at $\underline{z}$ and $\bar{z}$ are respectively given by:

$$\varepsilon_{\ell,R'(\underline{z})} \;=\; -\frac{1}{D}\left(\frac{\ell b}{\underline{z}}\varepsilon_{in}\right) \quad \textit{and} \quad \varepsilon_{\ell,R'(\bar{z})} \;=\; \frac{1}{D}\left(\frac{\ell b}{\underline{z}}\varepsilon_{in} + 1\right)$$

and the elasticities of labor effort with respect to the average tax rates at $\underline{z}$ and $\bar{z}$ are *respectively given by:*

$$\varepsilon_{\ell,R(\underline{z})} \;=\; -\frac{1}{D}\left(1-\varepsilon_{out}+\frac{(1-\ell)\underline{z}}{\mathbb{E}[1/v']}E\right) \quad \textit{and} \quad \varepsilon_{\ell,R(\bar{z})} \;=\; -\frac{1}{D}\left(-\varepsilon_{out}+\frac{(1-\ell)\underline{z}}{\mathbb{E}[1/v']}E\right),$$

where we denote:

$$D \;\equiv\; -\frac{\ell^2}{\underline{z}}\frac{\partial^2\Pi(\theta)}{\partial\ell^2} > 0 \quad \textit{and} \quad E \;\equiv\; \left(\frac{\ell b}{\underline{z}}\varepsilon_{in}+1\right)\frac{-u''(\bar{c})}{(u'(\bar{c}))^2} - \left(\frac{\ell b}{\underline{z}}\varepsilon_{in}\right)\frac{-u''(\underline{c})}{(u'(\underline{c}))^2}.$$

In particular, if the utility function u is logarithmic, then we have $E=1$ and $\varepsilon_{\ell,R(\bar{z})}<0$ if and only if $R'(\underline{z})>\frac{R(\underline{z})}{\underline{z}}$.

Proof of Lemma 8. The first-order condition (35) for labor effort, expressed at the perturbed tax schedule and to a first order as $\delta\to 0$, reads:

$$\begin{aligned}\theta = b+\delta\hat{b}+ \Bigg[&\frac{1}{[R'(\bar{z})+\delta(\hat{R}'(\bar{z})+R''(\bar{z})\hat{\bar{z}})]u'[R(\bar{z})+\delta(\hat{R}(\bar{z})+R'(\bar{z})\hat{\bar{z}})]}-\\ &\frac{1}{[R'(\underline{z})+\delta(\hat{R}'(\underline{z})+R''(\underline{z})\hat{\underline{z}})]u'[R(\underline{z})+\delta(\hat{R}(\underline{z})+R'(\underline{z})\hat{\underline{z}})]}\Bigg]\\ &\times(\ell+\delta\hat{\ell})(1-\ell-\delta\hat{\ell})h''(\ell+\delta\hat{\ell}).\end{aligned}$$

Suppose for simplicity that the tax schedule is piecewise linear, so that $R''(\underline{z})=R''(\bar{z})=0$. A first-order Taylor expansion of this expression around the initial equilibrium leads to

$$\begin{aligned}0 = &\left[\frac{\ell(1-\ell)h''(\ell)}{R'(\underline{z})u'(R(\underline{z}))}\right]\frac{\hat{R}'(\underline{z})}{R'(\underline{z})} - \left[\frac{\ell(1-\ell)h''(\ell)}{R'(\bar{z})u'(R(\bar{z}))}\right]\frac{\hat{R}'(\bar{z})}{R'(\bar{z})}\\ &+\left[\frac{\ell(1-\ell)h''(\ell)}{u'(R(\underline{z}))}\frac{u''(R(\underline{z}))}{u'(R(\underline{z}))}\right]\frac{\hat{R}(\underline{z})}{R'(\underline{z})} - \left[\frac{\ell(1-\ell)h''(\ell)}{u'(R(\bar{z}))}\frac{u''(R(\bar{z}))}{u'(R(\bar{z}))}\right]\frac{\hat{R}(\bar{z})}{R'(\bar{z})}\\ &+\left[\frac{\ell(1-\ell)h''(\ell)}{u'(R(\underline{z}))}\frac{u''(R(\underline{z}))}{u'(R(\underline{z}))}\right]\hat{\underline{z}} - \left[\frac{\ell(1-\ell)h''(\ell)}{u'(R(\bar{z}))}\frac{u''(R(\bar{z}))}{u'(R(\bar{z}))}\right]\hat{\bar{z}}+\hat{b}\\ &+\ell(1-\ell)h''(\ell)\left(\frac{1}{R'(\bar{z})u'(R(\bar{z}))}-\frac{1}{R'(\underline{z})u'(R(\underline{z}))}\right)\left(\frac{1-2\ell}{1-\ell}+\frac{\ell h'''(\ell)}{h''(\ell)}\right)\frac{\hat{\ell}}{\ell}.\end{aligned}$$

Recall that, by the first-order condition for effort (35) and the zero-profit condition (4),

$$\left[\frac{1}{R'(\bar{z})u'(R(\bar{z}))}-\frac{1}{R'(\underline{z})u'(R(\underline{z}))}\right]\ell(1-\ell)h''(\ell)=\theta-b=\frac{\underline{z}}{\ell}.$$

Moreover, we saw that

$$\hat{\underline{z}} = -\frac{\frac{\ell}{v'(\bar{z})}}{\frac{1-\ell}{v'(\underline{z})}+\frac{\ell}{v'(\bar{z})}}\frac{\hat{R}(\underline{z})}{R'(\underline{z})} + \frac{\frac{\ell}{v'(\underline{z})}}{\frac{1-\ell}{v'(\underline{z})}+\frac{\ell}{v'(\bar{z})}}\frac{\hat{R}(\bar{z})}{R'(\bar{z})} - \frac{\ell}{1-\ell}\frac{\ell(1-\ell)h''(\ell)}{v'(\underline{z})}\frac{\hat{\ell}}{\ell}$$

and

$$\hat{\bar{z}} = \frac{\frac{1-\ell}{v'(\bar{z})}}{\frac{1-\ell}{v'(\underline{z})}+\frac{\ell}{v'(\bar{z})}}\frac{\hat{R}(\underline{z})}{R'(\underline{z})} - \frac{\frac{1-\ell}{v'(\underline{z})}}{\frac{1-\ell}{v'(\underline{z})}+\frac{\ell}{v'(\bar{z})}}\frac{\hat{R}(\bar{z})}{R'(\bar{z})} + \frac{\ell(1-\ell)h''(\ell)}{v'(\bar{z})}\frac{\hat{\ell}}{\ell}$$

and hence

$$\hat{b} = \frac{\frac{1}{v'(\bar{z})}}{\frac{1-\ell}{v'(\underline{z})}+\frac{\ell}{v'(\bar{z})}}\frac{\hat{R}(\underline{z})}{R'(\underline{z})} - \frac{\frac{1}{v'(\underline{z})}}{\frac{1-\ell}{v'(\underline{z})}+\frac{\ell}{v'(\bar{z})}}\frac{\hat{R}(\bar{z})}{R'(\bar{z})} + \left[\frac{\underline{z}}{\ell} + \frac{\ell h''(\ell)}{v'(\underline{z})}\right]\frac{\hat{\ell}}{\ell}.$$

We thus obtain

$$\begin{aligned}
D\frac{\hat{\ell}}{\ell} = &- \frac{\ell}{\underline{z}}\left[\frac{\ell(1-\ell)h''(\ell)}{R'(\underline{z})u'(R(\underline{z}))}\right]\frac{\hat{R}'(\underline{z})}{R'(\underline{z})} + \frac{\ell}{\underline{z}}\left[\frac{\ell(1-\ell)h''(\ell)}{R'(\bar{z})u'(R(\bar{z}))}\right]\frac{\hat{R}'(\bar{z})}{R'(\bar{z})} \\
&- \frac{\ell}{\underline{z}}\left[\frac{\frac{\frac{1}{1-\ell}}{R'(\bar{z})u'(R(\bar{z}))} + \left(\frac{u''(R(\underline{z}))}{R'(\underline{z})(u'(R(\underline{z})))^3} - \frac{u''(R(\bar{z}))}{R'(\bar{z})(u'(R(\bar{z})))^3}\right)\ell(1-\ell)h''(\ell)}{\frac{1-\ell}{R'(\underline{z})u'(R(\underline{z}))} + \frac{\ell}{R'(\bar{z})u'(R(\bar{z}))}}\right](1-\ell)\frac{\hat{R}(\underline{z})}{R'(\underline{z})} \\
&+ \frac{\ell}{\underline{z}}\left[\frac{\frac{\frac{1}{\ell}}{R'(\underline{z})u'(R(\underline{z}))} - \left(\frac{u''(R(\underline{z}))}{R'(\underline{z})(u'(R(\underline{z})))^3} - \frac{u''(R(\bar{z}))}{R'(\bar{z})(u'(R(\bar{z})))^3}\right)\ell(1-\ell)h''(\ell)}{\frac{1-\ell}{R'(\underline{z})u'(R(\underline{z}))} + \frac{\ell}{R'(\bar{z})u'(R(\bar{z}))}}\right]\ell\frac{\hat{R}(\bar{z})}{R'(\bar{z})}
\end{aligned}$$

where

$$\begin{aligned}
D = &\frac{1-2\ell}{1-\ell} + \frac{\ell h'''(\ell)}{h''(\ell)} + \ell h''(\ell)\times \\
&+\left(\frac{\ell}{\underline{z}}\frac{1-\ell}{R'(\bar{z})u'(R(\bar{z}))} + \frac{\ell}{\underline{z}}\frac{\ell}{R'(\underline{z})u'(R(\underline{z}))} - \frac{\ell\frac{u''(R(\underline{z}))}{R'(\underline{z})(u'(R(\underline{z})))^3} + (1-\ell)\frac{u''(R(\bar{z}))}{R'(\bar{z})(u'(R(\bar{z})))^3}}{\frac{1}{R'(\bar{z})u'(R(\bar{z}))} - \frac{1}{R'(\underline{z})u'(R(\underline{z}))}}\right).
\end{aligned}$$

Now, recall that the firm's profit is equal to $\Pi(\theta) = \ell\theta - \underline{z} - \ell b$. Thus, we can write

$$\begin{aligned}
\frac{\partial\Pi(\theta)}{\partial\ell} &= \theta - b - \frac{\partial\underline{z}}{\partial\ell} - \ell\frac{\partial b}{\partial\ell} \\
&= \theta - b - \left[\frac{1}{R'(\bar{z})u'(R(\bar{z}))} - \frac{1}{R'(\underline{z})u'(R(\underline{z}))}\right]\ell(1-\ell)h''(\ell).
\end{aligned}$$

The second-order condition to the firm's maximization problem reads:

$$\frac{\partial^2 \Pi(\theta)}{\partial \ell^2} \leq 0. \tag{37}$$

Differentiating the previous expression leads to

$$\begin{aligned}\frac{\partial^2 \Pi(\theta)}{\partial \ell^2} = &- \frac{\partial b}{\partial \ell} - \left[\frac{u''(R(\underline{z}))}{(u'(R(\underline{z})))^2}\frac{\partial \underline{z}}{\partial \ell} - \frac{u''(R(\bar{z}))}{(u'(R(\bar{z})))^2}\frac{\partial \bar{z}}{\partial \ell}\right]\ell(1-\ell)h''(\ell) \\ &- \frac{1}{\ell}\left[\frac{1}{R'(\bar{z})u'(R(\bar{z}))} - \frac{1}{R'(\underline{z})u'(R(\underline{z}))}\right]\left[\frac{1-2\ell}{1-\ell} + \frac{\ell h'''(\ell)}{h''(\ell)}\right]\ell(1-\ell)h''(\ell).\end{aligned}$$

But recall that

$$\frac{\partial \underline{z}}{\partial \ell} = -\frac{\ell h''(\ell)}{R'(\underline{z})u'(R(\underline{z}))} \quad \text{and} \quad \frac{\partial b}{\partial \ell} = \left[\frac{1-\ell}{R'(\bar{z})u'(R(\bar{z}))} + \frac{\ell}{R'(\underline{z})u'(R(\underline{z}))}\right]h''(\ell).$$

Hence, we obtain

$$\begin{aligned}&-\frac{\ell^2}{\underline{z}}\frac{\partial^2 \Pi(\theta)}{\partial \ell^2} = \frac{1-2\ell}{1-\ell} + \frac{\ell h'''(\ell)}{h''(\ell)} + \ell h''(\ell)\times \\ &\left(\frac{\ell}{\underline{z}}\frac{1-\ell}{R'(\bar{z})u'(R(\bar{z}))} + \frac{\ell}{\underline{z}}\frac{\ell}{R'(\underline{z})u'(R(\underline{z}))} - \frac{\ell\frac{u''(R(\underline{z}))}{R'(\underline{z})(u'(R(\underline{z})))^3} + (1-\ell)\frac{u''(R(\bar{z}))}{R'(\bar{z})(u'(R(\bar{z})))^3}}{\frac{1}{R'(\bar{z})u'(R(\bar{z}))} - \frac{1}{R'(\underline{z})u'(R(\underline{z}))}}\right).\end{aligned}$$

where we used again equation (35) with $\theta - b = \underline{z}/\ell$. We can therefore rewrite the Gateaux derivative of labor effort as

$$\begin{aligned}\left(-\ell\frac{\partial^2 \Pi(\theta)}{\partial \ell^2}\right)\frac{\hat{\ell}}{\ell} = &-\left[\frac{\ell(1-\ell)h''(\ell)}{R'(\underline{z})u'(R(\underline{z}))}\right]\frac{\hat{R}'(\underline{z})}{R'(\underline{z})} + \left[\frac{\ell(1-\ell)h''(\ell)}{R'(\bar{z})u'(R(\bar{z}))}\right]\frac{\hat{R}'(\bar{z})}{R'(\bar{z})} \\ &- \left[\frac{\frac{\frac{1}{1-\ell}}{R'(\bar{z})u'(R(\bar{z}))} + \left(\frac{u''(R(\underline{z}))}{R'(\underline{z})(u'(R(\underline{z})))^3} - \frac{u''(R(\bar{z}))}{R'(\bar{z})(u'(R(\bar{z})))^3}\right)\ell(1-\ell)h''(\ell)}{\frac{1-\ell}{R'(\underline{z})u'(R(\underline{z}))} + \frac{\ell}{R'(\bar{z})u'(R(\bar{z}))}}\right](1-\ell)\frac{\hat{R}(\underline{z})}{R'(\underline{z})} \\ &+ \left[\frac{\frac{\frac{1}{\ell}}{R'(\underline{z})u'(R(\underline{z}))} - \left(\frac{u''(R(\underline{z}))}{R'(\underline{z})(u'(R(\underline{z})))^3} - \frac{u''(R(\bar{z}))}{R'(\bar{z})(u'(R(\bar{z})))^3}\right)\ell(1-\ell)h''(\ell)}{\frac{1-\ell}{R'(\underline{z})u'(R(\underline{z}))} + \frac{\ell}{R'(\bar{z})u'(R(\bar{z}))}}\right]\ell\frac{\hat{R}(\bar{z})}{R'(\bar{z})}\end{aligned}$$

where, by condition (37), the term multiplying $\hat{\ell}/\ell$ in the left hand side is positive.

Using the definition of ε_{out}, and noting that

$$\frac{\ell b}{\underline{z}}\varepsilon_{in} = \frac{\ell}{\underline{z}}\frac{\ell(1-\ell)h''(\ell)}{v'(\underline{z})} \quad \text{and} \quad \frac{\ell b}{\underline{z}}\varepsilon_{in} + 1 = \frac{\ell}{\underline{z}}\frac{\ell(1-\ell)h''(\ell)}{v'(\bar{z})},$$

leads to equation (36). Note finally that if the utility function is logarithmic, this expression simplifies to:

$$\left(-\frac{\ell^2}{\underline{z}}\frac{\partial^2\Pi(\theta)}{\partial\ell^2}\right)\frac{\hat{\ell}}{\ell} = -\left[\frac{\frac{R(\underline{z})}{R'(\underline{z})}}{\frac{R(\bar{z})}{R'(\bar{z})} - \frac{R(\underline{z})}{R'(\underline{z})}}\right]\frac{\hat{R}'(\underline{z})}{R'(\underline{z})} + \left[\frac{\frac{R(\bar{z})}{R'(\bar{z})}}{\frac{R(\bar{z})}{R'(\bar{z})} - \frac{R(\underline{z})}{R'(\underline{z})}}\right]\frac{\hat{R}'(\bar{z})}{R'(\bar{z})}$$
$$-\left[\frac{\frac{\ell}{1-\ell}\frac{R(\bar{z})}{\underline{z}R'(\bar{z})} + 1}{(1-\ell)\frac{R(\underline{z})}{R'(\underline{z})} + \ell\frac{R(\bar{z})}{R'(\bar{z})}}\right](1-\ell)\frac{\hat{R}(\underline{z})}{R'(\underline{z})} - \left[\frac{1 - \frac{R(\underline{z})}{\underline{z}R'(\underline{z})}}{(1-\ell)\frac{R(\underline{z})}{R'(\underline{z})} + \ell\frac{R(\bar{z})}{R'(\bar{z})}}\right]\ell\frac{\hat{R}(\bar{z})}{R'(\bar{z})}.$$

where we used again equation (35). This concludes the proof.

Extension to a locally nonlinear baseline tax schedule. Accounting for the terms involving R'' in the above Taylor expansion leads to the following more general expression for the response of labor effort to tax reforms:

$$D'\frac{\hat{\ell}}{\ell} = -\frac{\ell}{\underline{z}}\left[\frac{\ell(1-\ell)h''(\ell)}{R'(\underline{z})u'(R(\underline{z}))}\right]\frac{\hat{R}'(\underline{z})}{R'(\underline{z})} + \frac{\ell}{\underline{z}}\left[\frac{\ell(1-\ell)h''(\ell)}{R'(\bar{z})u'(R(\bar{z}))}\right]\frac{\hat{R}'(\bar{z})}{R'(\bar{z})}$$
$$-\frac{\ell}{\underline{z}}\left[\frac{A\cdot\frac{\frac{1}{1-\ell}}{R'(\bar{z})u'(R(\bar{z}))} + \left(\frac{u''(R(\underline{z}))}{R'(\underline{z})(u'(R(\underline{z})))^3} - \frac{u''(R(\bar{z}))}{R'(\bar{z})(u'(R(\bar{z})))^3}\right)\ell(1-\ell)h''(\ell)}{\frac{1-\ell}{R'(\underline{z})u'(R(\underline{z}))} + \frac{\ell}{R'(\bar{z})u'(R(\bar{z}))}}\right](1-\ell)\frac{\hat{R}(\underline{z})}{R'(\underline{z})}$$
$$+\frac{\ell}{\underline{z}}\left[\frac{A\cdot\frac{\frac{1}{\ell}}{R'(\underline{z})u'(R(\underline{z}))} - \left(\frac{u''(R(\underline{z}))}{R'(\underline{z})(u'(R(\underline{z})))^3} - \frac{u''(R(\bar{z}))}{R'(\bar{z})(u'(R(\bar{z})))^3}\right)\ell(1-\ell)h''(\ell)}{\frac{1-\ell}{R'(\underline{z})u'(R(\underline{z}))} + \frac{\ell}{R'(\bar{z})u'(R(\bar{z}))}}\right]\ell\frac{\hat{R}(\bar{z})}{R'(\bar{z})}$$

where we denote

$$D' = \frac{1-2\ell}{1-\ell} + \frac{\ell h'''(\ell)}{h''(\ell)} - \frac{\frac{\frac{\ell}{1-\ell}R''(\underline{z})}{(R'(\underline{z}))^3(u'(R(\underline{z})))^2} + \frac{R''(\bar{z})}{(R'(\bar{z}))^3(u'(R(\bar{z})))^2}}{\frac{1}{R'(\bar{z})u'(R(\bar{z}))} - \frac{1}{R'(\underline{z})u'(R(\underline{z}))}} + \ell h''(\ell)\times$$
$$+\left(\frac{\ell}{\underline{z}}\frac{1-\ell}{R'(\bar{z})u'(R(\bar{z}))} + \frac{\ell}{\underline{z}}\frac{\ell}{R'(\underline{z})u'(R(\underline{z}))} - \frac{\ell\frac{u''(R(\underline{z}))}{R'(\underline{z})(u'(R(\underline{z})))^3} + (1-\ell)\frac{u''(R(\bar{z}))}{R'(\bar{z})(u'(R(\bar{z})))^3}}{\frac{1}{R'(\bar{z})u'(R(\bar{z}))} - \frac{1}{R'(\underline{z})u'(R(\underline{z}))}}\right),$$

and

$$A = 1 - \left(\frac{\ell R''(\underline{z})}{(R'(\underline{z}))^2 u'(R(\underline{z}))} + \frac{(1-\ell) R''(\bar{z})}{(R'(\bar{z}))^2 u'(R(\bar{z}))} \right).$$

Note that this term appears in the second and third line of the right hand side of the expression for $\hat{\ell}/\ell$, which is otherwise identical to the formula derived above. ■

Proof of Proposition 3. Substituting expression (21) into (19) and (20) implies

$$\hat{\underline{z}} = \left[\frac{\frac{1-\ell}{v'(\underline{z})}}{\frac{1-\ell}{v'(\underline{z})} + \frac{\ell}{v'(\bar{z})}} - 1 \right] \frac{\hat{R}(\underline{z})}{R'(\underline{z})} + \frac{\frac{\ell}{v'(\underline{z})}}{\frac{1-\ell}{v'(\underline{z})} + \frac{\ell}{v'(\bar{z})}} \frac{\hat{R}(\bar{z})}{R'(\bar{z})} - \frac{\ell h''(\ell)}{v'(\underline{z})} \hat{\ell}$$

$$\hat{\bar{z}} = \frac{\frac{1-\ell}{v'(\bar{z})}}{\frac{1-\ell}{v'(\underline{z})} + \frac{\ell}{v'(\bar{z})}} \frac{\hat{R}(\underline{z})}{R'(\underline{z})} + \left[\frac{\frac{\ell}{v'(\bar{z})}}{\frac{1-\ell}{v'(\underline{z})} + \frac{\ell}{v'(\bar{z})}} - 1 \right] \frac{\hat{R}(\bar{z})}{R'(\bar{z})} + \frac{(1-\ell) h''(\ell)}{v'(\bar{z})} \hat{\ell}.$$

This system can be rewritten as follows:

$$\hat{\underline{z}} = - \left[1 - \frac{\frac{1-\ell}{v'(\underline{z})}}{\frac{1-\ell}{v'(\underline{z})} + \frac{\ell}{v'(\bar{z})}} \right] \frac{\hat{R}(\underline{z})}{R'(\underline{z})} + \frac{1}{1-\ell} \frac{\frac{1-\ell}{v'(\underline{z})}}{\frac{1-\ell}{v'(\underline{z})} + \frac{\ell}{v'(\bar{z})}} \ell \frac{\hat{R}(\bar{z})}{R'(\bar{z})} - \frac{1}{1-\ell} \frac{\ell(1-\ell) h''(\ell)}{v'(\underline{z})} \hat{\ell}$$

$$\hat{\bar{z}} = \frac{1}{\ell} \left[\frac{\frac{\ell}{v'(\bar{z})}}{\frac{1-\ell}{v'(\underline{z})} + \frac{\ell}{v'(\bar{z})}} \right] (1-\ell) \frac{\hat{R}(\underline{z})}{R'(\underline{z})} - \left[\frac{\frac{1-\ell}{v'(\underline{z})}}{\frac{1-\ell}{v'(\underline{z})} + \frac{\ell}{v'(\bar{z})}} \right] \frac{\hat{R}(\bar{z})}{R'(\bar{z})} + \frac{1}{\ell} \left[\frac{\ell(1-\ell) h''(\ell)}{v'(\bar{z})} \right] \hat{\ell}.$$

Defining ε_{out} and ε_{in} as in the text, and noting that, by equation (35),

$$\frac{\ell(1-\ell) h''(\ell)}{v'(\bar{z})} = \frac{\ell(1-\ell) h''(\ell)}{v'(\underline{z})} + \theta - b$$

leads to expressions (25) and (26). ■

Proof of Theorem 2. Suppose that there is a top tax bracket $[z^*, \infty)$ with tax rate τ, and that the tax rate below z^* is fixed at t. Let θ_i be the lowest type with high-level pay in the top bracket, and θ_t be the lowest type whose earnings are always in the top bracket. Types $[\theta_i, \theta_t)$ are called "intermediate workers" and types $[\theta_t, \infty)$ are called "top workers". Denote the share of intermediate workers among both the top and intermediate workers by σ^I.

Consider a uniform increase in the marginal tax rate τ in the top bracket $[z^*, \infty)$ by $\delta \hat{\tau} > 0$ with $\delta \to 0$. This perturbation is represented by $\hat{R}(z) = -\hat{\tau}(z - z^*) \mathbb{I}_{\{z \geq z^*\}}$ and $\hat{R}'(z) = -\hat{\tau} \mathbb{I}_{\{z \geq z^*\}}$, so that the tax liability levied after the reform on workers with income $z > z^*$ is $T(z^*) + (\tau + \delta\hat{\tau})(z - z^*)$. Let $\hat{\underline{z}}, \hat{\overline{z}}, \hat{\ell}$ be the changes (Gateaux derivatives) in base pay, high-level pay, and effort in response to this reform.

The Gateaux derivative of a generic variable X of the form $X = \int_{\theta_1}^{\theta_2} x \frac{dF(\theta)}{F(\theta_2)-F(\theta_1)}$ is given by $\hat{X} = \int_{\theta_1}^{\theta_2} \hat{x} \frac{dF(\theta)}{F(\theta_2)-F(\theta_1)}$. We define the elasticity of X with respect to $1-\tau$, keeping the sets of intermediate and top workers fixed, by $e_X = -\frac{1-\tau}{\hat{\tau}} \frac{\hat{X}}{X}$. There are two ways to interpret this variable. First, it is the aggregate elasticity of the variable x keeping the thresholds θ_1, θ_2 fixed, i.e., ignoring the impact of the tax change on composition of the two groups—e.g., the elasticity of mean earnings of top agents, keeping the set of top agents fixed. Alternatively, we can interpret e_X as the average elasticity of x weighted by x/X for a given set of agents—e.g., the elasticity of mean earnings of top agents, weighted by their relative earnings. To see this, note that we can write $e_X = -\int_{\theta_1}^{\theta_2} \frac{x}{X} \cdot \frac{1-\tau}{\hat{\tau}} \frac{\hat{x}}{x} \frac{dF(\theta)}{F(\theta_2)-F(\theta_1)}$.

The impact of the perturbation of the top tax rate $\hat{\tau}$ on the average revenue from top workers is given by

$$dTR^T = \int_{\theta_t}^{\infty} \left\{ \hat{\tau}[\ell\overline{z} + (1-\ell)\underline{z} - z^*] + \tau[\widehat{\ell\overline{z}} + \widehat{(1-\ell)\underline{z}}] \right\} \frac{dF(\theta)}{1-F(\theta_t)}.$$

Denoting average earnings of top workers by $Z^T = \int_{\theta_t}^{\infty} [\ell\overline{z} + (1-\ell)\underline{z}] \frac{dF(\theta)}{1-F(\theta_t)}$, we can rewrite the previous expression as

$$dTR^T = \hat{\tau} \left(Z^T - z^* - \frac{\tau}{1-\tau} Z^T e_{Z^T} \right).$$

Next, the impact of the reform on the tax payments of intermediate workers is given by

$$dTR^I = \int_{\theta_i}^{\theta_t} \left\{ \hat{\tau}(\overline{z} - z^*)\ell + t[(1-\ell)\hat{\underline{z}} + \tau\ell\hat{\overline{z}} + \hat{\ell}(T(\overline{z}) - T(\underline{z}))] \right\} \frac{dF(\theta)}{F(\theta_t) - F(\theta_i)},$$

where $T(\overline{z}) - T(\underline{z}) = \tau(\overline{z} - z^*) - t(z^* - \underline{z})$. Introduce the following notation: the mean effort ℓ of intermediate workers is $L^I = \int_{\theta_i}^{\theta_t} \ell \frac{dF(\theta)}{F(\theta_t)-F(\theta_i)}$, the mean frequency-adjusted high-level pay $\ell\overline{z}$ of intermediate workers is $\overline{Z}^I = \int_{\theta_i}^{\theta_t} \ell\overline{z} \frac{dF(\theta)}{F(\theta_t)-F(\theta_i)}$, and the mean frequency-adjusted base pay $(1-\ell)\underline{z}$ of intermediated workers is $\underline{Z}^I = \int_{\theta_i}^{\theta_t} (1 -$

$\ell)\bar{z}\frac{dF(\theta)}{F(\theta_t)-F(\theta_i)}$. We can rewrite the impact on intermediate workers as

$$
\begin{aligned}
dTR^I &= \int_{\theta_i}^{\theta_t} \left\{ \hat{\tau}(\bar{z}-z^*)\ell + \tau\widehat{\ell\bar{z}} + t\widehat{(1-\ell)\underline{z}} + (t-\tau)z^*\hat{\ell} \right\} \frac{dF(\theta)}{F(\theta_t)-F(\theta_i)} \\
&= \hat{\tau}\left[\bar{Z}^I - L^I z^* - \frac{\tau}{1-\tau}\bar{Z}^I e_{\bar{Z}^I} - \frac{t}{1-\tau}\underline{Z}^I e_{\underline{Z}^I} - \frac{t-\tau}{1-\tau} z^* L^I e_{L^I} \right].
\end{aligned}
$$

The average tax revenue impact on the intermediate and top workers is then

$$
\begin{aligned}
dTR &= (1-\sigma^I)dTR^T + \sigma^I dTR^I \\
&= Z^* - (1-\sigma^I+\sigma^I L^I)z^* - \frac{\tau}{1-\tau}\left[(1-\sigma^I)Z^T e_{Z^T} + \sigma^I \bar{Z}^I e_{\bar{Z}^I}\right] \\
&\quad - \frac{t}{1-\tau}\sigma^I \underline{Z}^I e_{\underline{Z}^I} - \frac{t-\tau}{1-\tau}\sigma^I z^* L^I e_{L^I},
\end{aligned}
$$

where $Z^* = (1-\sigma^I)Z^T + \sigma^I\overline{Z}^I$ denotes the mean frequency-adjusted earnings in the top bracket. It is easy to show that $(1-\sigma^I)Z^T e_{Z^T} + \sigma^I \bar{Z}^I e_{\bar{Z}^I} = Z^* e_{Z^*}$.

Next, denote the average welfare impact of the reform on intermediate and top workers by dW. We have

$$
\begin{aligned}
dW &= -\hat{\tau}\int_{\theta_i}^{\infty} \left\{ \ell\tilde{g}(\bar{z} \mid \theta)(\bar{z}-z^*) + (1-\ell)\tilde{g}(\underline{z} \mid \theta)(\underline{z}-z^*)\mathbb{I}_{\{\underline{z}>z^*\}} \right\} \frac{dF(\theta)}{1-F(\theta_i)} \\
&= -\hat{\tau}[Z^* - (1-\sigma^I+\sigma^I L^I)z^*]\tilde{\mathcal{G}},
\end{aligned}
$$

where $\tilde{\mathcal{G}}$ is the income-weighted average modified marginal social welfare weight in the top bracket. The top tax rate is optimal when

$$
dTR + dW = 0.
$$

Substituting the previous expressions and rearranging, we obtain

$$
\frac{\tau}{1-\tau} = \frac{1-\tilde{\mathcal{G}} - \frac{1}{Z^*-(1-\sigma^I+\sigma^I L^I)z^*}\left[\frac{t}{1-\tau}\sigma^I \underline{Z}^I e_{\underline{Z}^I} + \frac{t-\tau}{1-\tau}\sigma^I z^* L^I e_{L^I}\right]}{\rho\, e_{Z^*}}.
$$

where $\rho = \frac{Z^*}{Z^*-(1-\sigma^I+\sigma^I L^I)z^*}$ is the (observed) Pareto coefficient of the earnings distribution. Now define the earnings share of intermediate workers in the top bracket as $s^I = \frac{\sigma^I(\bar{Z}^I - L^I z^*)}{\sigma^I(\bar{Z}^I-L^I z^*)+(1-\sigma^I)(Z^T-z^*)}$ and the Pareto coefficient for intermediate earners as

$\rho^I = \frac{\bar{Z}^I/L^I}{\bar{Z}^I/L^I - z^*}$. We obtain

$$\frac{\tau}{1-\tau} = \frac{1 - \tilde{\mathcal{G}} - s^I\left(\frac{t}{1-\tau}\frac{\underline{Z}^I}{\bar{Z}^I}\rho^I e_{\underline{Z}_i} + \frac{t-\tau}{1-\tau}(\rho^I - 1)e_{L^I}\right)}{\rho e_{Z^*}}$$

which concludes the proof of equation (27).

Structural expression for $e_{\underline{Z}^I}$. By Proposition 3, the response of base pay to the tax reform $\hat{\tau}$ is given by

$$(1-\ell)\hat{\underline{z}} = -\varepsilon_{out}\ell(\overline{z} - z^*)\frac{\hat{\tau}}{1-\tau} - \overline{z}\varepsilon_{in}\hat{\ell} = -\left[\varepsilon_{out}\ell(\overline{z} - z^*) - \ell\overline{z}\varepsilon_{in}e_\ell\right]\frac{\hat{\tau}}{1-\tau},$$

where $e_\ell = -\frac{1-\tau}{\hat{\tau}}\frac{\hat{\ell}}{\ell}$ is the individual labor effort elasticity with respect to $1-\tau$. Noting that $\widehat{(1-\ell)\underline{z}} = \widehat{(1-\ell)} \cdot \underline{z} - \hat{\ell}\underline{z}$, we can express $e_{\underline{Z}^I}$ as

$$\begin{aligned}
e_{\underline{Z}^I} &= -\frac{1-\tau}{\hat{\tau}}\frac{1}{\underline{Z}^I}\int_{\theta_i}^{\theta_t}\widehat{(1-\ell)\underline{z}}\frac{dF(\theta)}{F(\theta_t) - F(\theta_i)} \\
&= \frac{1}{\underline{Z}^I}\int_{\theta_i}^{\theta_t}\left\{\varepsilon_{out}(\bar{z} - z^*)\ell - \ell\bar{z}\varepsilon_{in}e_\ell + (1-\ell)\underline{z}e_{1-\ell}\right\}\frac{dF(\theta)}{F(\theta_t) - F(\theta_i)},
\end{aligned}$$

where $e_{1-\ell}$ is the individual elasticity of $1-\ell$. Define the following averages over intermediate workers: $\overline{\varepsilon_{out}}^I$ is the average crowd-out parameter, $\overline{\varepsilon_{in}e_\ell}^I$ is the average product of the crowd-in parameter and the labor effort elasticity, and $\overline{e_{1-\ell}}^I$ is the average elasticity of one minus labor effort, all appropriately weighted:

$$\begin{aligned}
\overline{\varepsilon_{out}}^I &= \int_{\theta_i}^{\theta_t}\varepsilon_{out}\frac{\ell\bar{z} - \ell z^*}{\bar{Z}^I - L^I z^*}\frac{dF(\theta)}{F(\theta_t) - F(\theta_i)}, \\
\overline{\varepsilon_{in}e_\ell}^I &= \int_{\theta_i}^{\theta_t}\varepsilon_{in}e_\ell\frac{\ell\bar{z}}{\bar{Z}^I}\frac{dF(\theta)}{F(\theta_t) - F(\theta_i)}, \\
\overline{e_{1-\ell}}^I &= \int_{\theta_i}^{\theta_t}e_{1-\ell}\frac{(1-\ell)\underline{z}}{\underline{Z}^I}\frac{dF(\theta)}{F(\theta_t) - F(\theta_i)}.
\end{aligned}$$

We can then rewrite the elasticity of mean base pay as

$$e_{\underline{Z}_i} = \frac{\bar{Z}^I - L^I z^*}{\underline{Z}^I}\overline{\varepsilon_{out}}^I - \frac{\bar{Z}^I}{\underline{Z}^I}\overline{\varepsilon_{in}e_\ell}^I + \overline{e_{1-\ell}}^I = \kappa_1^{-1}\left(\overline{\varepsilon_{out}}^I - \rho^I\overline{\varepsilon_{in}e_\ell}^I\right) + \overline{e_{1-\ell}}^I.$$

This concludes the proof of equation (28). ■

B.2 Full Optimum Tax Schedule

Theorem 2 provides a formula for the optimal top-bracket tax rate. We now derive and analyze a formula (à la Diamond-Saez) for the full optimal non-linear tax schedule.

Consider an arbitrary tax reform $\hat{T}$ of a given baseline tax schedule T. The effect (Gateaux derivative) of this reform on the social objective is $WE = \frac{1}{\lambda}\int \hat{U}(\theta)dF(\theta)$, where the impact on individual expected utility $\hat{U}(\theta)$ is described by Lemma 6, and λ denotes the marginal value of public funds. The effect of the reform on tax revenue is

$$\begin{aligned} RE = &\int \left[\hat{T}(\overline{z}(\theta))(1-\ell(\theta)) + \hat{T}(\underline{z}(\theta))\ell(\theta)\right] dF(\theta) \\ &+ \int \left[T'(\underline{z}(\theta))\hat{\underline{z}}(\theta)(1-\ell(\theta)) + T'(\overline{z}(\theta))\hat{\overline{z}}(\theta)\ell(\theta)\right] dF(\theta) \\ &+ \int \left[T(\overline{z}(\theta)) - T(\underline{z}(\theta))\right]\hat{\ell}(\theta)dF(\theta) \end{aligned}$$

where the first integral is the mechanical effect of the reform, the second integral captures the responses of base pay and high-level pay, and the third captures the frequency responses.

We now express these effects in terms of the earnings distribution. We assume throughout that $\underline{z}(\theta)$ and $\bar{z}(\theta)$ are strictly increasing in θ. We can thus change variables from θ to $z = \bar{z}(\theta)$, e.g., $\ell(z)$, $\hat{\underline{z}}(z)$, $\hat{\bar{z}}(z)$, and $\Delta T(z) \equiv T(\bar{z}(\theta)) - T(\underline{z}(\theta))$. We use a different convention whenever a variable is naturally indexed by $\underline{z}$, in which case we use a change variables from θ to $\underline{\theta}$; this is the case below for the elasticities of base earnings with respect to taxes. Denote the c.d.f. of total earnings by F_z, and the (scaled)[44] c.d.f.s of base pay and high-level pay by $F_{\underline{z}}$ and $F_{\bar{z}}$, respectively (the corresponding p.d.f.s are denoted by $f_z, f_{\underline{z}}, f_{\bar{z}}$).

We can then write the welfare effect as $WE = -\int \tilde{g}(z)\hat{T}(z)dF_z(z)$, where $\tilde{g}(z)$ is the average modified marginal social welfare weight conditional on earnings z, defined

[44] The scaling is due to the fact that they would otherwise not converge to 1 as $z \to \infty$, but rather to the average values of $1-\ell$ and ℓ in the economy, respectively.

by:

$$\tilde{g}(z) = \frac{f_{\underline{z}}(z)}{f_z(z)}\tilde{g}(z \mid \theta_{\underline{z}}) + \frac{f_{\bar{z}}(z)}{f_z(z)}\tilde{g}(z \mid \theta_{\bar{z}}),$$

where $\theta_{\underline{z}}$ and $\theta_{\bar{z}}$ follow $\underline{z}(\theta_{\underline{z}}) = \bar{z}(\theta_{\bar{z}}) = z$, and where $\tilde{g}(z \mid \theta)$ are given in Corollary 1.[45]

Next, the revenue effect of the tax reform can be written as

$$RE = \int_0^\infty \hat{T}(z)dF_z(z) + \int_0^\infty T'(z)\hat{\underline{z}}(z)dF_{\underline{z}}(z) + \int_0^\infty T'(z)\hat{\bar{z}}(z) + \Delta T(z)\frac{\hat{\ell}(z)}{\ell(z)}dF_{\bar{z}}(z).$$

Define the elasticities of the variables $x \in \{\underline{z}, \bar{z}, \ell\}$ with respect to the relevant marginal and average tax rates as

$$\frac{\hat{x}(\theta)}{x(\theta)} = \varepsilon_{x,R'(\underline{z})}(\theta)\frac{\hat{R}'(\underline{z}(\theta))}{R'(\underline{z}(\theta))} + \varepsilon_{x,R'(\bar{z})}(\theta)\frac{\hat{R}'(\bar{z}(\theta))}{R'(\bar{z}(\theta))} + \varepsilon_{x,R(\underline{z})}(\theta)\frac{\hat{R}(\underline{z}(\theta))}{R(\underline{z}(\theta))} + \varepsilon_{x,R(\bar{z})}(\theta)\frac{\hat{R}(\bar{z}(\theta))}{R(\bar{z}(\theta))}.$$

These elasticities can be obtained from Lemma 8 (see extension of the result in the proof to allow for a fully non-linear initial tax schedule) and Proposition 3. As explained above we index these elasticities by earnings, e.g., $\varepsilon_{\underline{z},R'(\bar{z})}(z)$ stands for $\varepsilon_{\underline{z},R'(\bar{z})}(\theta)$ for θ s.t. $\underline{z}(\theta) = z$. Furthermore, define the following mappings linking the two levels of pay: $\underline{x} \equiv \underline{z} \circ \bar{z}^{-1}$ and $\overline{x} \equiv \bar{z} \circ \underline{z}^{-1}$, meaning that $\underline{x}(z) = \underline{z}(\theta)$ for θ s.t. $\bar{z}(\theta) = z$, and $\overline{x}(z) = \bar{z}(\theta)$ for θ s.t. $\underline{z}(\theta) = z$.

To derive the optimal tax formula, consider a reduction of the marginal tax rate at earnings level z^* (Saez (2001)). Following Sachs, Tsyvinski, and Werquin (2020), we can express this reform as $\hat{T}'(z) = -\delta(z^* - z)$ and $\hat{T}(z) = -\mathbb{I}_{\{z > z^*\}}$, where δ is the Dirac delta function. This reform generates the following effects:

1. Welfare effect: $WE = \int_{z^*}^\infty g(z)dF_z(z)$, with marginal social welfare weights $g(z)$ defined as in Theorem 2.

2. Mechanical effect: $ME = -(1 - F_z(z^*))$.

3. Behavioral effects due to the reduction of $T'(z^*)$:

[45] The MVPF is defined as the social welfare gain from reducing every agent's tax payment by one unit: $\lambda = \frac{1}{1+\mathcal{I}} \int_0^\infty \tilde{g}(z)u'(R(z))f_z(z)dz$ where $\mathcal{I} \equiv \int_0^\infty T'(z)\frac{\partial z}{\partial R}f_z(z)dz$ is the aggregate income effect that this tax cut induces.

(a) Effect on earnings at z^*

$$BE_{z^*,R'} = \frac{T'(z^*)}{1-T'(z^*)} z^* \left[\varepsilon_{\underline{z},R'(\underline{z})}(z^*) \cdot f_{\underline{z}}(z^*) + \varepsilon_{\bar{z},R'(\bar{z})}(z^*) \cdot f_{\bar{z}}(z^*)\right]$$

Note that we can express the above as $\frac{T'(z^*)}{1-T'(z^*)} \cdot z^* \cdot \bar{\varepsilon}_{z,R'(z)}(z^*) \cdot f_z(z^*)$, where $\bar{\varepsilon}_{z,R'(z)}(z^*)$ is the average elasticity of earnings at z^* with respect to the retention rate at z^*.

(b) Effects on base pay at $\underline{x}(z^*)$ and high-level pay at $\overline{x}(z^*)$

$$\begin{aligned} BE_{z\neq z^*,R'} &= \frac{T'(\overline{x}(z^*))}{1-T'(z^*)} \overline{x}(z^*) \cdot \varepsilon_{\bar{z},R'(\underline{z})}(\overline{x}(z^*)) \cdot f_{\bar{z}}(\overline{x}(z^*)) \\ &+ \frac{T'(\underline{x}(z^*))}{1-T'(z^*)} \underline{x}(z^*) \cdot \varepsilon_{\underline{z},R'(\bar{z})}(\underline{x}(z^*)) \cdot f_{\underline{z}}(\underline{x}(z^*)) \end{aligned}$$

(c) Effects on frequencies $\ell(z^*)$ and $\ell(\overline{x}(z^*))$

$$BE_{\ell,R'} = \frac{\Delta T(z^*)}{1-T'(z^*)} \varepsilon_{\ell,R'(\bar{z})}(z^*) \cdot f_{\bar{z}}(z^*) + \frac{\Delta T(\overline{x}(z^*))}{1-T'(z^*)} \varepsilon_{\ell,R'(\underline{z})}(\overline{x}(z^*)) \cdot f_{\bar{z}}(\overline{x}(z^*))$$

4. Behavioral effects due the reduction of $T(z)$ for $z > z^*$:

(a) Effects on base pay

$$BE_{\underline{z},R} = \int_{\underline{x}(z^*)}^{\infty} \frac{T'(z)}{R(\overline{x}(z))} z \cdot \varepsilon_{\underline{z},R(\bar{z})}(z) dF_{\underline{z}}(z) + \int_{z^*}^{\infty} \frac{T'(z)}{R(z)} z \cdot \varepsilon_{\underline{z},R(\underline{z})}(z) dF_{\underline{z}}(z)$$

(b) Effects on high-level pay

$$BE_{\bar{z},R} = \int_{z^*}^{\infty} \frac{T'(z)}{R(z)} z \cdot \varepsilon_{\bar{z},R(\bar{z})}(z) dF_{\bar{z}}(z) + \int_{\overline{x}(z^*)}^{\infty} \frac{T'(z)}{R(\underline{x}(z^*))} z \cdot \varepsilon_{\bar{z},R(\underline{z})}(z) dF_{\bar{z}}(z)$$

(c) Effects on frequency

$$BE_{\ell,R} = \int_{z^*}^{\infty} \frac{\Delta T(z)}{R(z)} \varepsilon_{\ell,R(\bar{z})}(z) dF_{\bar{z}}(z) + \int_{\overline{x}(z^*)}^{\infty} \frac{\Delta T(z)}{R(\underline{x}(z^*))} \varepsilon_{\ell,R(\underline{z})}(z) dF_{\bar{z}}(z)$$

Summing all the effects, equating the sum to zero and rearranging yields the Diamond-

Saez formula with performance pay:

$$\frac{T'(z^*)}{1-T'(z^*)} = \frac{1-\mathcal{G}-\frac{1}{1-F_z(z^*)}\cdot[BE_{z\neq z^*,R'}+BE_{\ell,R'}+BE_{\underline{z},R}+BE_{\bar{z},R}+BE_{\ell,R}]}{\rho(z^*)\cdot\bar{\varepsilon}_{z,R'(z)}(z^*)}$$

where $\rho(z^*) = \frac{z^* f_z(z^*)}{1-F_z(z^*)}$ is the local Pareto parameter of the earnings distribution and $\mathcal{G} = \frac{\int_{z^*}^{\infty} g(z)dF_z(z)}{1-F_z(z^*)}$ is the average marginal social welfare weight of workers with earnings above z^*.

The presence of performance pay modifies the standard Diamond-Saez formula by adjusting the marginal social welfare weights and adding several new fiscal externalities. To build intuition about how these modifications affect optimal tax rates, suppose that the ability distribution is bounded with strictly positive infimum: $\theta \in [\underline{\theta}, \overline{\theta}]$, with $\underline{\theta} > 0$ and $\overline{\theta} < \infty$. As a result, the support of base pay and high-level pay do not fully overlap: sufficiently low earnings $z < \overline{z}_{lb}$ can only be reached with base pay, while sufficiently high earnings $z > \underline{z}_{ub}$ can only be reached with high-level pay. We describe how tax rates should be chosen at such extremities.[46]

For simplicity, we assume that the baseline tax schedule $T(\cdot)$ is strictly increasing and progressive: $T'(z) > \frac{T(z)}{z}$ for all $z \geq 0$, and that the marginal tax rate is (approximately) constant at the two extremities of the earnings distribution. Thus, the signs of the effort elasticities with respect to $R(\underline{z})$ and $R(\bar{z})$ are as described in Lemma 7. Consider the range of earnings which can be reached only with high-level pay (respectively, base pay). It is straightforward to show that:

1. The marginal social welfare weights are adjusted upwards (downwards) relative to the standard formula, leading to lower (higher) tax rates. Intuitively, for the high extremity, the crowd-out implies that the consumption gains from the lower taxation of high-level pay are partially transferred, via the endogenous earnings adjustment, to the base pay, where the marginal utility of consumption is higher.

2. The fiscal externalities due to frequency responses $BE_{\ell,R'}$ and $BE_{\ell,R}$ are strictly positive (negative), leading to lower (higher) tax rates.

[46]We can characterize the tax rate in the lower extremity using a slightly different (mirror image) tax reform than the Saezian reform described above, namely, a decrease in the marginal tax rate at z^* such that the level of taxes increases uniformly below z^* but is unchanged above.

3. The fiscal externality $BE_{z \neq z^*, R'}$ is negative in both cases, leading to higher tax rates.

4. The signs of the fiscal externalities $BE_{\underline{z},R}$ and $BE_{\bar{z},R}$ depend on the relative strengths of the crowding-out and the crowding-in. If they approximately offset each other, as in Section 1, then these fiscal externalities are approximately equal to zero.

Thus, as long as the crowd-out and crowd-in approximately offset each other and the impact of $BE_{z \neq z^*, R'}$ is approximately uniform across the earnings distribution, performance pay tends to reduce to optimal tax progressivity: It leads to relatively lower tax rates at high earnings levels (that can only be reached with high-level pay) and relatively higher tax rates at low earnings levels (that can only be reached with base pay). This is intuitively consistent with our results on the optimal separation of tax rates on bonuses and base pay. In Proposition 4 we demonstrated that there are efficiency gains from taxing bonuses at lower rate than base pay, since labor effort is much more responsive to the tax on bonuses. The tax rate on earnings in the upper extremity, where no one earns a base pay, is effectively a tax rate on bonuses. Thus, reducing it brings strong efficiency gains due to higher effort. A mirror image of this argument applies to the tax rate in the lower extremity of the earnings distribution, which can be understood as a tax rate on base pay and should be, consequently, set at a higher level.

In the previous argument, we restricted attention to the regions of the earnings distribution where all earnings come either exclusively from base pay or exclusively from high-level pay. We conjecture that this reasoning holds more generally when both forms of pay coincide, in which case the optimal tax rates should be set higher (lower) in the regions of the earnings distribution where base pay (high-level pay) is relatively more prevalent. Given that base pay is necessarily more prevalent than high-level pay at low earnings levels, this implies that performance pay makes the optimal tax schedule more regressive than in the standard setting, consistent with our findings regarding the optimal overall rate of progressivity (Theorem 1).

B.3 Incidence and Optimal Taxation with Separate Tax Schedules

Lemma 9 *Suppose that the tax system is separable between base pay and bonuses, so that $R(\underline{z}, b) \equiv R(\underline{z}) + P(b)$ for all $b \geq 0$. A tax reform $(\hat{R}, \hat{P})$ leads to the following incidence:*

$$\hat{\underline{z}} = -(1-\varepsilon'_{out})\frac{\hat{R}(\underline{z})}{R'(\underline{z})} + \varepsilon'_{out}\ell\frac{\hat{P}(b)}{P'(b)} - b\varepsilon'_{in}\hat{\ell} \tag{38}$$

$$\ell\hat{b} = (1-\varepsilon'_{out})\frac{\hat{R}(\underline{z})}{R'(\underline{z})} - \varepsilon'_{out}\ell\frac{\hat{P}(b)}{P'(b)} + \left(b\varepsilon'_{in} + \frac{\underline{z}}{\ell}\right)\hat{\ell} \tag{39}$$

and

$$\frac{\hat{U}}{R'(\underline{z})u'(R(\underline{z}))} = \varepsilon'_{out}\left[\frac{\hat{R}(\underline{z})}{R'(\underline{z})} + \ell\frac{\hat{P}(b)}{P'(b)}\right], \tag{40}$$

where the crowd-out and moral-hazard elasticities are defined by

$$\varepsilon'_{out} = \frac{\frac{1}{R'(\underline{z})u'(R(\underline{z}))}}{\frac{1-\frac{R'(\underline{z})}{P'(b)}\ell}{R'(\underline{z})u'(R(\underline{z}))} + \frac{\ell}{P'(b)u'(R(\underline{z})+P(b))}} \quad \text{and} \quad \varepsilon'_{in} = \frac{\ell h''(\ell)}{bR'(\underline{z})u'(R(\underline{z}))}.$$

Suppose for simplicity that the utility of consumption is logarithmic. Then the impact of the tax reform on labor effort is given by

$$\begin{aligned}
&\left(-\frac{1}{h''(\ell)}\frac{\partial^2\Pi(\theta)}{\partial\ell^2}\right)\frac{\hat{\ell}}{\ell} \qquad (41)\\
&= \left[\left(1-\frac{R'(\underline{z})}{P'(b)}\right)\frac{\frac{R(\underline{z})}{R'(\underline{z})}}{\frac{R(\underline{z})}{R'(\underline{z})}+\ell\frac{P(b)}{P'(b)}} - (1-\ell)\frac{\frac{P(b)}{P'(b)}}{\frac{R(\underline{z})}{R'(\underline{z})}+\ell\frac{P(b)}{P'(b)}} - \frac{1}{\ell h''(\ell)}\frac{\frac{P(b)}{P'(b)}}{\frac{R(\underline{z})}{R'(\underline{z})}+\ell\frac{P(b)}{P'(b)}}\right]\frac{\hat{R}(\underline{z})}{R'(\underline{z})}\\
&+ \left[\ell\left(1-\frac{R'(\underline{z})}{P'(b)}\right)\frac{\frac{R(\underline{z})}{R'(\underline{z})}}{\frac{R(\underline{z})}{R'(\underline{z})}+\ell\frac{P(b)}{P'(b)}} - \ell(1-\ell)\frac{\frac{P(b)}{P'(b)}}{\frac{R(\underline{z})}{R'(\underline{z})}+\ell\frac{P(b)}{P'(b)}} + \frac{1}{\ell h''(\ell)}\frac{\frac{R(\underline{z})}{R'(\underline{z})}}{\frac{R(\underline{z})}{R'(\underline{z})}+\ell\frac{P(b)}{P'(b)}}\right]\frac{\hat{P}(b)}{P'(b)}\\
&- \frac{R(\underline{z})}{R'(\underline{z})}\frac{\hat{R}'(\underline{z})}{R'(\underline{z})} + \left[\frac{R(\underline{z})}{P'(b)} + (1-\ell)\frac{P(b)}{P'(b)}\right]\frac{\hat{P}'(b)}{P'(b)}.
\end{aligned}$$

Proof of Lemma 9. Equations (16) and (17) can be rewritten as follows:

$$\begin{aligned} u(R(\underline{z})) - h(\ell) &= U(\theta) - \ell h'(\ell) \\ u(R(\underline{z}) + P(b)) - h(\ell) &= U(\theta) + (1-\ell)h'(\ell). \end{aligned}$$

After a perturbation $(\delta\hat{R}, \delta\hat{P})$, to a first order as $\delta \to 0$, this system becomes

$$u[R(\underline{z} + \delta\hat{\underline{z}}) + \delta\hat{R}(\underline{z})] - h(\ell + \delta\hat{\ell}) = U(\theta) + \delta\hat{U} - (\ell + \delta\hat{\ell})h'(\ell + \delta\hat{\ell})$$

and

$$\begin{aligned} &u[R(\underline{z} + \delta\hat{\underline{z}}) + P(b + \delta\hat{b}) + \delta(\hat{R}(\underline{z}) + \hat{P}(b))] - h(\ell + \delta\hat{\ell}) \\ =&U(\theta) + \delta\hat{U} + (1 - \ell - \delta\hat{\ell})h'(\ell + \delta\hat{\ell}). \end{aligned}$$

Linearizing this system around the initial equilibrium leads to

$$u'(R(\underline{z}))\hat{R}(\underline{z}) + R'(\underline{z})u'(R(\underline{z}))\hat{\underline{z}} - h'(\ell)\hat{\ell} = \hat{U} - [h'(\ell) + \ell h''(\ell)]\hat{\ell}$$

and

$$\begin{aligned} &u'(R(\underline{z}) + P(b))[\hat{R}(\underline{z}) + \hat{P}(b) + R'(\underline{z})\hat{\underline{z}} + P'(b)\hat{b}] - h'(\ell)\hat{\ell} \\ =&\hat{U} + [-h'(\ell) + (1-\ell)h''(\ell)]\hat{\ell}. \end{aligned}$$

Rearranging terms leads to

$$\hat{\underline{z}} = -\frac{\hat{R}(\underline{z})}{R'(\underline{z})} + \frac{\hat{U}}{R'(\underline{z})u'(R(\underline{z}))} - \frac{\ell h''(\ell)}{R'(\underline{z})u'(R(\underline{z}))}\hat{\ell}$$

and

$$\begin{aligned} \hat{b} =& -\frac{\hat{R}(\underline{z})}{P'(b)} - \frac{\hat{P}(b)}{P'(b)} - \frac{R'(\underline{z})}{P'(b)}\hat{\underline{z}} + \frac{\hat{U}}{P'(b)u'(R(\underline{z}) + P(b))} + \frac{(1-\ell)h''(\ell)}{P'(b)u'(R(\underline{z}) + P(b))}\hat{\ell} \\ =& -\frac{\hat{P}(b)}{P'(b)} + \left[\frac{1}{P'(b)u'(R(\underline{z}) + P(b))} - \frac{1}{P'(b)u'(R(\underline{z}))}\right]\hat{U} \\ &+ \left[\frac{1-\ell}{P'(b)u'(R(\underline{z}) + P(b))} + \frac{\ell}{P'(b)u'(R(\underline{z}))}\right]h''(\ell)\hat{\ell}. \end{aligned}$$

The free-entry condition implies

$$\hat{\underline{z}} + \ell\hat{b} + b\hat{\ell} = \theta\hat{\ell}.$$

Substituting the expressions for $\hat{\underline{z}}$ and $\hat{b}$ into this condition leads to

$$\left[\frac{1-\ell\frac{R'(\underline{z})}{P'(b)}}{R'(\underline{z})u'(R(\underline{z}))} + \frac{\ell}{P'(b)u'(R(\underline{z})+P(b))}\right]\hat{U} = \frac{\hat{R}(\underline{z})}{R'(\underline{z})} + \ell\frac{\hat{P}(b)}{P'(b)}$$
$$+\left[\theta - b - \left(\frac{1-\ell}{P'(b)u'(R(\underline{z})+P(b))} - \frac{1-\ell\frac{R'(\underline{z})}{P'(b)}}{R'(\underline{z})u'(R(\underline{z}))}\right)\ell h''(\ell)\right]\hat{\ell}.$$

But the first-order condition for labor effort (18) can be rewritten as

$$\theta = b + \left[\frac{1}{P'(b)u'(R(\underline{z})+P(b))} - \frac{\frac{1-\ell\frac{R'(\underline{z})}{P'(b)}}{1-\ell}}{R'(\underline{z})u'(R(\underline{z}))}\right]\ell(1-\ell)h''(\ell). \tag{42}$$

Thus, the Gateaux derivative of expected utility is given by

$$\hat{U} = \frac{1}{\frac{1-\ell\frac{R'(\underline{z})}{P'(b)}}{R'(\underline{z})u'(R(\underline{z}))} + \frac{\ell}{P'(b)u'(R(\underline{z})+P(b))}}\left[\frac{\hat{R}(\underline{z})}{R'(\underline{z})} + \ell\frac{\hat{P}(b)}{P'(b)}\right].$$

Substitute this expression into the above equations to get:

$$\hat{\underline{z}} = \left[\frac{\frac{1}{R'(\underline{z})u'(R(\underline{z}))}}{\frac{1-\ell\frac{R'(\underline{z})}{P'(b)}}{R'(\underline{z})u'(R(\underline{z}))} + \frac{\ell}{P'(b)u'(R(\underline{z})+P(b))}} - 1\right]\frac{\hat{R}(\underline{z})}{R'(\underline{z})}$$
$$+\left[\frac{\frac{1}{R'(\underline{z})u'(R(\underline{z}))}}{\frac{1-\ell\frac{R'(\underline{z})}{P'(b)}}{R'(\underline{z})u'(R(\underline{z}))} + \frac{\ell}{P'(b)u'(R(\underline{z})+P(b))}}\right]\ell\frac{\hat{P}(b)}{P'(b)} - \frac{\ell h''(\ell)}{R'(\underline{z})u'(R(\underline{z}))}\hat{\ell}$$

and

$$\ell\hat{b} = -\left[\frac{\frac{1}{R'(\underline{z})u'(R(\underline{z}))}}{\frac{1-\ell\frac{R'(\underline{z})}{P'(b)}}{R'(\underline{z})u'(R(\underline{z}))} + \frac{\ell}{P'(b)u'(R(\underline{z})+P(b))}} - 1\right]\frac{\hat{R}(\underline{z})}{R'(\underline{z})}$$

$$= -\left[\frac{\frac{1}{R'(\underline{z})u'(R(\underline{z}))}}{\frac{1-\ell\frac{R'(\underline{z})}{P'(b)}}{R'(\underline{z})u'(R(\underline{z}))} + \frac{\ell}{P'(b)u'(R(\underline{z})+P(b))}}\right]\ell\frac{\hat{P}(b)}{P'(b)}$$

$$+\left[\frac{1-\ell}{P'(b)u'(R(\underline{z})+P(b))} + \frac{\ell\frac{R'(\underline{z})}{P'(b)}}{R'(\underline{z})u'(R(\underline{z}))}\right]\ell h''(\ell)\hat{\ell}.$$

Using equation (42) to substitute for $\frac{\ell(1-\ell)h''(\ell)}{P'(b)u'(R(\underline{z})+P(b))}$, the above expressions can be rewritten as (38), (39), and (40).

Suppose that the utility function is logarithmic. The crowd-out elasticity is then equal to

$$\varepsilon'_{out} = \frac{\frac{R(\underline{z})}{R'(\underline{z})}}{\frac{R(\underline{z})}{R'(\underline{z})} + \ell\frac{P(b)}{P'(b)}} = \frac{p_{\underline{z}}}{p_{\underline{z}} + \frac{\ell b}{\underline{z}}p_b},$$

where $\frac{\ell b}{\underline{z}}$ is the ratio of expected bonus to base pay, and $p_{\underline{z}}, p_b$ are tax progressivity parameters defined as the ratios of average to marginal tax rates of the base pay and bonus tax schedules:

$$p_{\underline{z}} \equiv \frac{R(\underline{z})}{\underline{z}R'(\underline{z})}, \quad p_b \equiv \frac{P(b)}{bP'(b)}.$$

For more general utility functions, we can approximate the difference in inverse marginal utilities as a function of the difference in consumption levels and the risk aversion (resp., prudence) via first- (resp., second-) order Taylor expansions around $b = 0$. We get

$$\varepsilon'_{out} \approx \frac{\frac{1}{R'(\underline{z})u'(R(\underline{z}))}}{\frac{1}{R'(\underline{z})u'(R(\underline{z}))} + \frac{\ell P(b)}{P'(b)}\left[-\frac{u''(R(\underline{z}))}{(u'(R(\underline{z})))^2}\right]} = \frac{\frac{R(\underline{z})}{R'(\underline{z})}}{\frac{R(\underline{z})}{R'(\underline{z})} + \frac{\ell P(b)}{P'(b)}\left[-\frac{R(\underline{z})u''(R(\underline{z}))}{u'(R(\underline{z}))}\right]} = \frac{p_{\underline{z}}}{p_{\underline{z}} + \frac{\ell b}{\underline{z}}p_b\sigma},$$

where $\sigma \equiv -\frac{R(\underline{z})u''(R(\underline{z}))}{u'(R(\underline{z}))}$ is a risk aversion coefficient.

Next, suppose that $\underline{z}, b$ are located in locally linear portions of the tax schedule, so that $R''(\underline{z}) = P''(b) = 0$. After the perturbation, the first-order condition for effort reads

$$
\begin{aligned}
\theta = {} & b+\delta\hat{b}+(\ell+\delta\hat{\ell})h''(\ell+\delta\hat{\ell})\times \\
& \left[\frac{1-\ell-\delta\hat{\ell}}{\{P'(b)+\delta\hat{P}'(b)\}\{u'(R(\underline{z})+P(b))+\delta[\hat{R}(\underline{z})+\hat{P}(b)+R'(\underline{z})\hat{\underline{z}}+P'(b)\hat{b}]u''(R(\underline{z})+P(b))\}}\right. \\
& \left.-\frac{1-(\ell+\delta\hat{\ell})\frac{R'(\underline{z})+\delta\hat{R}'(\underline{z})}{P'(b)+\delta\hat{P}'(b)}}{\{R'(\underline{z})+\delta\hat{R}'(\underline{z})\}\{u'(R(\underline{z}))+\delta[\hat{R}(\underline{z})+R'(\underline{z})\hat{\underline{z}}]u''(R(\underline{z}))\}}\right].
\end{aligned}
$$

To a first-order as $\delta\to 0$, this implies

$$
\begin{aligned}
\theta = {} & b+\delta\hat{b}+\ell h''(\ell)\left[1+\delta\left(1+\frac{\ell h'''(\ell)}{h''(\ell)}\right)\frac{\hat{\ell}}{\ell}\right]\times \\
& \left[\frac{(1-\ell)\left[1-\delta\frac{\ell}{1-\ell}\frac{\hat{\ell}}{\ell}\right]}{P'(b)u'(R(\underline{z})+P(b))\left[1+\delta\left(\frac{\hat{P}'(b)}{P'(b)}+[\hat{R}(\underline{z})+\hat{P}(b)+R'(\underline{z})\hat{\underline{z}}+P'(b)\hat{b}]\frac{u''(R(\underline{z})+P(b))}{u'(R(\underline{z})+P(b))}\right)\right]}\right. \\
& \left.-\frac{\left(1-\ell\frac{R'(\underline{z})}{P'(b)}\right)\left[1-\delta\frac{\ell\frac{R'(\underline{z})}{P'(b)}}{1-\ell\frac{R'(\underline{z})}{P'(b)}}\left(\frac{\hat{\ell}}{\ell}+\frac{\hat{R}'(\underline{z})}{R'(\underline{z})}-\frac{\hat{P}'(b)}{P'(b)}\right)\right]}{R'(\underline{z})u'(R(\underline{z}))\left[1+\delta\left(\frac{\hat{R}'(\underline{z})}{R'(\underline{z})}+[\hat{R}(\underline{z})+R'(\underline{z})\hat{\underline{z}}]\frac{u''(R(\underline{z}))}{u'(R(\underline{z}))}\right)\right]}\right]
\end{aligned}
$$

i.e.,

$$
\begin{aligned}
\theta = {} & b+\ell h''(\ell)\left(\frac{1-\ell}{P'(b)u'(R(\underline{z})+P(b))}-\frac{1-\ell\frac{R'(\underline{z})}{P'(b)}}{R'(\underline{z})u'(R(\underline{z}))}\right) \\
& +\delta\ell h''(\ell)\left[\left(\frac{1-2\ell}{P'(b)u'(R(\underline{z})+P(b))}-\frac{1-2\ell\frac{R'(\underline{z})}{P'(b)}}{R'(\underline{z})u'(R(\underline{z}))}\right)+\right. \\
& \qquad\left.\left(\frac{1-\ell}{P'(b)u'(R(\underline{z})+P(b))}-\frac{1-\ell\frac{R'(\underline{z})}{P'(b)}}{R'(\underline{z})u'(R(\underline{z}))}\right)\frac{\ell h'''(\ell)}{h''(\ell)}\right]\frac{\hat{\ell}}{\ell} \\
& +\delta\ell h''(\ell)\left(\frac{1-\ell\frac{R'(\underline{z})}{P'(b)}}{R'(\underline{z})}\frac{u''(R(\underline{z}))}{(u'(R(\underline{z})))^2}-\frac{1-\ell}{P'(b)}\frac{u''(R(\underline{z})+P(b))}{(u'(R(\underline{z})+P(b)))^2}\right)\left[\hat{R}(\underline{z})+R'(\underline{z})\hat{\underline{z}}\right] \\
& -\delta\ell h''(\ell)\left(\frac{1-\ell}{P'(b)}\frac{u''(R(\underline{z})+P(b))}{(u'(R(\underline{z})+P(b)))^2}\right)\left[\hat{P}(b)+P'(b)\hat{b}\right]+\delta\hat{b} \\
& +\delta\ell h''(\ell)\left[\frac{1}{R'(\underline{z})u'(R(\underline{z}))}\frac{\hat{R}'(\underline{z})}{R'(\underline{z})}-\left(\frac{\ell}{P'(b)u'(R(\underline{z}))}+\frac{1-\ell}{P'(b)u'(R(\underline{z})+P(b))}\right)\frac{\hat{P}'(b)}{P'(b)}\right].
\end{aligned}
$$

Use the first-order condition for effort and assume that the utility is log to get

$$\left[\left(1-\frac{R'(\underline{z})}{P'(b)}\right)\frac{R(\underline{z})}{R'(\underline{z})}+(2\ell-1)\frac{P(b)}{P'(b)}-\left(\left(\frac{R'(\underline{z})}{P'(b)}-1\right)\frac{R(\underline{z})}{R'(\underline{z})}+(1-\ell)\frac{P(b)}{P'(b)}\right)\frac{\ell h'''(\ell)}{h''(\ell)}\right]\frac{\hat{\ell}}{\ell}$$
$$=\left(\frac{R'(\underline{z})}{P'(b)}-1\right)\left[\frac{\hat{R}(\underline{z})}{R'(\underline{z})}+\hat{\underline{z}}\right]+(1-\ell)\frac{\hat{P}(b)}{P'(b)}+\left(1-\ell+\frac{1}{\ell h''(\ell)}\right)\hat{b}$$
$$+\left(\frac{R(\underline{z})}{R'(\underline{z})}\right)\frac{\hat{R}'(\underline{z})}{R'(\underline{z})}-\left(\frac{R(\underline{z})}{P'(b)}+(1-\ell)\frac{P(b)}{P'(b)}\right)\frac{\hat{P}'(b)}{P'(b)}.$$

Next, recall the incidence of the tax reform on $\hat{\underline{z}}$ and $\hat{b}$ when the utility is logarithmic is given by

$$\hat{\underline{z}}=-\frac{\ell\frac{P(b)}{P'(b)}}{\frac{R(\underline{z})}{R'(\underline{z})}+\ell\frac{P(b)}{P'(b)}}\frac{\hat{R}(\underline{z})}{R'(\underline{z})}+\frac{\frac{R(\underline{z})}{R'(\underline{z})}}{\frac{R(\underline{z})}{R'(\underline{z})}+\ell\frac{P(b)}{P'(b)}}\ell\frac{\hat{P}(b)}{P'(b)}-\frac{R(\underline{z})}{R'(\underline{z})}\ell h''(\ell)\hat{\ell}$$
$$\ell\hat{b}=\frac{\ell\frac{P(b)}{P'(b)}}{\frac{R(\underline{z})}{R'(\underline{z})}+\ell\frac{P(b)}{P'(b)}}\frac{\hat{R}(\underline{z})}{R'(\underline{z})}-\frac{\frac{R(\underline{z})}{R'(\underline{z})}}{\frac{R(\underline{z})}{R'(\underline{z})}+\ell\frac{P(b)}{P'(b)}}\ell\frac{\hat{P}(b)}{P'(b)}+\left(\frac{R(\underline{z})}{P'(b)}+(1-\ell)\frac{P(b)}{P'(b)}\right)\ell h''(\ell)\hat{\ell}.$$

Substitute these expressions into the previous equation to get

$$D\frac{\hat{\ell}}{\ell}=\left[\left(1-\frac{R'(\underline{z})}{P'(b)}\right)\frac{\frac{R(\underline{z})}{R'(\underline{z})}}{\frac{R(\underline{z})}{R'(\underline{z})}+\ell\frac{P(b)}{P'(b)}}-(1-\ell)\frac{\frac{P(b)}{P'(b)}}{\frac{R(\underline{z})}{R'(\underline{z})}+\ell\frac{P(b)}{P'(b)}}-\frac{1}{\ell h''(\ell)}\frac{\frac{P(b)}{P'(b)}}{\frac{R(\underline{z})}{R'(\underline{z})}+\ell\frac{P(b)}{P'(b)}}\right]\frac{\hat{R}(\underline{z})}{R'(\underline{z})}$$
$$+\left[\ell\left(1-\frac{R'(\underline{z})}{P'(b)}\right)\frac{\frac{R(\underline{z})}{R'(\underline{z})}}{\frac{R(\underline{z})}{R'(\underline{z})}+\ell\frac{P(b)}{P'(b)}}-\ell(1-\ell)\frac{\frac{P(b)}{P'(b)}}{\frac{R(\underline{z})}{R'(\underline{z})}+\ell\frac{P(b)}{P'(b)}}+\frac{1}{\ell h''(\ell)}\frac{\frac{R(\underline{z})}{R'(\underline{z})}}{\frac{R(\underline{z})}{R'(\underline{z})}+\ell\frac{P(b)}{P'(b)}}\right]\frac{\hat{P}(b)}{P'(b)}$$
$$-\frac{R(\underline{z})}{R'(\underline{z})}\frac{\hat{R}'(\underline{z})}{R'(\underline{z})}+\left[\frac{R(\underline{z})}{P'(b)}+(1-\ell)\frac{P(b)}{P'(b)}\right]\frac{\hat{P}'(b)}{P'(b)}$$

where

$$D=-\left(1-2\frac{R'(\underline{z})}{P'(b)}\right)\frac{R(\underline{z})}{R'(\underline{z})}-(3\ell-2)\frac{P(b)}{P'(b)}$$
$$-\left[\left(\frac{R'(\underline{z})}{P'(b)}-1\right)\ell\frac{R(\underline{z})}{R'(\underline{z})}-(1-\ell)\left(\frac{R(\underline{z})}{P'(b)}+(1-\ell)\frac{P(b)}{P'(b)}\right)\right]\ell h''(\ell)$$
$$+\left[\left(\frac{R'(\underline{z})}{P'(b)}-1\right)\frac{R(\underline{z})}{R'(\underline{z})}+(1-\ell)\frac{P(b)}{P'(b)}\right]\frac{\ell h'''(\ell)}{h''(\ell)}.$$

Now, differentiate the firm's profit $\Pi(\theta) = \ell\theta - \underline{z} - \ell b$ to get

$$\begin{aligned}
\frac{\partial \Pi(\theta)}{\partial \ell} &= \theta - b - \frac{\partial \underline{z}}{\partial \ell} - \ell \frac{\partial b}{\partial \ell} \\
&= \theta - b + \ell h''(\ell) \left[\frac{1 - \ell \frac{R'(\underline{z})}{P'(b)}}{R'(\underline{z}) u'(R(\underline{z}))} - \frac{1-\ell}{P'(b) u'(R(\underline{z}) + P(b))} \right]
\end{aligned}$$

where we used

$$\begin{aligned}
\frac{\partial \underline{z}}{\partial \ell} &= -\ell h''(\ell) \frac{1}{R'(\underline{z}) u'(R(\underline{z}))} \\
\ell \frac{\partial b}{\partial \ell} &= \ell h''(\ell) \left[\frac{1-\ell}{P'(b) u'(R(\underline{z}) + P(b))} + \frac{\ell}{P'(b) u'(R(\underline{z}))} \right].
\end{aligned}$$

Differentiating once more gives the following expression for $\frac{\partial^2 \Pi(\theta)}{\partial \ell^2}$:

$$\begin{aligned}
&- \frac{\partial b}{\partial \ell} + h''(\ell) \left(1 + \frac{\ell h'''(\ell)}{h''(\ell)} \right) \left[\frac{1 - \ell \frac{R'(\underline{z})}{P'(b)}}{R'(\underline{z}) u'(R(\underline{z}))} - \frac{1-\ell}{P'(b) u'(R(\underline{z}) + P(b))} \right] \\
&- \ell h''(\ell) \left[\frac{1}{P'(b) u'(R(\underline{z}))} + \left(1 - \ell \frac{R'(\underline{z})}{P'(b)} \right) \frac{u''(R(\underline{z}))}{(u'(R(\underline{z})))^2} \frac{\partial \underline{z}}{\partial \ell} \right] \\
&+ \ell h''(\ell) \left[\frac{1}{P'(b) u'(R(\underline{z}) + P(b))} + (1-\ell) \frac{u''(R(\underline{z}) + P(b))}{(u'(R(\underline{z}) + P(b)))^2} \left(\frac{R'(\underline{z})}{P'(b)} \frac{\partial \underline{z}}{\partial \ell} + \frac{\partial b}{\partial \ell} \right) \right]
\end{aligned}$$

i.e., when the utility is logarithmic,

$$\begin{aligned}
\frac{1}{h''(\ell)} \frac{\partial^2 \Pi(\theta)}{\partial \ell^2} = &\left(1 - 2 \frac{R'(\underline{z})}{P'(b)} \right) \frac{R(\underline{z})}{R'(\underline{z})} + (3\ell - 2) \frac{P(b)}{P'(b)} \\
&+ \left[(2\ell - 1) \frac{R(\underline{z})}{P'(b)} - \ell \frac{R(\underline{z})}{R'(\underline{z})} - (1-\ell)^2 \frac{P(b)}{P'(b)} \right] \ell h''(\ell) \\
&- \left[\left(\frac{1}{P'(b)} - \frac{1}{R'(\underline{z})} \right) R(\underline{z}) + (1-\ell) \frac{P(b)}{P'(b)} \right] \frac{\ell h'''(\ell)}{h''(\ell)}.
\end{aligned}$$

As a result, the second-order condition of the firm's problem implies

$$D = -\frac{1}{h''(\ell)} \frac{\partial^2 \Pi(\theta)}{\partial \ell^2} \geq 0.$$

Equation (41) follows. ■

Proof of Proposition 4. Suppose that the tax reform satisfies

$$\frac{\hat{R}(\underline{z})}{R'(\underline{z})} = -\ell\frac{\hat{P}(b)}{P'(b)}.$$

Recall that the utility is log and $R'(\underline{z}) = P'(b) = 1-\tau$ in the baseline tax system. Lemma 9 gives the impact of this perturbation on the worker's base pay:

$$\hat{\underline{z}} = -\frac{\hat{R}(\underline{z})}{1-\tau} - \frac{R(\underline{z})}{1-\tau}\ell h''(\ell)\hat{\ell},$$

on the bonus:

$$\ell\hat{b} = \frac{\hat{R}(\underline{z})}{1-\tau} + \frac{R(\underline{z})+(1-\ell)P(b)}{1-\tau}\ell h''(\ell)\hat{\ell},$$

and on expected utility:

$$\hat{U}(\theta) = 0.$$

If the tax rates are perturbed for types $[\theta^*,\infty)$, the impact of the reform on government revenue $\hat{\mathcal{R}}$ is given by

$$\begin{aligned}&\int_{\theta^*}^{\infty}\left[\hat{T}(\underline{z}(\theta)) + \ell(\theta)\hat{T}_B(b(\theta)) + T_B(b(\theta))\hat{\ell}(\theta) + T'(\underline{z}(\theta))\hat{\underline{z}}(\theta) + T_B'(b(\theta))\ell(\theta)\hat{b}(\theta)\right]dF(\theta)\\ &= \int_{\theta^*}^{\infty}\left[T_B(b(\theta)) + \frac{\tau}{1-\tau}P(b(\theta))\ell(\theta)(1-\ell(\theta))h''(\ell(\theta))\right]\hat{\ell}(\theta)dF(\theta)\end{aligned}$$

where $T(\underline{z}) \equiv \underline{z} - R(\underline{z})$ and $T_B(b) \equiv b - P(b)$, and where the second equality uses the expressions we have derived above for the incidence of the reform around a baseline tax system where $T'(\underline{z}(\theta)) = T_B'(b(\theta)) = \tau$. Since the terms in square brackets are positive, it follows that $\hat{\mathcal{R}} > 0$ if $\hat{\ell}(\theta) > 0$ for all θ.

Now, the incidence of tax reforms on labor effort is given by equation (41). Apply this formula with $\frac{\hat{R}(\underline{z})}{R'(\underline{z})} = -\ell\frac{\hat{P}(b)}{P'(b)}$ to obtain

$$D\frac{\hat{\ell}}{\ell} = \frac{1}{\ell h''(\ell)}\frac{\hat{P}(b)}{P'(b)} + \left[\frac{R(\underline{z})}{P'(b)} + (1-\ell)\frac{P(b)}{P'(b)}\right]\frac{\hat{P}'(b)}{P'(b)} - \frac{R(\underline{z})}{R'(\underline{z})}\frac{\hat{R}'(\underline{z})}{R'(\underline{z})},$$

where $D = -\frac{1}{h''(\ell)}\frac{\partial^2\Pi(\theta)}{\partial\ell^2} \geq 0$. It follows that, if the tax reform lowers the marginal and total tax rate on bonuses, so that $\hat{P}(b) > 0$ and $\hat{P}'(b) > 0$, labor effort unambiguously increases ($\hat{\ell} > 0$) if the reform also implies $\hat{R}'(\underline{z}) < 0$; that is, if the marginal tax rate on base pay increases. Assuming that the tax schedule is initially linear on

$[0, \infty)$ ensures that this is satisfied, since in that case $\ell(\theta)$ and $b(\theta)/\underline{z}(\theta)$ are constant (by Proposition 1); thus, for a linear downward perturbation of the bonus tax rate $\hat{P}(b) = b$, we have $\hat{R}(\underline{z}(\theta)) = -\ell b(\theta) \propto -\underline{z}(\theta)$, i.e., the base pay tax rate is perturbed upwards linearly. ■

Proof of Theorem 3. Suppose bonuses are taxed with a schedule $T_b(b)$ that has a tax rate τ_b in the top bracket $[b^*, \infty)$. Denote by θ^* the type that earns the bonus b^*. The base pay of top bonus earners is taxed at a fixed rate t. Denote the average bonus tax rate at b^* by $t_b \equiv T_b(b^*)/b^*$. The elasticities of aggregate variables with respect to $1 - \tau_b$ used below are constructed by keeping the set of top agents fixed, as in the proof of Theorem 2).

Consider an increase in the top bonus tax rate by $\delta\hat{\tau}_b$ with $\delta \to 0$, i.e., a perturbation $\hat{T}_b(b) = -\hat{\tau}_b(b - b^*)\mathbb{I}_{\{b \geq b^*\}}$ and $\hat{T}_b'(b) = -\hat{\tau}_b\mathbb{I}_{\{b \geq b^*\}}$. At the optimum, the first-order change in social welfare caused by the reform is equal to zero:

$$\int_{\theta^*}^{\infty} \left[\hat{\tau}_b(b - b^*)\ell(1 - \tilde{g}(\bar{z} \mid \theta)) + \tau_z\hat{\underline{z}} + \tau_b\ell\hat{b} + T_b(b)\hat{\ell}\right] \frac{dF(\theta)}{1 - F(\theta^*)} = 0,$$

where, using the expression for $\hat{U}$ derived in Lemma 9, we define the modified social marginal welfare weights as

$$\tilde{g}(\bar{z} \mid \theta) \equiv \frac{1}{\lambda} \frac{\frac{1}{(1-\tau_b)u'(R(\underline{z},b))}}{\frac{1-\frac{1-t}{1-\tau_b}\ell(\theta)}{(1-t)u'(R(\underline{z},0))} + \frac{\ell(\theta)}{(1-\tau_b)u'(R(\underline{z},b))}} \alpha(\theta)u'(R(\underline{z}, b)).$$

Note that, if $\tau_b \geq t$, then $\tilde{g}(\bar{z} \mid \theta) > g(\bar{z} \mid \theta)$. By continuity, there exists $\underline{t} < t$ such that, as long as $\tau_b \geq \underline{t}$, this inequality continues to hold.

Denote the average effort ℓ over top bonus earners by $L = \int_{\theta^*}^{\infty} \ell \frac{dF(\theta)}{1-F(\theta^*)}$, the average frequency-adjusted bonus ℓb by $B = \int_{\theta^*}^{\infty} \ell b \frac{dF(\theta)}{1-F(\theta^*)}$, and the average base pay by $\underline{\mathcal{Z}} = \int_{\theta^*}^{\infty} \underline{z} \frac{dF(\theta)}{1-F(\theta^*)}$. Additionally, denote the bonus-weighted average of the modified marginal social welfare weights in the top bracket by $\tilde{\mathcal{G}} = \int_{\theta^*}^{\infty} \frac{\ell b - \ell b^*}{B - L b^*} \tilde{g}(\bar{z} \mid \theta) \frac{dF(\theta)}{1-F(\theta^*)}$.

We can now rewrite the previous equation as follows. The mechanical effect is equal to $\hat{\tau}_b(B - Lb^*)$. The welfare effect is $-\hat{\tau}_b(B - Lb^*)\tilde{\mathcal{G}}$. The behavioural effects due to bonus and effort changes are equal to

$$\int_{\theta^*}^{\infty}\left[\tau_b \ell \hat{b} + T_b(b)\hat{\ell}\right]\frac{dF(\theta)}{1-F(\theta^*)} = \int_{\theta^*}^{\infty}\left[\tau_b \widehat{\ell b} + (T_b(b)-\tau_b b)\hat{\ell}\right]\frac{dF(\theta)}{1-F(\theta^*)}$$
$$= -\hat{\tau}_b\left(\frac{\tau_b}{1-\tau_b}Be_B - \frac{t_b-\tau_b}{1-\tau_b}Lb^*e_L\right),$$

where we used $T_b(b)-\tau_b b = \tau_b(b-b^*)+t_b b^* - \tau_b b = (t_b-\tau_b)b^*$. The behavioural effect due to base pay responses is equal to:

$$\int_{\theta^*}^{\infty}\tau_z \hat{\underline{z}}\frac{dF(\theta)}{1-F(\theta^*)} = -\hat{\tau}_b\frac{\tau_z}{1-\tau_b}\underline{\mathcal{Z}}e_{\underline{\mathcal{Z}}}.$$

Combining the terms, we get

$$\frac{\tau_b}{1-\tau_b} = \frac{1-\tilde{\mathcal{G}}\frac{\tau_z}{1-\tau_b}\frac{\underline{\mathcal{Z}}}{B}\rho_b e_{\underline{\mathcal{Z}}} - \frac{t_b-\tau_b}{1-\tau_b}(1-\rho_b)e_L}{\rho_b e_B},$$

where $\rho_b = \frac{B/L}{B/L-b^*}$ is the empirical Pareto coefficient of top bonuses.

Structural expression for $e_{\underline{\mathcal{Z}}}$. Using the incidence formulas for the case of separate taxation (Lemma 9), we can write

$$\hat{\underline{z}} = -\frac{\hat{\tau}_b}{1-\tau_b}\varepsilon_{out}\ell(b-b^*) - b\varepsilon_{in}\hat{\ell} = -\frac{\hat{\tau}_b}{1-\tau_b}\left[\varepsilon_{out}\ell(b-b^*) - \ell b\varepsilon_{in}e_\ell\right].$$

Plugging this expression into the definition of $e_{\underline{\mathcal{Z}}} = -\frac{1-\tau_b}{\hat{\tau}_b}\frac{1}{\underline{\mathcal{Z}}}\int_{\theta^*}^{\infty}\hat{\underline{z}}\frac{dF(\theta)}{1-F(\theta^*)}$, we get

$$e_{\underline{\mathcal{Z}}} = \frac{(B-Lb^*)}{\underline{\mathcal{Z}}}\underbrace{\int_{\theta^*}^{\infty}\frac{\ell(b-b^*)}{B-Lb^*}\varepsilon_{out}\frac{dF(\theta)}{1-F(\theta^*)}}_{=\overline{\varepsilon_{out}}} - \frac{B}{\underline{\mathcal{Z}}}\underbrace{\int_{\theta^*}^{\infty}\frac{\ell b}{B}\varepsilon_{in}e_\ell\frac{dF(\theta)}{1-F(\theta^*)}}_{=\overline{\varepsilon_{in}e_\ell}}$$

This expression easily leads to equation (31). ■

C Alternative Models of Performance Pay

C.1 Linear Contracts: Piece Rates and Commissions

Preferences are represented by the utility function $U(c,\ell) = -\frac{1}{\gamma}\exp(-\gamma(c-h(\ell)))$, where h is convex. The income tax schedule is affine: $c = T_0 + (1-\tau)z$. Providing effort ℓ yields output $\theta(\ell+\eta)$, where $\eta \sim \mathcal{N}(0,\sigma_\eta^2)$. The firm observes the worker's output but not her effort nor performance shock. Following Holmstrom and Milgrom (1987), we can restrict attention to linear contracts, i.e., pre-tax earnings are given as a function of observed output by $z = z_0 + \beta\theta(\ell+\eta)$, for some $(z_0,\beta) \in \mathbb{R}^2$. The firm maximizes expected profits $\theta\ell - \mathbb{E}z$ subject to the incentive constraint

$$\ell = \arg\max_{l \geq 0} \mathbb{E}[U(c,l)] \tag{43}$$

and the participation constraint $\mathbb{E}[U(c,\ell] \geq U(\theta)$. The free-entry condition holds and determines the equilibrium reservation values.

We show below that the incentive compatibility constraint (43) implies $h'(\ell) = (1-\tau)\beta\theta$. In other words, if the firm wants to elicit an effort level ℓ from the worker, it must design a contract such that the sensitivity of pay to performance is equal to

$$\beta = \frac{1}{\theta}\frac{h'(\ell)}{1-\tau}.$$

This equation shows the worker's exposure to output risk, measured by the slope of the equilibrium contract, has a similar expression as in our baseline model, and identical crowd-out and crowd-in elasticities $\varepsilon_{\beta,1-p} = -1$ and $\varepsilon_{\beta,\ell} = 1/\varepsilon_\ell^F$. In Section D in the Appendix we derive expressions for the demogrant z_0 and the equilibrium expected utility $U(\theta)$.

The optimal effort level is chosen to maximize the firm's profit. We find that ℓ satisfies

$$h'(\ell) = \frac{(1-\tau)\theta}{1+\gamma h''(\ell)\sigma_\eta^2}. \tag{44}$$

Suppose in particular that $h(\ell) = \frac{\ell^2}{2}$. We then get $\beta = \frac{1}{\theta}\frac{\ell}{1-\tau}$ and $\ell = \frac{(1-\tau)\theta}{1+\gamma\sigma_\eta^2}$. Thus, $\beta = \frac{1}{1+\gamma\sigma_\eta^2}$ is independent of the tax rate. More generally, the net effect of the tax rate on the pass-through is given by

$$\frac{d\ln\beta}{d\ln(1-\tau)} = -1 + \frac{\varepsilon_{\ell,1-\tau}}{\varepsilon_\ell^F},$$

and the elasticity of labor effort with respect to the retention rate $1-\tau$ is given by

$$\varepsilon_{\ell,1-\tau} = \frac{\partial\ln\ell}{\partial\ln(1-\tau)} = \frac{\varepsilon_\ell^F}{1+(1-\beta)\frac{h'(\ell)h'''(\ell)}{h''(\ell)^2}},$$

where $\varepsilon_\ell^F = \frac{h'(\ell)}{\ell h''(\ell)}$ is Frisch elasticity. These expressions imply that an increase in the tax rate leads to an increase in the pass through β, so that the crowd-out dominates the crowd-in, if and only if the labor effort elasticity $\varepsilon_{\ell,1-\tau}$ is smaller than the Frisch elasticity ε_ℓ^F, or equivalently whenever $h'''(\ell) > 0$. If the disutility of effort is isoelastic, this is the case iff $\varepsilon_\ell^F < 1$.

The framework of Holmstrom and Milgrom (1987) allows us to verify that our main prediction—the offsetting of the crowd-out and crowd-in effects—is robust to the degree of risk aversion of workers. Suppose that the Frisch elasticity is constant, in which case the effort elasticity becomes

$$\varepsilon_{\ell,1-\tau} = \frac{\varepsilon_\ell^F}{1 + (1-\beta)(1-\varepsilon_\ell^F)}.$$

This elasticity depends on $\beta = \frac{1}{\theta}\frac{h'(\ell)}{1-\tau}$, which is increasing in the level of effort. Note further that by the first-order condition (44), effort is strictly decreasing in the coefficient of absolute risk aversion γ. Intuitively, motivating effort requires exposing workers to earnings risk, and more risk-averse workers require higher compensation for this risk—a higher z_0—which is costly to the firm. Thus, the firm optimally chooses a lower level of effort when γ is higher. This comparative statics allows us to sharply characterise how risk aversion affects both the labor effort elasticity and the degree to which the crowd-in offsets the crowd-out.

Corollary 2 *The effort elasticity $\varepsilon_{\ell,1-\tau}$ is a monotonic function of the coefficient of absolute risk aversion γ, and takes values between ε_ℓ^F when $\gamma = 0$ and $\frac{\varepsilon_\ell^F}{2-\varepsilon_\ell^F}$ when $\gamma \to \infty$. Thus, $\frac{d\ln\beta}{d\ln(1-\tau)}$ takes values between 0 when $\gamma = 0$ and $-\frac{1-\varepsilon_\ell^F}{2-\varepsilon_\ell^F}$ when $\gamma \to \infty$.*

Suppose that, in line with the existing evidence, Frisch elasticity is equal $\varepsilon_\ell^F = 0.5$. We know that, regardless of the degree of risk aversion, the crowd-in will offset at least two-thirds of the crowd-out: $\frac{d\ln\beta}{d\ln(1-\tau)} > -1/3$. Furthermore, the lower is coeff. of absolute risk aversion γ, the higher is this offset rate, reaching 100 percent when the risk aversion vanishes.

C.2 Convex Contracts: Stock-Options

We now build on the model of performance pay proposed by Edmans and Gabaix (2011). This framework gives rise to convex optimal contracts and has been used to describe forms of executive compensation such as stock options. Here, we focus on a simple version of the model, and we refer to our earlier Working Paper (Doligalski, Ndiaye, and Werquin 2020) for a thorough analysis of taxation a general environment that allows for arbitrary utility function, distribution of performance shocks, and tax schedule.

The setup is similar to our baseline model of Section 1, except that agents can now draw continuous performance shocks. A worker with ability θ who provides effort ℓ produces output $\theta(\ell+\eta)$, where $\eta \in \mathbb{R}$ is a random variable with mean 0. As in Edmans and Gabaix (2011), we impose the following assumption.

Assumption 3 *The agent chooses effort ℓ* after *observing the realization of her performance shock η. The firm recommends the same effort level $\ell(\theta)$ for all agents with the same ability θ.*

Importantly, we assume here that the worker is committed to stay with an employer regardless of the realisation of the performance shock. We relax this assumption in section C.4. Since the design of the contract ensures that effort is incentive compatible, the firm is able to infer the underlying type η from the worker's output. We thus denote the earnings schedule by $z(\theta, \eta)$. The firm's problem is to maximize expected profit (1) subject to the participation constraint (3) and the incentive compatibility constraint, which reads:

$$\ell(\theta) \in \arg\max_{\hat{\ell}} u(R(z(\theta, \eta + \hat{\ell} - \ell(\theta)))) - h(\hat{\ell}), \quad \forall \eta. \tag{45}$$

That is, when the worker exerts effort $\hat{\ell}$, the employer assumes that she has exerted the recommend effort $\ell(\theta)$ and deduces that η is $\eta + \hat{\ell} - \ell(\theta)$ and pays her according to that calculation. Incentive compatibility then implies that $\hat{\ell} = \ell(\theta)$ is optimal. Notice that, in contrast to our baseline framework of Section 1, the effort level $\ell(\theta)$ must maximize utility state-by-state (i.e., for each performance shock realization η) rather than in expectation. This is a consequence of the timing Assumption 3. Finally, the free-entry condition (4) holds.

Assumption 4 *The utility of consumption is logarithmic, $u(c) = \log c$. The Frisch elasticity of labor supply $\varepsilon_\ell^F \equiv h'(\ell)/\ell h''(\ell)$ is constant. The performance shocks are normally distributed, $\eta \sim \mathcal{N}(0, \sigma_\eta^2)$. The tax schedule has a constant rate of progressivity (CRP), $T(z) = z - \frac{1-\tau}{1-p} z^{1-p}$.*

We denote by $\beta \equiv \partial \log z(\theta,\eta)/\partial\eta$ the pass-through of performance shocks to log-earnings. The following proposition characterizes the equilibrium labor contract.

Proposition 5 *The earnings schedule is log-linear and given by:*

$$\log z(\theta,\eta) = \log(\theta\ell) + \beta\eta - \frac{1}{2}\beta^2\sigma_\eta^2 \quad \text{with} \quad \beta = \frac{h'(\ell)}{1-p}. \tag{46}$$

Effort ℓ is independent of θ and satisfies:

$$\ell = [(1-p)(1-\varepsilon_{\beta,\ell}\beta^2\sigma_\eta^2)]^{\varepsilon_\ell^F/(1+\varepsilon_\ell^F)}, \tag{47}$$

where $\varepsilon_{\beta,\ell} \equiv \frac{\partial \log \beta}{\partial \log \ell} = 1/\varepsilon_\ell^F$. Expected utility is given by

$$U(\theta) = \log(R(\theta\ell)) - h(\ell) - \frac{1}{2}(1-p)\beta^2\sigma_\eta^2. \tag{48}$$

Proposition 5 shows that earnings risk, measured by the pass-through parameter β, is constant and has the exact same expression as in our discrete model (equation (6)), namely $\beta = h'(\ell)/(1-p)$. As in Section 1, this property follows immediately by taking the first-order condition in the incentive compatibility constraint (45). This implies in turn that the crowd-out and crowd-in elasticities are given by $\varepsilon_{\beta,1-p} = -1$ and $\varepsilon_{\beta,\ell} = 1/\varepsilon_\ell^F$. Lemma 2 and the subsequent discussion on the relative magnitude of these two forces thus applies identically to this framework. Only the expression for the labor effort elasticity is different, namely,

$$\varepsilon_{\ell,1-p} = \frac{\varepsilon_\ell^F}{1+\varepsilon_\ell^F} \cdot \frac{1+\varepsilon_{\beta,\ell}\beta^2\sigma_\eta^2}{1+\frac{1-\varepsilon_\ell^F}{1+\varepsilon_\ell^F}\varepsilon_{\beta,\ell}\,\beta^2\sigma_\eta^2}.$$

This expression shows that the labor effort elasticity is strictly larger in the presence of moral hazard ($\varepsilon_{\beta,\ell} > 0$) than in the benchmark model with exogenous risk ($\varepsilon_{\beta,\ell} = 0$), due to the marginal cost of incentives (MCI) in the first-order condition for effort.

We can now derive the optimal rate of progressivity in this framework. We obtain the following result.

Theorem 4 *Suppose that the social welfare objective is utilitarian. The optimal rate* *of progressivity satisfies*

$$\frac{p}{(1-p)^2} = \frac{\sigma_\theta^2 + (1+\varepsilon_{\beta,1-p})\beta^2\sigma_\eta^2}{\left[1+\frac{g}{(1-g)p}\right]\varepsilon_{\ell,1-p} + (1-p)\varepsilon_{\beta,\ell}\varepsilon_{\ell,1-p}\beta^2\sigma_\eta^2}. \tag{49}$$

Thus, the optimal rate of progressivity is strictly smaller in the model with endogenous private insurance than in the benchmark environment with exogenous risk where $\varepsilon_{\beta,1-p} = \varepsilon_{\beta,\ell} = 0$.

Interestingly, the optimal tax progressivity in our baseline setting (equation (13)) coincides with formula (49) up to a second order as $\beta \to 0$. We derive further theoretical and quantitative results in Doligalski et al. (2020).

C.3 Dynamic Contracts: Career Incentives

We now extend our results to a dynamic model of the labor market based on the model of Edmans et al. (2012).[47] Workers are indexed by their constant productivity θ. They live for $S \geq 2$ periods and discount the future at rate r. Preferences are separable, logarithmic in consumption and isoelastic in effort. Productivity θ is lognormally distributed with mean μ_θ and variance σ_θ^2. The government levies a CRP income tax given by $R_t(z) = \frac{1-\tau_t}{1-p} z^{1-p}$. The rate of progressivity p is time-independent while the intercept τ_t ensures that the budget is balanced in each period. Private savings are ruled out, so that $c_t = R_t(z_t)$.

We denote the history of a random variable x up to time $t \leq S$ by x^t. Flow output at time t is given by $y_t = \theta(\ell_t + \eta_t)$ where $\{\eta_t\}_{1 \leq t \leq S}$ are i.i.d. random variables. We assume that η_t are normally distributed with mean 0 and variance σ_η^2. As in Section C.2, we assume that the agent chooses period-t effort ℓ_t after observing the realization of the history of performance shocks up to and including time t, η^t. Firms discount future profits at rate r. In each period they observe the agent's productivity and history of output realizations. A labor contract specifies for each t a recommended effort level $\ell_t(\theta)$ and an earnings function $z_t(\theta, \eta^t)$. The firm maximizes its expected profit

$$\Pi(\theta) = \max_{\{\ell_t(\theta), z_t(\theta,\eta^t)\}_{1\leq t\leq S}} \sum_{t=1}^{S} \left(\frac{1}{1+r}\right)^{t-1} \left\{\theta\ell_t - \mathbb{E}_0\left[z_t(\theta,\eta^t)\right]\right\}$$

[47] Our results of Section 1 also extend to the dynamic framework of Sannikov (2008), in which the one-shot deviation principle implies that the sensitivity of utility to output shocks is, again, given by the marginal disutility of effort $h'(\ell)$ (see equation (4) on p. 962).

subject to the incentive constraint:

$$\begin{aligned} & \mathbb{E}_1\left[\sum_{t=1}^{S}\beta^{t-1}(u(R_t(z_t(\theta,\eta^t)))-h(\tilde{\ell}_t(\eta^t)))\right] \qquad (50)\\ \leq\ & \mathbb{E}_1\left[\sum_{t=1}^{S}\beta^{t-1}(u(R_t(z_t(\theta,\eta^t)))-h(\ell_t(\theta)))\right], \quad \forall\{\tilde{\ell}_t(\eta^t)\}_{1\leq t\leq S} \end{aligned}$$

and the participation constraint:

$$\mathbb{E}_0\left[\sum_{t=1}^{S}\beta^{t-1}(u(R(z_t(\theta,\eta^t)))-h(\ell_t(\theta)))\right]\geq U(\theta).$$

The free-entry condition (4) holds.

Proposition 6 *Let* $\sum_{s=0}^{S-t}\left(\frac{1}{1+r}\right)^s \equiv 1/\delta_t$*, and denote the present value of effort by* $L\equiv\sum_{s=1}^{S}\left(\frac{1}{1+r}\right)^{s-1}\ell_s$*. Define the sequence of pass-through parameters* $\{\beta_t\}_{1\leq t\leq S}$ *by*

$$\beta_t=\delta_t\,\frac{h'(\ell_t)}{1-p}. \qquad (51)$$

The earnings schedule satisfies

$$\log(z_t(\theta,\eta^t))=\log(z_{t-1}(\theta,\eta^{t-1}))+\beta_t\eta_t-\frac{1}{2}\beta_t^2\sigma_\eta^2, \qquad (52)$$

where initial earnings are given by $z_0\equiv\delta_1\theta L$*. Period-*$t$ *effort level* ℓ_t *is independent of* θ *and satisfies*

$$\ell_t=\left[(1-p)\left(\frac{\ell_t}{\delta_1 L}-\frac{1}{\delta_t}\varepsilon_{\beta_t,\ell_t}\beta_t^2\sigma_\eta^2\right)\right]^{\varepsilon_\ell^F/(1+\varepsilon_\ell^F)}$$

where $\varepsilon_{\beta_t,\ell_t}=1/\varepsilon_\ell^F$ *is the elasticity of the pass-through parameter* β_t *with respect to effort* ℓ_t*. Expected utility is given by*

$$U(\theta)=\sum_{t=1}^{S}\left(\frac{1}{1+r}\right)^{t-1}\left[u(R(\delta_1\theta L))-h(\ell_t)-\frac{1}{2\delta_t}\beta_t^2\sigma_\eta^2\right].$$

Equation (52) shows that, as in the static setting of Section C.2, earnings in each period t are a log-linear function of the performance shock η_t in that period. Note that $\delta_S = 1$ in the last period, so that β_S is exactly the same as in the static model. In earlier periods we have $\delta_t < 1$ for all $t \leq S-1$, so that the pass-through of output risk is smaller than in the static environment. This is because an increase in output realization in a given period, either due to effort or to random shocks, boosts log-earnings in the current and all future periods equally. Indeed, since the agent is risk-averse it is efficient to spread the rewards over her entire horizon. In other words, a given increase in lifetime utility necessary to elicit higher effort requires a higher increase in flow utility if there are fewer remaining periods over which to smooth these benefits. As a result, the sequence $\{\delta_t\}_{1\leq t\leq S}$ is strictly increasing and the degree of performance pay gets stronger over time. Nevertheless, the pass-through of performance shocks to log-earnings β_t keeps the same expression as in the static model. Thus, our insight that tax progressivity affects the private contract via offsetting crowd-out and crowd-in forces carries over to this dynamic environment.

Theorem 5 *Suppose that the planner is utilitarian. The optimal rate of progressivity is given by*

$$\frac{p}{(1-p)^2} = \frac{\sigma_\theta^2}{\varepsilon_{L,1-p} + (1-p)\sum_{s=1}^{S}\left(\frac{1}{1+r}\right)^{s-1}\frac{\delta_1}{\delta_s}\varepsilon_{\beta_s,\ell_s}\varepsilon_{\ell_s,1-p}\beta_s^2\sigma_\eta^2} \tag{53}$$

where $\varepsilon_{L,1-p}$ is the elasticity of the present discounted value of effort with respect to progressivity, and $\varepsilon_{\beta_s,\ell_s} = 1/\varepsilon_\ell^F$.

Equation (53) is similar to its static counterpart (49). Assuming first that private insurance is exogenous ($\varepsilon_{\beta_s,\ell_s} = \varepsilon_{\beta_s,1-p} = 0$ for all $s \geq 1$), note that the relevant labor effort elasticity is that of the present-value of effort, $\varepsilon_{L,1-p}$. With endogenous earnings risk, the optimal rate of progressivity accounts for the negative fiscal externality due to the crowding-in of private insurance (second term in the denominator). The only difference with the static expression is that the relevant discount factor is not $(1/(1+r))^{s-1}$ but $(1/(1+r))^{s-1}\delta_1/\delta_s$. Since δ_s is increasing over time, this implies that the fiscal externalities caused by the future performance-pay effects are discounted at a higher rate than the standard deadweight losses from distorting effort.

C.4 No Commitment of Workers, Competitive Screening and Adverse Selection

In this section, we consider a modified version of the Edmans and Gabaix (2011) model in which workers cannot commit to stay with their employer once they privately observe their idiosyncratic performance shock η. Effectively, the performance shock becomes a hidden type, and the framework becomes one of adverse selection: workers with heterogenous performance shocks are screened by competitive firms that offer a menu of contracts for workers to select from.

Our results are as follows. First, the equilibrium in this model, taking the income tax schedule as given, always exists and is unique. Second, the equilibrium is very simple: labor effort satisfies the standard first-order condition with the tax rate as the only distortion, workers' earnings are equal to their realized output and there is no private insurance within firms. Below we describe the setup of the model, derive our results, and discuss them in the context of other models of adverse selection.

Setup. The setting is as in Edmans and Gabaix (2011), described in Section C.2. Workers are characterized by a fixed ability θ and a performance shock η. They choose labor effort $\ell \geq 0$ and produce output $y = \theta(\ell + \eta)$. The utility over consumption c and effort ℓ is given by $u(c) - h(\ell)$, where u is strictly increasing and concave, and h is strictly increasing and strictly convex, both continuously differentiable. Competitive firms observe workers' ability θ. Thus, when describing the equilibrium we will focus on a market for a particular realization of ability θ. Furthermore, firms observe the individual's output, but do not observe the performance shock nor labor effort. This means that the performance shock realization is a hidden type. We define a labor contract (y, z) as a pair of output y and earnings z. Firms are competitive: free entry ensures that they cannot make positive profits in equilibrium. To be precise, we use the equilibrium notion of Miyazaki (1977)-Wilson (1977)-Spence (1978), defined formally below. Defining an equilibrium as in M. Rothschild and Stiglitz (1976) would make no difference: the two equilibrium notions always coincide in this setting.

Definition 1 ***(Miyazaki-Wilson-Spence equilibrium)*** *A set of contracts is an equilibrium if (i) firms make zero profits on their overall portfolio of contracts offered, and (ii) there is no other potential contract which would make positive profits, if offered, after all contracts rendered unprofitable by its introduction have been withdrawn.*

We make a number of simplifying assumptions for ease of exposition. These assumptions can be relaxed without affecting the main results.

1. The performance shock η takes two values: η_H with probability $\lambda \in (0,1)$ and η_L with the remaining probability, where $\eta_H > \eta_L \geq 0$.

2. The utility function is quasilinear: $u(c) = c$.

3. The income tax schedule is affine with tax rate t and lump-sum transfer T.

Since we consider a labor market with asymmetric information, the equilibrium set of contracts needs to satisfy the incentive-compatibility (IC) constraints. Denote the contract designed for type η_i, $i \in \{L, H\}$, by (y_i, z_i), with the associated effort level $\ell_i = \frac{y_i}{\theta} - \eta_i$. The IC constraints then ensure that both types of workers have incentives to select the contract that is designed for them:

$$(1-t)z_i - h(\ell_i) \geq (1-t)z_{-i} - h(\ell_{-i} + \eta_{-i} - \eta_i) \quad \forall i \in \{L, H\}. \tag{54}$$

Consider the following candidate for the equilibrium—it is the one we would expect to see in the absence of information frictions. Namely, firms break even on each type of worker and labor effort follows the standard first-order condition. Lemma 10 below shows that in such a candidate equilibrium, the IC constraints are always slack. Proposition 7, which follows, demonstrates further that the candidate is the unique equilibrium and always exists.

Lemma 10 *Consider a candidate equilibrium in which a worker with performance shock η_i, $i \in \{L, H\}$, exerts effort ℓ^* which satisfies $(1-t)\theta = h'(\ell^*)$ and receives pre-tax earnings $z_i = \theta(\ell^* + \eta_i)$. The incentive-compatibility constraints are slack.*

Proof. Rewrite the IC constraint of type L as $h(\ell^* + \eta_H - \eta_L) - h(\ell^*) \geq \theta(1-t)(\eta_H - \eta_L)$. We can bound the left-hand side from below:

$$\begin{aligned} h(\ell^* + \eta_H - \eta_L) - h(\ell^*) &= \int_0^{\eta_H - \eta_L} h'(\ell^* + x)dx \\ &> \int_0^{\eta_H - \eta_L} h'(\ell^*)dx = h'(\ell^*)(\eta_H - \eta_L) = \theta(1-t)(\eta_H - \eta_L), \end{aligned} \tag{55}$$

where the inequality follows from the strict convexity of $h(\cdot)$ and the last equality follows from the first-order condition for effort. Thus, the IC constraint of the L type is always slack. It is straightforward to show the analogous result for type H. ■

Proposition 7 *There exists a unique equilibrium, which is given by the candidate equilibrium from Lemma 10.*

Proof. First we will prove the properties of the equilibrium conditional on existence, then we will prove the existence and uniqueness. The proposition effectively postulates that in equilibrium *(i)* firms break even on each contract and *(ii)* labor effort satisfies the standard first-order condition.

First, suppose that condition *(i)* is not satisfied. Without loss of generality, suppose that the incumbent firm makes positive profits on worker type i by offering a contract (y_i, z_i) with $y_i > z_i$ and is incurring losses on the other type. A profitable deviation for the competitor is to offer a single contract $(\tilde{y}, \tilde{z}) = (y_i, \frac{y_i+z_i}{2})$ to both types. Type i is always better-off accepting the contract, while type $-i$ may or may not be better-off. The competitor makes positive profits on each worker who accepts the contract regardless of their type, since output is higher than wage payments: $\tilde{y} > \tilde{z}$.

Second, suppose that condition *(ii)* is not satisfied for type i. The profitable deviation is then to offer a contract $(\tilde{y}, \tilde{z}) = (\theta(\ell^* + \eta_i)), \theta(\ell^* + \eta_i) - \epsilon)$ for sufficiently small $\epsilon > 0$. It is straightforward to show that for sufficiently small $\epsilon > 0$ such contracts attract type i. Since the contracted output is strictly higher than the wage payment, the competitor makes positive profits on any worker who accepts the new contract. Hence, it does not matter whether type $-i$ is attracted by the contract or not.

The candidate equilibrium of Lemma 10 exists by construction. To prove that it is indeed an equilibrium, we will show that there exists no profitable deviation starting from it. At the candidate equilibrium each worker receives a utility-maximizing contract subject to the employer making no losses on that worker. Consequently, there exists no contract which yields higher utility to a worker and generates positive profits. Thus, there exists no profitable deviation. The uniqueness follows from the strict concavity of h, implying a unique solution ℓ^* to the first-order condition $\theta(1-t) = h'(\ell)$. ■

We find that the equilibrium in the model without workers' commitment and asymmetric information about the performance shock coincides with the allocation that we would expect to see in the absence of information frictions. Thus, the labor market operates exactly as in the standard Mirrlees model with two-dimensional heterogeneity (θ, η). In equilibrium workers are paid exactly what they produce, which can be interpreted as workers having performance-based contracts, e.g., a 100 percent commission or piece rate. However, the exact slope of the contract is not uniquely pinned down. The performance shock is fully passed through to earnings, which means that there is no private insurance within the firm.

It is useful to compare these results to other settings with competitive labor market screening. In particular, Miyazaki (1977) finds rat-race effects (upward effort distortion of high types) and cross-subsidization from high to low types, while Stantcheva (2014), who studies tax policy in Miyazaki's environment, demonstrates that the government can redistribute more than in the standard Mirrlees model. Using the Nash equilibrium notion as in M. Rothschild and Stiglitz (1976) could even lead to a non-existence of equilibrium.

Our model has none of these features. The fundamental difference is in the information available to the firms: in Miyazaki (1977) and Stantcheva (2014), firms observe the worker's effort (hours worked), while in our setting firms observe the worker's output. Screening workers via effort is costly (IC constraints are binding) and complex: it involves additional effort distortions and cross-subsidization between types. By contrast, screening workers via output is easy and costless—IC constraints are slack. In equilibrium with observable output the firm can simply promise each worker a pay equal to the realized output and let workers choose their effort level.

D Proofs of Section C

D.1 Proofs of Section C.1

The incentive constraint reads

$$\ell = \arg\max_{l} -\frac{1}{\gamma}\mathbb{E}\left[e^{-\gamma[T_0+(1-\tau)(z_0+\beta\theta(l+\eta))-h(l)]}\right]$$

Taking the first-order condition implies

$$\mathbb{E}\left[\{(1-\tau)\beta\theta - h'(\ell)\}e^{-\gamma[T_0+(1-\tau)(z_0+\beta\theta(\ell+\eta)-h(\ell)]}\right] = 0$$

and hence

$$(1-\tau)\beta\theta = h'(\ell).$$

The slope of the optimal contract is thus given by $\beta = \frac{1}{\theta}\frac{h'(\ell)}{1-\tau}$. Expected utility is given by

$$\begin{aligned}
\mathbb{E}[U(c,\ell)] =& \mathbb{E}\left[-\frac{1}{\gamma}e^{-\gamma[T_0+(1-\tau)z_0+(1-\tau)\beta\theta(\ell+\eta)-h(\ell)]}\right] \\
=& -\frac{1}{\gamma}e^{-\gamma T_0}e^{-\gamma(1-\tau)z_0}e^{-\gamma(\ell h'(\ell)-h(\ell))}\mathbb{E}\left[e^{-\gamma h'(\ell)\eta}\right] \\
=& -\frac{1}{\gamma}e^{-\gamma T_0}e^{-\gamma(1-\tau)z_0}e^{-\gamma(\ell h'(\ell)-h(\ell))}e^{\frac{1}{2}\gamma^2(h'(\ell))^2\sigma_\eta^2}.
\end{aligned}$$

The participation constraint then implies

$$z_0 = -\frac{\log(-\gamma U(\theta))}{\gamma(1-\tau)} - \frac{T_0}{1-\tau} - \frac{\ell h'(\ell)-h(\ell)}{1-\tau} + \frac{1}{2}\frac{\gamma}{1-\tau}(h'(\ell))^2\sigma_\eta^2$$

Free entry implies $0 = (1-\beta)\theta\ell - z_0$, and hence

$$z_0 = (1-\beta)\theta\ell.$$

Thus expected utility is equal to

$$\begin{aligned}
U(\theta) =& -\frac{1}{\gamma}e^{-\gamma T_0}e^{-\gamma(1-\tau)z_0}e^{-\gamma(\ell h'(\ell)-h(\ell))}e^{\frac{1}{2}\gamma^2(h'(\ell))^2\sigma_\eta^2} \\
=& -\frac{1}{\gamma}e^{-\gamma(1-\tau)(1-\beta)\theta\ell}e^{-\gamma T_0}e^{-\gamma(\ell h'(\ell)-h(\ell))}e^{\frac{1}{2}\gamma^2(h'(\ell))^2\sigma_\eta^2} \\
=& -\frac{1}{\gamma}e^{-\gamma(1-\tau)\theta\ell}e^{\gamma\ell h'(\ell)}e^{-\gamma T_0}e^{-\gamma(\ell h'(\ell)-h(\ell))}e^{\frac{1}{2}\gamma^2(h'(\ell))^2\sigma_\eta^2} \\
=& -\frac{1}{\gamma}e^{-\gamma[T_0+(1-\tau)\theta\ell-h(\ell)]}e^{\frac{1}{2}\gamma^2(h'(\ell))^2\sigma_\eta^2} = U(R(\theta\ell),\ell)e^{\frac{1}{2}\gamma^2(h'(\ell))^2\sigma_\eta^2}.
\end{aligned}$$

Firm profits are given by

$$\begin{aligned}\Pi =&(1-\beta)\theta\ell - z_0\\ =&\left(1-\frac{1}{\theta}\frac{h'(\ell)}{1-\tau}\right)\theta\ell + \frac{\ell h'(\ell)-h(\ell)}{1-\tau} - \frac{1}{2}\frac{\gamma}{1-\tau}(h'(\ell))^2\sigma_\eta^2 + \frac{\log(-\gamma U(\theta))}{\gamma(1-\tau)} + \frac{T_0}{1-\tau}.\end{aligned}$$

The optimal choice of effort maximizes the firm's profits:

$$0 = \theta\left(1-\frac{1}{\theta}\frac{h'(\ell)}{1-\tau}\right) - \frac{\ell h''(\ell)}{1-\tau} + \frac{\ell h''(\ell)}{1-\tau} - \frac{\gamma}{1-\tau}h'(\ell)h''(\ell)\sigma_\eta^2$$

so that

$$h'(\ell) = \frac{(1-\tau)\theta}{1+\gamma h''(\ell)\sigma_\eta^2}.$$

This first-order condition implies that

$$\frac{\partial \ell}{\partial 1-\tau} = \frac{\theta}{h''(\ell)+\gamma\sigma_\eta^2(h''(\ell)^2+h'(\ell)h'''(\ell))}$$

which leads to the following effort elasticity

$$\varepsilon_{\ell,1-\tau} = \frac{\partial \ln \ell}{\partial \ln(1-\tau)} = \frac{\varepsilon_\ell^F}{1+(1-\beta)\frac{h'(\ell)h'''(\ell)}{h''(\ell)^2}},$$

where $\varepsilon_\ell^F = \frac{h'(\ell)}{\ell h''(\ell)}$ is the Frisch elasticity. Assuming that the Frisch elasticity is constant, the effort elasticity becomes $\varepsilon_{\ell,1-\tau} = \frac{\varepsilon_\ell^F}{1+(1-\beta)(1-\varepsilon_\ell^F)}$.

D.2 Proofs of Section C.2

Proof of Proposition 5. Consider first the general case of a concave utility function u and a nonlinear retention function R. Given the earnings contract $\{z(\theta,\eta) : \eta \in \mathbb{R}\}$, an agent with ability θ and performance shock η chooses effort $\ell(\theta)$ to maximize utility $v(z(\theta,\eta)) - h(\ell(\theta))$ with $v = u \circ R$. Equation (45) implies that $\frac{\partial z(\theta,\eta)}{\partial \eta} = \frac{\partial z(\theta,\eta)}{\partial \ell}$ so that the first-order condition reads

$$v'(z(\theta,\eta))\frac{\partial z(\theta,\eta)}{\partial \eta} = h'(\ell(\theta)). \tag{56}$$

This equation pins down the slope of the earnings schedule that the firm must implement in order to induce the effort level $\ell(\theta)$. Integrating this incentive constraint over η given $\ell(\theta)$ leads to

$$v(z(\theta,\eta)) = h'(\ell(\theta))\eta + k, \tag{57}$$

for some constant $k \in \mathbb{R}$. Since in equilibrium the participation constraint (3) must hold with equality, the agent's expected utility must be equal to his reservation value $U(\theta)$. Therefore, the value of k must be chosen by the firm such that the agent's participation constraint holds with equality. Imposing the participation constraint with $\mathbb{E}\eta = 0$ implies

$$k = U(\theta) + h(\ell(\theta)). \tag{58}$$

The previous two equations fully characterize the wage contract given the desired effort level $\ell(\theta)$ and the reservation value $U(\theta)$. They imply that, for a given pair $(a(\theta), U(\theta))$, the wage given performance shock η satisfies:

$$v(z(\theta,\eta)) = h'(\ell(\theta))\eta + [U(\theta) + h(\ell(\theta))]. \tag{59}$$

The first-order condition for effort is obtained by taking the first-order condition with respect to $\ell(\theta)$ in the firm's problem, taking as given the earnings contract required to satisfy the workers' incentive and participation constraints.

Suppose now that the tax schedule is CRP, so that $R(z) = \frac{1-\tau}{1-p} z^{1-p}$. Equation (57) then implies that in order to induce agents with ability θ to choose the same effort ℓ regardless of their noise realization η, the earnings contract must satisfy:

$$\log(z(\theta,\eta)) = \frac{\ell^{\frac{1}{\varepsilon}}}{1-p}\eta - \frac{1}{1-p}\log\frac{1-\tau}{1-p} + \frac{k}{1-p}, \tag{60}$$

for some $k \in \mathbb{R}$. Thus, log-earnings are linear in the performance shock $\eta = \frac{z}{\theta} - \ell$ that the firm infers upon observing realized output z. Imposing that the agent's participation constraint holds with equality pins down the value of k as a function of $U(\theta)$. Namely, equation (58) implies:

$$k = U(\theta) + \frac{1}{1+\frac{1}{\varepsilon_\ell^F}} \ell^{1+\frac{1}{\varepsilon_\ell^F}}$$

and hence

$$\log(z(\theta,\eta)) = \frac{\ell^{\frac{1}{\varepsilon_\ell^F}}}{1-p}\eta + \frac{1}{1-p}\frac{1}{1+\frac{1}{\varepsilon_\ell^F}}\ell^{1+\frac{1}{\varepsilon_\ell^F}} - \frac{1}{1-p}\log\frac{1-\tau}{1-p} + \frac{U(\theta)}{1-p}. \tag{61}$$

Below we derive the equilibrium value of the reservation utility $U(\theta)$ and obtain the equilibrium wage given (ℓ,η):

$$\log(z(\theta,\eta)) = \log(\theta\ell) + \frac{\ell^{\frac{1}{\varepsilon_\ell^F}}}{1-p}\eta - \frac{1}{2}\left(\frac{\ell^{\frac{1}{\varepsilon_\ell^F}}}{1-p}\right)^2\sigma_\eta^2. \tag{62}$$

Define the sensitivity of the before-tax and after-tax wages to output in the optimal contract by the semi-elasticities $\beta(\theta,\eta) \equiv \frac{1}{z(\theta,\eta)}\frac{\partial z(\theta,\eta)}{\partial\eta}$ and $\beta^c(\theta,\eta) \equiv \frac{1}{R(z(\theta,\eta))}\frac{\partial R(z(\theta,\eta))}{\partial\eta}$, respectively. We have $\beta(\theta,\eta) = \frac{\ell^{1/\varepsilon_\ell^F}}{1-p}$ and $\beta^c(\theta,\eta) = \ell^{1/\varepsilon_\ell^F}$. Both $\beta(\theta,\eta)$ and $\beta^c(\theta,\eta)$ depend on the tax schedule through its effect on optimal effort, and there is an additional crowding-out effect on the before-tax sensitivity.

Next, since $v'(z) = \frac{R'(z)}{R(z)} = \frac{1-p}{z}$ the firm's first-order condition reads

$$\begin{aligned}\theta &= \mathbb{E}\left[\frac{h'(\ell)}{v'(z(\theta,\eta))} + \frac{h''(\ell)}{v'(z(\theta,\eta))}\eta\right] \\ &= \frac{\ell^{\frac{1}{\varepsilon_\ell^F}}}{1-p}\mathbb{E}[z(\theta,\eta)] + \frac{1}{\varepsilon_\ell^F}\frac{\ell^{\frac{1}{\varepsilon_\ell^F}-1}}{1-p}\mathbb{E}[z(\theta,\eta)\eta].\end{aligned}$$

We have

$$\begin{aligned}\mathbb{E}[z(\theta,\eta)] &= \mathbb{E}\left[e^{\frac{\ell^{1/\varepsilon_\ell^F}}{1-p}\eta}\right]e^{\frac{1}{1-p}\frac{1}{1+1/\varepsilon_\ell^F}\ell^{1+1/\varepsilon_\ell^F} - \frac{1}{1-p}\log\frac{1-\tau}{1-p} + \frac{U(\theta)}{1-p}} \\ &= e^{\frac{1}{2}\frac{\ell^{2/\varepsilon_\ell^F}}{(1-p)^2}\sigma_\eta^2}e^{\frac{1}{1-p}\frac{1}{1+1/\varepsilon_\ell^F}\ell^{1+1/\varepsilon_\ell^F} - \frac{1}{1-p}\log\frac{1-\tau}{1-p} + \frac{U(\theta)}{1-p}}.\end{aligned}$$

where we used the fact that that η is normally distributed with mean 0 and variance σ_η^2 so that $\mathbb{E}[e^{x\eta}] = e^{\frac{1}{2}x^2\sigma_\eta^2}$ for any x. Moreover, we have $\mathbb{E}[\eta e^{x\eta}] = x\sigma^2 e^{\frac{1}{2}x^2\sigma_\eta^2}$ for any x. Indeed, let φ the (normal) pdf of η. We have $\varphi'(\eta) = -\frac{\eta}{\sigma_\eta^2}\varphi(\eta)$, so that $\mathbb{E}[\eta e^{x\eta}] = \int \eta e^{x\eta}\varphi(\eta)d\eta = -\sigma_\eta^2\int e^{x\eta}\varphi'(\eta)d\eta = x\sigma_\eta^2\int e^{x\eta}\varphi(\eta)d\eta = x\sigma_\eta^2 e^{\frac{1}{2}x^2\sigma_\eta^2}$, where the third equality follows from an integration by parts.

$$\begin{aligned}\mathbb{E}[z(\theta,\eta)\eta] &= \mathbb{E}\Big[\eta e^{\frac{\ell^{1/\varepsilon_\ell^F}}{1-p}\eta}\Big] e^{\frac{1}{1-p}\frac{1}{1+1/\varepsilon_\ell^F}\ell^{1+1/\varepsilon_\ell^F}-\frac{1}{1-p}\log\frac{1-\tau}{1-p}+\frac{U(\theta)}{1-p}} \\ &= \frac{\ell^{1/\varepsilon_\ell^F}}{1-p}\sigma_\eta^2 e^{\frac{1}{2}\frac{\ell^{2/\varepsilon_\ell^F}}{(1-p)^2}\sigma_\eta^2} e^{\frac{1}{1-p}\frac{1}{1+1/\varepsilon_\ell^F}\ell^{1+1/\varepsilon_\ell^F}-\frac{1}{1-p}\log\frac{1-\tau}{1-p}+\frac{U(\theta)}{1-p}}.\end{aligned}$$

Plugging these expressions into the firm's first order condition leads to

$$\theta\ell = \left[\frac{\ell^{1+1/\varepsilon_\ell^F}}{1-p} + \frac{1}{\varepsilon_\ell^F}\frac{\ell^{2/\varepsilon_\ell^F}}{(1-p)^2}\sigma_\eta^2\right] e^{\frac{1}{2}\frac{\ell^{2/\varepsilon_\ell^F}}{(1-p)^2}\sigma_\eta^2} e^{\frac{1}{1-p}\frac{1}{1+1/\varepsilon_\ell^F}\ell^{1+1/\varepsilon_\ell^F}-\frac{1}{1-p}\log\frac{1-\tau}{1-p}+\frac{U(\theta)}{1-p}}$$

and hence

$$\frac{\ell^{1+\frac{1}{\varepsilon_\ell^F}}}{1-p} + \frac{1}{\varepsilon_\ell^F}\frac{\ell^{2/\varepsilon_\ell^F}}{(1-p)^2}\sigma_\eta^2 = \theta\ell e^{-\frac{1}{1-p}\frac{1}{1+1/\varepsilon_\ell^F}\ell^{1+1/\varepsilon_\ell^F}-\frac{1}{2}\frac{\ell^{2/\varepsilon_\ell^F}}{(1-p)^2}\sigma_\eta^2+\frac{1}{1-p}\log\frac{1-\tau}{1-p}-\frac{U(\theta)}{1-p}}.$$

Now use the free-entry condition and the expression derived above for $\mathbb{E}[z(\theta,\eta)]$ to get

$$e^{\frac{1}{1-p}\frac{1}{1+1/\varepsilon_\ell^F}\ell^{1+1/\varepsilon_\ell^F}+\frac{1}{2}\frac{\ell^{2/\varepsilon_\ell^F}}{(1-p)^2}\sigma_\eta^2-\frac{1}{1-p}\log\frac{1-\tau}{1-p}+\frac{U(\theta)}{1-p}} = \theta\ell. \tag{63}$$

Combining this equation with the first-order condition for optimal effort therefore leads to:

$$\ell^{1+1/\varepsilon_\ell^F} + \frac{1}{\varepsilon_\ell^F}\frac{\ell^{2/\varepsilon_\ell^F}}{1-p}\sigma_\eta^2 = 1-p. \tag{64}$$

Using the definition $\beta \equiv \frac{\ell^{\frac{1}{\varepsilon_\ell^F}}}{1-p}$ for the pass-through easily leads to (47). Note that if $\varepsilon_\ell^F = 1$, we obtain optimal effort in closed form:

$$\ell = \left(\frac{1}{1-p} + \frac{\sigma_\eta^2}{(1-p)^2}\right)^{-1/2}. \tag{65}$$

Taking logs in equation (63) easily leads to (48).

Differentiating equation (64) with respect to $(1-p)$ leads to

$$\left[\left(1+\frac{1}{\varepsilon_\ell^F}\right)\ell^{1/\varepsilon_\ell^F} + \frac{2\sigma_\eta^2}{(1-p)(\varepsilon_\ell^F)^2}\ell^{2/\varepsilon_\ell^F-1}\right]\frac{\partial\ell}{\partial(1-p)} - \frac{\sigma_\eta^2}{(1-p)^2\varepsilon_\ell^F}\ell^{2/\varepsilon_\ell^F} = 1,$$

and hence

$$\left[\left(1+\frac{1}{\varepsilon_\ell^F}\right)\ell^{1/\varepsilon_\ell^F+1}+\frac{2\sigma_\eta^2}{(1-p)(\varepsilon_\ell^F)^2}\ell^{2/\varepsilon_\ell^F}\right]\varepsilon_{\ell,1-p}-\frac{\sigma_\eta^2}{(1-p)\varepsilon_\ell^F}\ell^{2/\varepsilon_\ell^F}=1-p.$$

Using the first-order condition again to substitute for $1-p$ leads to

$$\varepsilon_{\ell,1-p}=\frac{\ell^{1/\varepsilon_\ell^F+1}+\frac{2\sigma_\eta^2}{(1-p)\varepsilon_\ell^F}\ell^{2/\varepsilon_\ell^F}}{\left(1+\frac{1}{\varepsilon_\ell^F}\right)\ell^{1/\varepsilon_\ell^F+1}+\frac{2\sigma_\eta^2}{(1-p)(\varepsilon_\ell^F)^2}\ell^{2/\varepsilon_\ell^F}}.$$

We finally express this elasticity in terms of the pass-through elasticities. We have $\beta=\frac{\ell^{1/\varepsilon_\ell^F}}{1-p}$ and $\varepsilon_{\beta,\ell}=1/\varepsilon_\ell^F$. We can thus write

$$\varepsilon_{\ell,1-p}=\frac{\ell^{1/\varepsilon_\ell^F+1}+2(1-p)\varepsilon_{\beta,\ell}\beta^2\sigma_\eta^2}{\left(1+\frac{1}{\varepsilon_\ell^F}\right)\ell^{1/\varepsilon_\ell^F+1}+\frac{2}{\varepsilon_\ell^F}(1-p)\varepsilon_{\beta,\ell}\beta^2\sigma_\eta^2}.$$

But the first-order condition for labor effort reads

$$\ell^{1+1/\varepsilon_\ell^F}=(1-p)(1-\varepsilon_{\beta,\ell}\beta^2\sigma_\eta^2).$$

Substituting into the previous equation and rearranging terms leads to

$$\varepsilon_{\ell,1-p}=\frac{1+\varepsilon_{\beta,\ell}\beta^2\sigma_\eta^2}{\left(1+\frac{1}{\varepsilon_\ell^F}\right)+\left(\frac{1}{\varepsilon_\ell^F}-1\right)\varepsilon_{\beta,\ell}\beta^2\sigma_\eta^2}.$$

This easily yields the expression given in the text. ∎

Proof of Theorem 4. Recall that the earnings schedule of agents with ability θ can be written as

$$\log(z(\theta,\eta))=\log(\theta\ell)+\beta\eta-\frac{1}{2}(\beta\sigma_\eta)^2$$

and their expected utility as

$$U(\theta)=\log\frac{1-\tau}{1-p}+(1-p)\log(\theta\ell)-\frac{1}{2}(1-p)(\beta\sigma_\eta)^2-h(\ell).$$

Utilitarian social welfare is therefore equal to

$$\int_\Theta U(\theta)\mathrm{d}F(\theta) = (1-p)\mu_\theta + (1-p)\log\ell - (1-p)\frac{\beta^2\sigma_\eta^2}{2} - h(\ell) + \log\frac{1-\tau}{1-p}.$$

The first-order condition for effort, taking taxes as given, reads

$$0 = \frac{\partial U(\theta)}{\partial \ell} = (1-p)\frac{1}{\ell} - (1-p)\beta\sigma_\eta^2\frac{\partial\beta}{\partial\ell} - h'(\ell).$$

Now recall that expected pre-tax and post-tax earnings are respectively given by $\mathbb{E}[z(\theta,\eta)] = \theta\ell$ and $\mathbb{E}[(z(\theta,\eta))^{1-p}] = (\theta\ell)^{1-p}e^{-\frac{p\ell^{2/\varepsilon_\ell^F}\sigma_\eta^2}{2(1-p)}}$, so that government revenue is equal to

$$\int_\Theta \mathbb{E}[R(z(\theta,\eta))]f(\theta)d\theta = \ell e^{\mu_\theta+\frac{\sigma_\theta^2}{2}} - \frac{1-\tau}{1-p}e^{-\frac{p\ell^{2/\varepsilon_\ell^F}\sigma_\eta^2}{2(1-p)}}\ell^{1-p}e^{(1-p)\mu_\theta+(1-p)^2\frac{\sigma_\theta^2}{2}}.$$

Budget balance thus requires

$$\frac{1-\tau}{1-p} = \frac{\ell e^{\mu_\theta+\frac{\sigma_\theta^2}{2}} - G}{e^{-\frac{p\ell^{2/\varepsilon_\ell^F}\sigma_\eta^2}{2(1-p)}}\ell^{1-p}e^{(1-p)\mu_\theta+(1-p)^2\frac{\sigma_\theta^2}{2}}} = \frac{(1-g)\ell e^{\mu_\theta+\frac{\sigma_\theta^2}{2}}}{e^{-\frac{p\ell^{2/\varepsilon_\ell^F}\sigma_\eta^2}{2(1-p)}}\ell^{1-p}e^{(1-p)\mu_\theta+(1-p)^2\frac{\sigma_\theta^2}{2}}}.$$

As a result, maximizing with respect to $1-p$ leads to:

$$\begin{aligned}0 = {}& \mu_\theta + \log\ell + (1-p)\frac{1}{\ell}\frac{\partial\ell}{\partial(1-p)} - h'(\ell)\frac{\partial\ell}{\partial(1-p)} - \frac{\beta^2\sigma_\eta^2}{2}\\ & - (1-p)\beta\sigma_\eta^2\left[\frac{\partial\beta}{\partial(1-p)} + \frac{\partial\beta}{\partial\ell}\frac{\partial\ell}{\partial(1-p)}\right] + \frac{\partial\log\frac{1-\tau}{1-p}}{\partial(1-p)},\end{aligned}$$

with

$$\begin{aligned}\frac{\partial\log\frac{1-\tau}{1-p}}{\partial(1-p)} = {}& \frac{g}{1-g}\frac{\partial\log\ell}{\partial(1-p)} - \mu_\theta - (1-p)\sigma_\theta^2 - \log\ell + p\frac{1}{\ell}\frac{\partial\ell}{\partial(1-p)}\\ & - \left(\frac{1}{2}-p\right)\beta^2\sigma_\eta^2 + p(1-p)\beta\sigma_\eta^2\left[\frac{\partial\beta}{\partial(1-p)} + \frac{\partial\beta}{\partial\ell}\frac{\partial\ell}{\partial(1-p)}\right].\end{aligned}$$

We therefore obtain

$$
\begin{aligned}
0 = &\left[(1-p)\frac{1}{\ell} - h'(\ell) - (1-p)\beta\sigma_\eta^2\frac{\partial\beta}{\partial\ell}\right]\frac{\partial\ell}{\partial(1-p)} + p\frac{1}{\ell}\frac{\partial\ell}{\partial(1-p)} + \frac{g}{1-g}\frac{\partial\log\ell}{\partial(1-p)} \\
&- (1-p)\sigma_\theta^2 - (1-p)\beta^2\sigma_\eta^2 - (1-p)^2\beta\sigma_\eta^2\frac{\partial\beta}{\partial(1-p)} + p(1-p)\beta\sigma_\eta^2\frac{\partial\beta}{\partial\ell}\frac{\partial\ell}{\partial(1-p)}.
\end{aligned}
$$

Using the first-order condition for effort leads to

$$
\begin{aligned}
0 = &\frac{1}{1-p}\left[p + \frac{g}{1-g}\right]\varepsilon_{\ell,1-p} + p\beta^2\sigma_\eta^2\varepsilon_{\beta,\ell}\varepsilon_{\ell,1-p} \\
&- (1-p)[\sigma_\theta^2 + \psi^2\sigma_\eta^2] - (1-p)\beta^2\sigma_\eta^2\varepsilon_{\beta,1-p}.
\end{aligned}
$$

Rearranging this equation leads to the result. ■

D.3 Proofs of Section C.3

Lemma 11 *The earnings process $z_t(\theta, \eta^t)$ is a martingale. That is, expected period-t earnings are equal to realized period-$(t-1)$ earnings,*

$$\mathbb{E}_{t-1}[z_t(\theta, \eta^{t-1}, \eta_t)] = z_{t-1}(\theta, \eta^{t-1}).$$

Proof of Lemma 11. Starting from an incentive compatible allocation, consider the following variations in retained earnings and utility:

$$\hat{u}_{t-1} = v(z_{t-1}(\theta, \eta^{t-1})) - \frac{1}{1+r}\Delta$$

and

$$\hat{u}_t = v(z_t(\theta, \eta^{t-1}, \eta_t)) + \Delta$$

and $\hat{u}_s = v(z_s(\theta, \eta^s))$ for all $s \notin \{t-1, t\}$. These perturbations preserve utility and incentive compatibility since for all ℓ_{t-1},

$$\hat{u}_{t-1} - h(\ell_{t-1}) + \frac{1}{1+r}\mathbb{E}_{t-1}[\hat{u}_t] = v(z_{t-1}(\theta, \eta^{t-1})) - h(\ell_{t-1}) + \frac{1}{1+r}\mathbb{E}_{t-1}[v(z_t(\theta, \eta^{t-1}, \eta_t))].$$

The optimal allocation must be unaffected by such deviations, so that

$$0 = \arg\min_{\Delta} \mathbb{E}\left[\sum_{s=1}^{S}(1+r)^{-t}(z_s - v^{-1}(\hat{u}_s))\right].$$

The associated first-order condition evaluated at $\Delta = 0$ reads

$$\mathbb{E}\left[\frac{1}{v'(z_t(\theta,\eta^{t-1},\eta_t))} \mid z^{t-1}\right] = \frac{1}{v'(z_{t-1}(\theta,\eta^{t-1}))}.$$

The inverse Euler equation (see Golosov et al. (2003)) holds in our setting. With log utility and a CRP tax schedule, this equation can be rewritten as

$$(1-p)\mathbb{E}[z_t(\theta,\eta^{t-1},\eta_t) \mid z^{t-1}] = (1-p)z_{t-1}(\theta,\eta^{t-1}),$$

which leads to the result. ■

Proof of Proposition 6. We provide a heuristic proof of this proposition; the formal argument follows the same steps as in Edmans et al. (2012). Assume that a unique level of effort is implemented at each time t, that this effort level is independent of past output noise, and that local incentive constraints are sufficient conditions. Consider a local deviation in effort ℓ_t after history (η^{t-1},η_t). By incentive compatibility the effect of such a deviation on the worker's lifetime utility U should be zero,

$$\mathbb{E}_{t-1}\left[\frac{\partial U}{\partial z_t}\frac{\partial z_t}{\partial \ell_t} + \frac{\partial U}{\partial \ell_t}\right] = 0.$$

Since $\frac{\partial z_t}{\partial \ell_t} = \theta$, we obtain

$$\mathbb{E}_{t-1}\left[\frac{\partial U}{\partial z_t}\right] = -\frac{1}{\theta}\frac{\partial U}{\partial \ell_t} \tag{66}$$

Applying incentive compatibility for effort in the final period we obtain:

$$v'(z_S(\theta,\eta^S))\frac{\partial z(\theta,\eta^{S-1},\eta_S)}{\partial \eta_S} = h'(\ell_S(\theta)).$$

Fixing η^{S-1} and integrating this incentive constraint over η_S leads to

$$v(z_S(\theta,\eta^S)) = h'(\ell_S(\theta))\eta_S + g^{S-1}(\eta^{S-1})$$

for some function of past output $g^{S-1}(\eta^{S-1})$. This implies in particular that

$$\frac{\partial v(z_S(\theta,\eta^S))}{\partial \eta_{S-1}} = \frac{\partial g^{S-1}(\eta^{S-1})}{\partial \eta_{S-1}}.$$

Analogously, the incentive constraint for effort in the second to last period reads

$$v'(z_{S-1}(\theta,\eta^{S-1}))\frac{\partial z_{S-1}(\theta,\eta^{S-1})}{\partial \eta_{S-1}} + \frac{1}{1+r}v'(z_S(\theta,\eta^S))\frac{\partial z_S(\theta,\eta^S)}{\partial \eta_{S-1}} = h'(\ell_{S-1}(\theta)).$$

Integrating over η_{S-1} and using the previous equation implies

$$v(z_{S-1}(\theta,\eta^{S-1})) + \frac{1}{1+r}g^{S-1}(\eta^{S-1}) = h'(\ell_{S-1}(\theta))\eta_{S-1} + g^{S-2}(\eta^{S-2}).$$

We now want to show that $g^{S-1}(\eta^{S-1})$ is a linear function of η_{S-1}. Since the utility function is logarithmic and the tax schedule is CRP, we obtain

$$(1-p)\log(z_S(\theta,\eta^S)) = h'(\ell_S(\theta))\eta_S + g^{S-1}(\eta^{S-1}) - \log\frac{1-\tau_S}{1-p}$$

and

$$\begin{aligned}&(1-p)\log(z_{S-1}(\theta,\eta^{S-1}))\\ &= h'(\ell_{S-1}(\theta))\eta_{S-1} - \frac{1}{1+r}g^{S-1}(\eta^{S-1}) + g^{S-2}(\eta^{S-2}) - \log\frac{1-\tau_{S-1}}{1-p}.\end{aligned}$$

Now recall that the inverse Euler equation reads

$$\mathbb{E}_{S-1}[z_S(\theta,\eta^S)] = z_{S-1}(\theta,\eta^{S-1}).$$

Using the previous expressions, this equality can be rewritten as

$$\begin{aligned}&\mathbb{E}_{S-1}\left[e^{\frac{1}{1-p}h'(\ell_S(\theta))\eta_S}\right]e^{\frac{1}{1-p}g^{S-1}(\eta^{S-1})}\\ &= \left(\frac{1-\tau_S}{1-\tau_{S-1}}\right)^{\frac{1}{1-p}} e^{\frac{1}{1-p}h'(\ell_{S-1}(\theta))\eta_{S-1}}e^{-\frac{1}{1+r}\frac{1}{1-p}g^{S-1}(\eta^{S-1})+\frac{1}{1-p}g^{S-2}(\eta^{S-2})}.\end{aligned}$$

This in turn implies

$$\left(1+\frac{1}{1+r}\right)g^{S-1}(\eta^{S-1})$$
$$= h'(\ell_{S-1}(\theta))\eta_{S-1} + g^{S-2}(\eta^{S-2}) - \frac{1}{2}\frac{(h'(\ell_S(\theta)))^2}{1-p}\sigma_\eta^2 + \frac{1}{1-p}\log\frac{1-\tau_S}{1-\tau_{S-1}}.$$

Therefore, $g^{S-1}(\eta^{S-1})$, and in turn $v(z_{S-1}(\theta,\eta^{S-1}))$, is linear in η_{S-1}. Moreover, the last-period utility is linear in both η_S and η_{S-1}. By induction, we can show that the utility in each period is a linear function of the performance shock in every past period. Now suppose for simplicity that $S=2$, $r=0$, $\theta=1$, so that $\delta_1=\frac{1}{2}$ and $\delta_2=1$. From the arguments above we guess a log-linear specification for earnings:

$$\log z_1 = \beta_1\eta_1 + k_1$$
$$\log z_2 = \beta_{21}\eta_1 + \beta_2\eta_2 + k_1 + k_2.$$

The martingale property derived above requires $z_1 = \mathbb{E}_1[z_2]$, so that for all η_1, $e^{\beta_1\eta_1+k_1} = e^{\beta_{21}\eta_1+k_1}\mathbb{E}[e^{\beta_2\eta_2+k_2} \mid \eta_1]$. This requires $\beta_1=\beta_{21}$ and $e^{-k_2}=\mathbb{E}[e^{\beta_2\eta_2}\mid\eta_1]$. Now, the total utility of the agent is given by

$$U = (1-p)[2\beta_1\eta_1 + \beta_2\eta_2 + 2k_1 + k_2]$$
$$- h(\ell_1) - h(\ell_2) + \log\frac{1-\tau_1}{1-p} + \log\frac{1-\tau_2}{1-p}.$$

The incentive constraint for effort (66) implies

$$\beta_1 = \frac{h'(\ell_1)}{2(1-p)}, \quad \text{and} \quad \beta_2 = \frac{h'(\ell_2)}{1-p}$$

and therefore

$$k_2 = -\frac{h'(\ell_2)}{1-p} - \frac{\sigma_\eta^2}{2}\left(\frac{h'(\ell_2)}{1-p}\right)^2.$$

Replacing in the expression for log earnings leads to

$$\log z_1 = k_1' + \frac{h'(\ell_1)}{2(1-p)}\eta_1 - \frac{\sigma_\eta^2}{2}\left(\frac{h'(\ell_1)}{2(1-p)}\right)^2$$

and

$$\log z_2 = k_1' + \frac{h'(\ell_1)}{2(1-p)}\eta_1 - \frac{\sigma_\eta^2}{2}\left(\frac{h'(\ell_1)}{2(1-p)}\right)^2 + \left(\frac{h'(\ell_2)}{1-p}\right)\eta_2 - \frac{\sigma_\eta^2}{2}\left(\frac{h'(\ell_2)}{1-p}\right)^2,$$

where $k_1' \equiv k_1 + \beta_1\ell_1 - \frac{\sigma_\eta^2}{2}\beta_1^2$. This constant is pinned down by the zero profit condition $\mathbb{E}[z_1 + z_2] = \ell_1 + \ell_2$, that is, $2e^{k_1'} = \ell_1 + \ell_2$. This implies

$$k_1' = \log\frac{\ell_1 + \ell_2}{2},$$

which concludes the proof of equation (52). The expressions for optimal effort and utility are derived in the next proof. ■

Proof of Theorem 5. Recall that the earnings schedule is given by

$$\begin{aligned}\log z_1 &= \log(\delta_1\theta L) + \beta_1\eta_1 - \frac{\beta_1^2\sigma_\eta^2}{2},\\ \log z_t &= \log z_{t-1} + \beta_t\eta_t - \frac{\beta_t^2\sigma_\eta^2}{2}.\end{aligned}$$

The expected utility of workers with productivity θ is therefore equal to

$$\begin{aligned}U(\theta) = (1-p)&\left[\frac{1}{\delta_1}\log(\delta_1\theta L) - \sum_{s=1}^{S}\left(\frac{1}{1+r}\right)^{s-1}\frac{1}{\delta_s}\frac{\beta_s^2\sigma_\eta^2}{2}\right]\\ &-\sum_{s=1}^{S}\left(\frac{1}{1+r}\right)^{s-1}h(\ell_s) + \sum_{s=1}^{S}\left(\frac{1}{1+r}\right)^{s-1}\log\frac{1-\tau_s}{1-p},\end{aligned}$$

from which the expression given in the text easily follows. Thus, utilitarian social welfare is

$$\begin{aligned}\int_\Theta U(\theta)\mathrm{d}F(\theta) = (1-p)&\left[\frac{1}{\delta_1}\log(\delta_1 L) + \frac{1}{\delta_1}\mu_\theta - \sum_{s=1}^{S}\left(\frac{1}{1+r}\right)^{s-1}\frac{1}{\delta_s}\frac{\beta_s^2\sigma_\eta^2}{2}\right]\\ &-\sum_{s=1}^{S}\left(\frac{1}{1+r}\right)^{s-1}h(\ell_s) + \sum_{s=1}^{S}\left(\frac{1}{1+r}\right)^{s-1}\log\frac{1-\tau_s}{1-p}.\end{aligned}$$

The first-order condition for optimal effort reads

$$0 = \frac{\partial U(\theta)}{\partial \ell_t} = (1-p)\left[\frac{1}{\delta_1}\frac{1}{L}\frac{\partial L}{\partial \ell_t} - \left(\frac{1}{1+r}\right)^{t-1}\frac{1}{\delta_t}\beta_t\sigma_\eta^2\frac{\partial \beta_t}{\partial \ell_t}\right] - \left(\frac{1}{1+r}\right)^{t-1} h'(\ell_t)$$
$$= (1-p)\left[\frac{1}{\delta_1}\frac{\left(\frac{1}{1+r}\right)^{t-1}\ell_t}{L} - \left(\frac{1}{1+r}\right)^{t-1}\frac{1}{\delta_t}\beta_t^2\sigma_\eta^2\varepsilon_{\beta_t,\ell_t}\right]\frac{1}{\ell_t} - \left(\frac{1}{1+r}\right)^{t-1} h'(\ell_t),$$

which easily leads to the equation given in the text. Now, the expected present value of pre-tax and post-tax earnings in period t are given by $\mathbb{E}[z_t] = \delta_1\theta L$ and

$$\mathbb{E}[z_t^{1-p}] = (\delta_1\theta L)^{1-p}\mathbb{E}\left[e^{\sum_{s=1}^t (1-p)\beta_s\eta_s}\right] e^{-\sum_{s=1}^t (1-p)\frac{\beta_s^2\sigma_\eta^2}{2}} = (\delta_1\theta L)^{1-p} e^{-p(1-p)\sum_{s=1}^t \frac{\beta_s^2\sigma_\eta^2}{2}}$$

respectively, so that expected government revenue in period t is equal to

$$\int_\Theta \mathbb{E}[T(z_t)]\mathrm{d}F(\theta)$$
$$= \delta_1 L e^{\mu_\theta + \frac{\sigma_\theta^2}{2}} - \frac{1-\tau_t}{1-p}(\delta_1 L)^{1-p} e^{-p(1-p)\sum_{s=1}^t \frac{\beta_s^2\sigma_\eta^2}{2}} e^{(1-p)\mu_\theta + (1-p)^2\frac{\sigma_\theta^2}{2}}.$$

Imposing period-by-period budget balance therefore requires

$$\frac{1-\tau_t}{1-p} = \frac{(\delta_1 L)^p e^{\mu_\theta + \frac{\sigma_\theta^2}{2}}}{e^{-p(1-p)(\sum_{s=1}^t \beta_s^2)\frac{\sigma_\eta^2}{2}} e^{(1-p)\mu_\theta + (1-p)^2\frac{\sigma_\theta^2}{2}}}.$$

Substituting this expression into the social welfare function $\int_\Theta U(\theta)\mathrm{d}F(\theta)$ implies that social welfare is equal to

$$\frac{1}{\delta_1}\left[\log(\delta_1 L) + \mu_\theta + (1-(1-p)^2)\frac{\sigma_\theta^2}{2}\right] - \sum_{s=1}^S \left(\frac{1}{1+r}\right)^{s-1} h(\ell_s)$$
$$+ p(1-p)\sum_{s=1}^S \left(\frac{1}{1+r}\right)^{s-1}\sum_{i=1}^s \frac{\beta_i^2\sigma_\eta^2}{2} - (1-p)\sum_{s=1}^S \left(\frac{1}{1+r}\right)^{s-1}\frac{1}{\delta_s}\frac{\beta_s^2\sigma_\eta^2}{2}$$
$$= \frac{1}{\delta_1}\left[\log(\delta_1 L) + \mu_\theta + (1-(1-p)^2)\frac{\sigma_\theta^2}{2}\right] - \sum_{s=1}^S \left(\frac{1}{1+r}\right)^{s-1} h(\ell_s)$$
$$- (1-p)^2\sum_{s=1}^S \left(\frac{1}{1+r}\right)^{s-1}\frac{1}{\delta_s}\frac{\beta_s^2\sigma_\eta^2}{2}.$$

We can now maximize this expression with respect to $1-p$ to get

$$\sum_{s=1}^{S}\left[\frac{1}{\delta_1}\frac{1}{L}\left(\frac{1}{1+r}\right)^{s-1}-\left(\frac{1}{1+r}\right)^{s-1}h'(\ell_s)\right]\frac{\partial \ell_s}{\partial(1-p)}$$
$$-(1-p)\sum_{s=1}^{S}\left(\frac{1}{1+r}\right)^{s-1}\frac{1}{\delta_s}\varepsilon_{\beta_s,\ell_s}\varepsilon_{\ell_s,1-p}\beta_s^2\sigma_\eta^2$$
$$=(1-p)\left[\frac{1}{\delta_1}\sigma_\theta^2+\sum_{s=1}^{S}\left(\frac{1}{1+r}\right)^{s-1}\frac{1}{\delta_s}\beta_s^2\sigma_\eta^2\right]+(1-p)\sum_{s=1}^{S}\left(\frac{1}{1+r}\right)^{s-1}\frac{1}{\delta_s}\varepsilon_{\beta_s,1-p}\beta_s^2\sigma_\eta^2.$$

Using the first-order condition for effort derived above to simplify the left hand side of this expression implies

$$\frac{p}{1-p}\frac{1}{\delta_1 L}\sum_{s=1}^{S}\left(\frac{1}{1+r}\right)^{s-1}\ell_s\varepsilon_{\ell_s,1-p}+p\sum_{s=1}^{S}\left(\frac{1}{1+r}\right)^{s-1}\frac{1}{\delta_s}\varepsilon_{\beta_s,\ell_s}\varepsilon_{\ell_s,1-p}\beta_s^2\sigma_\eta^2$$
$$=(1-p)\left[\frac{1}{\delta_1}\sigma_\theta^2+\sum_{s=1}^{S}\left(\frac{1}{1+r}\right)^{s-1}\frac{1}{\delta_s}(1+\varepsilon_{\beta_s,1-p})\beta_s^2\sigma_\eta^2\right].$$

But the elasticity of the present discounted value of effort is equal to

$$\varepsilon_{L,1-p}\equiv\frac{1-p}{L}\frac{\partial\sum_{s=1}^{S}\left(\frac{1}{1+r}\right)^{s-1}\ell_s}{\partial(1-p)}=\sum_{s=1}^{S}\left(\frac{1}{1+r}\right)^{s-1}\frac{\ell_s}{L}\varepsilon_{\ell_s,1-p}.$$

Moreover, we have $1+\varepsilon_{\beta_s,1-p}=0$. Substituting these two expressions into the previous equation and rearranging terms leads to

$$\frac{p}{(1-p)^2}\left[\frac{1}{\delta_1}\varepsilon_{L,1-p}+(1-p)\sum_{s=1}^{S}\left(\frac{1}{1+r}\right)^{s-1}\frac{1}{\delta_s}\varepsilon_{\beta_s,\ell_s}\varepsilon_{\ell_s,1-p}\beta_s^2\sigma_\eta^2\right]=\frac{1}{\delta_1}\sigma_\theta^2.$$

This concludes the proof. ■

References

Abrahám, A., Alvarez-Parra, F., & Forstner, S. (2016). The effects of moral hazard on wage inequality in a frictional labor market. *Unpublished working paper*.

Ábrahám, Á., Koehne, S., & Pavoni, N. (2016). Optimal income taxation when asset taxation is limited. *Journal of Public Economics*, *136*, 14–29.

Alvarez, F., & Jermann, U. J. (2000). Efficiency, equilibrium, and asset pricing with risk of default. *Econometrica*, *68*(4), 775–797.

Attanasio, O., & Rıos-Rull, J.-V. (2000). Consumption smoothing in island economies: Can public insurance reduce welfare? *European Economic Review*, *44*(7), 1225–1258.

Bandiera, O., Barankay, I., & Rasul, I. (2005). Social preferences and the response to incentives: Evidence from personnel data. *The Quarterly Journal of Economics*, *120*(3), 917–962.

Bell, B., & Van Reenen, J. (2013). Extreme wage inequality: pay at the very top. *American economic review*, *103*(3), 153–57.

Bell, B., & Van Reenen, J. (2014). Bankers and their bonuses. *The Economic Journal*, *124*(574), F1–F21.

Bénabou, R., & Tirole, J. (2016). Bonus culture: Competitive pay, screening, and multitasking. *Journal of Political Economy*, *124*(2), 305–370.

Besley, T., & Ghatak, M. (2013). Bailouts and the optimal taxation of bonus pay. *American Economic Review*, *103*(3), 163–67.

Boadway, R., & Sato, M. (2015). Optimal income taxation with risky earnings: A synthesis. *Journal of Public Economic Theory*, *17*(6), 773–801.

Brown, C., & Andrabi, T. (2020). Inducing positive sorting through performance pay: Experimental evidence from pakistani schools. *University of California at Berkeley Working Paper*.

Chang, Y., & Park, Y. (2017). *Optimal taxation with private insurance* (Tech. Rep.).

Chetty, R., Guren, A., Manoli, D., & Weber, A. (2011). Are micro and macro labor supply elasticities consistent? a review of evidence on the intensive and extensive margins. *American Economic Review*, *101*(3), 471–75.

Chetty, R., & Saez, E. (2010). Optimal taxation and social insurance with endogenous private insurance. *American Economic Journal. Economic Policy*, *2*(2), 85.

Colombi, R. (1990). A new model of income distribution: The pareto-lognormal distribution. In *Income and wealth distribution, inequality and poverty* (pp. 18–32). Springer.

Craig, A. C. (2019). Optimal income taxation with spillovers from employer learning.

Cremer, H., & Pestieau, P. (1996). Redistributive taxation and social insurance. *International Tax and Public Finance*, *3*(3), 281–295.

Cullen, J. B., & Gruber, J. (2000). Does unemployment insurance crowd out spousal labor supply? *Journal of labor Economics*, *18*(3), 546–572.

Cutler, D. M., & Gruber, J. (1996a). Does public insurance crowd out private insurance? *The Quarterly Journal of Economics*, *111*(2), 391–430.

Cutler, D. M., & Gruber, J. (1996b). The effect of medicaid expansions on public insurance, private insurance, and redistribution. *The American Economic Review*, *86*(2), 378–383.

Doligalski, P. (2019). Optimal income taxation and commitment on the labor market.

Doligalski, P., Ndiaye, A., & Werquin, N. (2020). Redistribution with performance pay. *Available at SSRN 3581882*.

Eaton, J., & Rosen, H. S. (1980). Optimal redistributive taxation and uncertainty. *The Quarterly Journal of Economics*, *95*(2), 357–364.

Edmans, A., & Gabaix, X. (2011). Tractability in incentive contracting. *The Review of Financial Studies*, *24*(9), 2865–2894.

Edmans, A., Gabaix, X., Sadzik, T., & Sannikov, Y. (2012). Dynamic ceo compensation. *The Journal of Finance*, *67*(5), 1603–1647.

Ferey, A., Haufler, A., & Perroni, C. (2022). Incentives, globalization, and redistribution. *Mimeo*.

Findeisen, S., & Sachs, D. (2016). Education and optimal dynamic taxation: The role of income-contingent student loans. *Journal of Public Economics*, *138*, 1–21.

Frydman, C., & Molloy, R. S. (2011). Does tax policy affect executive compensation? evidence from postwar tax reforms. *Journal of Public Economics*, *95*(11-12), 1425–1437.

Gietl, D., & Haufler, A. (2018). Bonus taxes and international competition for bank managers. *European Economic Review*, *110*, 41–60.

Gittleman, M., & Pierce, B. (2013). How prevalent is performance-related pay in the united states? current incidence and recent trends. *National Institute Economic Review*, *226*, R4–R16.

Golosov, M., Kocherlakota, N., & Tsyvinski, A. (2003). Optimal indirect and capital taxation. *Review of Economic Studies*, 569–587.

Golosov, M., & Tsyvinski, A. (2007). Optimal taxation with endogenous insurance markets. *The Quarterly Journal of Economics*, *122*(2), 487–534.

Grigsby, J., Hurst, E., & Yildirmaz, A. (2019). *Aggregate nominal wage adjustments: New evidence from administrative payroll data* (Tech. Rep.). National Bureau of Economic Research.

Hall, B. J., & Liebman, J. B. (2000). The taxation of executive compensation. *Tax policy and the economy*, *14*, 1–44.

Haufler, A., & Nishimura, Y. (2022). Taxing mobile and overconfident top earners. *International Tax and Public Finance, Forthcoming*.

Heathcote, J., Storesletten, K., & Violante, G. L. (2017). Optimal tax progressivity: An analytical framework. *The Quarterly Journal of Economics*, *132*(4), 1693–1754.

Heathcote, J., & Tsujiyama, H. (2021). Optimal income taxation: Mirrlees meets ramsey. *Journal of Political Economy*, *129*(11), 3141–3184.

Hendren, N. (2016). The policy elasticity. *Tax Policy and the Economy*, *30*(1), 51–89.

Holmstrom, B. (1984). *A note on the first-order approach in principal agent models* (mimeo).

Holmstrom, B., & Milgrom, P. (1987). Aggregation and linearity in the provision of intertemporal incentives. *Econometrica: Journal of the Econometric Society*, 303–328.

Kapička, M., & Neira, J. (2019). Optimal taxation with risky human capital. *American Economic Journal: Macroeconomics*.

Kaplow, L. (1991). Incentives and government relief for risk. *Journal of risk and Uncertainty*, *4*(2), 167–175.

Keane, M. P. (2011). Labor supply and taxes: A survey. *Journal of Economic Literature*, *49*(4), 961–1075.

Krueger, D., & Perri, F. (2011). Public versus private risk sharing. *Journal of Economic Theory*, *146*(3), 920–956.

Lazear, E. P. (2000). Performance pay and productivity. *American Economic Review*, *90*(5), 1346–1361.

Lazear, E. P. (2018). Compensation and incentives in the workplace. *Journal of Economic Perspectives*, *32*(3), 195–214.

Leaver, C., Ozier, O., Serneels, P., & Zeitlin, A. (2021). Recruitment, effort, and retention effects of performance contracts for civil servants: Experimental evidence from rwandan primary schools. *American economic review*, *111*(7), 2213–46.

Lehmann, E., Parmentier, A., & Van der Linden, B. (2011). Optimal income taxation with endogenous participation and search unemployment. *Journal of Public Economics*, *95*(11-12), 1523–1537.

Lemieux, T., MacLeod, W. B., & Parent, D. (2009). Performance pay and wage inequality. *The Quarterly Journal of Economics*, *124*(1), 1–49.

Levitt, S. D., & Syverson, C. (2008). Market distortions when agents are better informed: The value of information in real estate transactions. *The Review of Economics and Statistics*, *90*(4), 599–611.

Makris, M., & Pavan, A. (2021). Taxation under learning by doing. *Journal of Political Economy*, *129*(6), 1878–1944.

Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *The review of economic studies*, 175–208.

Miyazaki, H. (1977). The rat race and internal labor markets. *The Bell Journal of Economics*, 394–418.

Netzer, N., & Scheuer, F. (2007). Taxation, insurance, and precautionary labor. *Journal of Public Economics*, *91*(7-8), 1519–1531.

Park, Y. (2014). Optimal taxation in a limited commitment economy. *Review of Economic Studies*, *81*(2), 884–918.

Piketty, T., & Saez, E. (2003). Income inequality in the united states, 1913–1998. *The Quarterly journal of economics*, *118*(1), 1–41.

Raj, A. (2019). *Optimal income taxation in the presence of networks of altruism* (Tech. Rep.). Working Paper.

Rogerson, W. P. (1985). The first-order approach to principal-agent problems. *Econometrica: Journal of the Econometric Society*, 1357–1367.

Rose, N. L., & Wolfram, C. (2002). Regulating executive pay: Using the tax code to influence chief executive officer compensation. *Journal of Labor Economics*, *20*(S2), S138–S175.

Rothschild, C., & Scheuer, F. (2016). Optimal taxation with rent-seeking. *The Review of Economic Studies*, *83*(3), 1225–1262.

Rothschild, M., & Stiglitz, J. (1976). Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. *The Quarterly Journal of*

Economics, *90*(4), 629–649.

Sachs, D., Tsyvinski, A., & Werquin, N. (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), 205–29.

Saez, E., & Stantcheva, S. (2016). Generalized social marginal welfare weights for optimal tax theory. *American Economic Review*, *106*(1), 24–45.

Sannikov, Y. (2008). A continuous-time version of the principal-agent problem. *The Review of Economic Studies*, *75*(3), 957–984.

Scheuer, F. (2013). Adverse selection in credit markets and regressive profit taxation. *Journal of Economic Theory*, *148*(4), 1333–1360.

Schoeni, R. F. (2002). Does unemployment insurance displace familial assistance? *Public Choice*, *110*(1), 99–119.

Shearer, B. (2004). Piece rates, fixed wages and incentives: Evidence from a field experiment. *The Review of Economic Studies*, *71*(2), 513–534.

Shourideh, A. (2014). Optimal taxation of wealthy individuals. *Working paper*.

Sleet, C., & Yazici, H. (2017). Taxation, redistribution and frictional labor supply. In *2017 meeting papers298 (society for economic dynamics, 2017).*

Spence, M. (1978). Product differentiation and performance in insurance markets. *Journal of Public Economics*, *10*(3), 427–447.

Stantcheva, S. (2014). Optimal income taxation with adverse selection in the labour market. *The Review of Economic Studies*, rdu005.

Stantcheva, S. (2017). Optimal taxation and human capital policies over the life cycle. *Journal of Political Economy*, *125*(6), 1931–1990.

Thanassoulis, J. (2012). The case for intervening in bankers' pay. *The Journal of Finance*, *67*(3), 849–895.

Varian, H. R. (1980). Redistributive taxation as social insurance. *Journal of public Economics*, *14*(1), 49–68.

Wilson, C. (1977). A model of insurance markets with incomplete information. *Journal of Economic Theory*, *16*(2), 167–207.