Washington Center
for Equitable Growth
Evidence for a stronger economy

Working paper series

Redistribution with Performance Pay

Pawel Doligalski

Abdoulaye Ndiaye

Nicolas Werquin

May 2020

https://equitablegrowth.org/working-papers/redistribution-with-performance-pay/

Redistribution with Performance Pay[*]

Paweł Doligalski

University of Bristol

Abdoulaye Ndiaye

New York University

Nicolas Werquin

Toulouse School of Economics

May 21, 2020

Abstract

Half of the jobs in the U.S. feature pay-for-performance. We study nonlinear income taxation in a model where such labor contracts arise as a result of moral hazard frictions within firms. We derive novel formulas for the incidence of arbitrarily nonlinear reforms of a given tax code on both average earnings and their sensitivity to output risk. We show theoretically and quantitatively that, following an increase in tax progressivity, the higher sensitivity of earnings to performance caused by the crowding-out of private insurance is almost fully offset by a countervailing performance-pay effect driven by labor supply responses. As a result, earnings risk is hardly affected by policy. We then turn to the normative analysis of a government that levies taxes and transfers to redistribute income across workers with different levels of uninsurable productivity. We find that setting taxes without accounting for the endogeneity of private insurance is close to optimal. Thus, the common concern that standard models of taxation underestimate the cost of redistribution is, in the context of performance-based compensation, overblown.

Introduction

What do fruit harvesters, real estate brokers, bankers and CEOs have in common? All of them are paid based on their performance. Performance-pay contracts have become increasingly popular across the income distribution. Empirically, a large share – roughly half – of all the jobs in the U.S. involves performance-based compensation (Lemieux, MacLeod, and Parent (2009)) in the form of piece rates, commissions, bonuses, and stock options. These contracts are qualitatively different from usual wage contracts. Indeed, the structure of earnings is designed not only to compensate the employee for completing the job, but also to provide incentives for effort in the first place. When wages are highly sensitive to performance – incentives are high-powered – employees are generously rewarded for better outcomes, but at the same time they are also more exposed to risk. Crucially, we expect both the level and the performance-sensitivity of these contracts to be endogenous to the tax policy implemented by the government. Yet despite the prevalence of these compensation schemes, they have not been systematically studied in the taxation literature. We fill this gap. In a general and tractable framework we derive in closed form the incidence of tax reforms on the earnings and utility that performance-pay workers receive in equilibrium. We also derive the impact of taxes on government revenue and social welfare, as well as the optimal rate of tax progressivity in the presence of such realistic labor contracts.

A widespread concern is that traditional models of income taxation in the tradition of Mirrlees (1971) substantially overstate the optimal level of taxes, by assuming that heterogeneity in wage rates is exogenous and policy-invariant.[1] Instead, when wage risk is endogenous, increasing the progressivity of income taxes should lead to a crowding-out of private insurance provided by firms, that is, a one-for-one spread of the pre-tax earnings distribution. Theoretically, this crowding-out has been shown to be of critical importance in various contexts – in particular by Attanasio and Ríos-Rull (2000), Golosov and Tsyvinski (2007), and Krueger and Perri (2011) – where it severely limits the ability of governments to provide social insurance. Empirically, evidence of such crowding-out has been highlighted in several markets, for instance unemployment or health insurance – see Cullen and Gruber (2000); Schoeni (2002);

---

[1]This is the case both in the static (for instance Saez (2001)) and dynamic (for instance Golosov, Kocherlakota, and Tsyvinski (2003); Farhi and Werning (2013); Golosov, Troshkin, and Tsyvinski (2016)) frameworks.

Cutler and Gruber (1996a,b). Yet the empirical literature that studies the impact of income taxes on the structure of performance-pay contracts often fails to find significant crowding-out effects, see for instance Rose and Wolfram (2002); Frydman and Molloy (2011). Our paper reconciles these findings by highlighting a countervailing force that keeps earnings risk practically unaffected by tax policy. This novel "performance-pay" effect is driven by labor supply adjustments. Under a more progressive tax code, the worker's optimal level of effort is lower. The firm elicits this labor supply reduction by providing more insurance (crowding-*in*). We find that this performance-pay effect almost fully offsets the crowding-out.

We set up a model in which income inequality arises from two distinct sources, namely, innate ability differences, and ex-post performance shocks that affect the output of equally talented workers. While the former source of wage disparities cannot be insured by private markets, the latter is very much shaped in the labor market. In the presence of moral hazard frictions, wage risk has a productive role: employers choose the amount of risk faced by their employees through performance-based pay contracts in order to strike a balance between insurance and incentives for effort. Our modeling of labor markets is based on those of Edmans and Gabaix (2011) for our static setting, and Edmans, Gabaix, Sadzik, and Sannikov (2012) for our dynamic setting. Their frameworks have been very successful at explaining the empirical features of actual performance-based contracts (see Edmans and Gabaix (2016)). We extend them to incorporate sophisticated nonlinear policy instruments. The key technical breakthrough is that we allow for *arbitrarily nonlinear* tax instruments. Previous models of moral hazard were tractable only under very restricted forms of the utility of consumption – for instance, Holmstrom and Milgrom (1987) impose exponential utility functions. This makes it impossible to consider a wide class of tax schedules – typically, they would have to be restricted to being affine – since nonlinear taxes effectively modify the concavity of the utility that workers receive from their salaries. Instead, the analysis of Edmans and Gabaix (2011) remains tractable for very general utility functions. Therefore it allows us to study the incidence of arbitrary tax reforms (say, increasing taxes on the rich, or altering the shape of the EITC) of any initial tax schedule (say, the U.S. tax code). Our analysis is thus very general and can be used for both positive and normative investigation. The government has an effective role to play despite the fact that private insurance markets are constrained efficient. Indeed, while firms optimally provide *insurance* against ex-post output risk, the government

2

uses tax policy for *redistribution* between workers with different ex-ante ability.

We start with a positive analysis of the incidence of tax reforms on the workers' labor contracts and the distribution of utilities. In standard models with exogenous wage risk, taxes affect earnings only by modifying individual labor effort decisions. In our framework, wage risk is endogenous to policy as well. We show that it responds to tax changes via two channels: a *crowding-out* effect, and a *performance-pay* effect. On the one hand, the crowding-out effect is the optimal response of firms to a change in social insurance: they adjust the earnings contract endogenously so that the workers' incentives for effort and participation constraints remain satisfied after the reform. Thus, following an improvement in social insurance (higher tax progressivity), firms respond by spreading the pre-tax earnings schedule. The performance-pay effect, on the other hand, arises from the optimal labor supply adjustment to the tax reform. As in standard models of income taxation, workers' optimal effort is lower in response to an increase in marginal tax rates or tax progressivity. But eliciting a lower effort level in the presence of moral hazard frictions is achieved by lowering the sensitivity of pre-tax earnings to performance, that is, by compressing the wage distribution. This effect counteracts the direct crowding-out of private insurance that the tax reform induces. Crucially, because our model is tractable, we are able to derive this tax incidence analysis entirely in closed form for an arbitrary baseline tax system and arbitrary tax reforms.

We show both theoretically and quantitatively in a calibrated version of our model that the two earnings risk adjustments almost fully offset each other in response to an increase in the progressivity of the tax code. Taken separately these effects are both significant, but summing them implies that taxes barely affect the sensitivity of pay to compensation. Moreover, this result is robust to the value of the labor supply elasticity. The fundamental reason is that the sensitivity of the contract to performance is proportional to the marginal disutility of labor. As a result, in order to elicit a given increase in labor effort, the firm must increase the pass-through of output risk to earnings proportionally to the *inverse* of the labor supply elasticity. Therefore, if labor effort is relatively inelastic, the change in performance-sensitivity necessary to elicit the optimal effort change must be large, and vice versa. We evaluate the robustness of this result to other canonical tax reforms and show that in all cases, the performance-pay offsets at least fifty percent, and in some cases even dominates, the direct crowding-out effect.

Armed with this tax incidence analysis, we then derive the impact of tax reforms on government revenue and social welfare, as well as the optimal level of tax progressivity. This analysis extends Chetty and Saez (2010) to our environment with arbitrarily nonlinear taxes. In addition to the standard effects obtained in the benchmark model with exogenous wage risk, the crowding-out and performance-pay effects create fiscal externalities: given an initially progressive tax code, a spread (respectively, contraction) of the pre-tax earnings distribution impacts positively (resp., negatively) the government budget. Moreover, the crowding-out effect has a first-order negative impact on social welfare. This is because, following a tax reform, firms adjust wages in a way that renders tax cuts less accurately targeted than in a model with exogenous risk. This modifies the relevant social welfare weights in the direction of less redistribution. We then impose a number of functional form assumptions to make the analysis as transparent as possible and obtain sharper results. In particular, we assume that the nonlinear tax schedule is restricted to having a constant rate of progressivity (as in, for instance, Heathcote, Storesletten, and Violante (2017)). Within this class of tax schedules, we derive the optimal rate of progressivity in closed form and show that it is smaller than when wage risk is considered exogenous. However, the welfare losses from setting taxes suboptimally by ignoring the endogeneity of wage risk are quantitatively limited, equivalent to a mere 0.24% drop in consumption. This is because only roughly half of the jobs in our calibration are performance-pay, which reduces the aggregate welfare losses from ignoring the endogeneity of wage risk to a quarter of what they would be if all jobs were subject to agency frictions. We conclude that the common concern that standard models overstate optimal tax policy by ignoring the endogeneity of private insurance is – in the context of performance-pay jobs – overblown.

Literature Review. The two papers that are closest to ours are Golosov and Tsyvinski (2007) and Chetty and Saez (2010). Golosov and Tsyvinski (2007) study an economy in which firms insure their workers subject to unobservable productivity and hidden asset trades. They show that tax reforms generate a large crowding out effect which reduces the gains from public insurance. The government optimally refrains from providing insurance and instead uses tax policy to correct the externality generated by hidden trades. In our environment, markets are constrained efficient. Instead, we study how the redistributive motive for government intervention interacts

with the endogenous private insurance on the labor market. Chetty and Saez (2010) derive a sufficient statistics formula for the optimal linear tax in the presence of linear private insurance contracts. We extend their analysis in two ways. First, and most importantly, rather than following an approach based purely on endogenous sufficient statistics – in particular, the elasticity of crowd-out with respect to tax policy – we study a tractable structural microfoundation for the equilibrium labor contracts. This allows us to characterize analytically the effects of government policy on private insurance contracts via crowding-out and performance-pay responses, and derive explicit theoretical formulas for tax incidence and optimal taxes. Second, we allow for arbitrarily nonlinear taxes in the equilibrium with nonlinear incentive contracts, and we show that several novel effects arise from theses nonlinearities.

Our paper is motivated by the large literature that studies performance-pay contracts as an optimal way for firms to incentivize workers' effort in the presence of moral hazard frictions. On the theoretical side, our baseline framework is the model of Edmans and Gabaix (2011) for our static setting, and that of Edmans et al. (2012) for our dynamic setting. These models have been very successful at explaining the structure of performance-pay contracts of CEOs (Frydman and Jenter (2010); Edmans and Gabaix (2016); Edmans, Gabaix, and Jenter (2017)). On the empirical side, there is growing reduced-form and structural evidence that moral hazard in labor markets is pervasive (Foster and Rosenzweig (1994); Prendergast (1999); Shearer (2004); Lazear and Oyer (2010); Bandiera, Barankay, and Rasul (2011); Ábrahám, Alvarez-Parra, and Forstner (2016a)), that employers are important providers of insurance for their employees (Guiso, Pistaferri, and Schivardi (2005); Lamadon (2016); Friedrich, Laun, Meghir, and Pistaferri (2019); Lamadon, Mogstad, and Setzler (2019)), and that the fraction of jobs with explicit pay-for-performance is high and rising (Lemieux, MacLeod, and Parent (2009); Bloom and Van Reenen (2010); Bell and Van Reenen (2014); Grigsby, Hurst, and Yildirmaz (2019)). Analogous to Kaplow (1991), our key contribution to this large literature is to analyze the effects of *policy* in such environments where the worker-firm relationship is modeled as a moral hazard problem. Crucially, our policy instruments are very general, yet our analysis remains tractable. Our results can help guide future empirical analysis on the impact of taxes on the level and structure of performance-pay packages in the spirit of Rose and Wolfram (2002); Frydman and Molloy (2011); Bird (2018); Dale-Olsen (2012).

Several other papers study optimal taxation with endogenous earnings risk. These

papers focus on risk generated by human capital accumulation (Kapicka and Neira (2013); Findeisen and Sachs (2016); Stantcheva (2017); Makris and Pavan (2017)), job search (Sleet and Yazici (2017)), or wage randomization in response to excessive tax regressivity (Doligalski (2019)). Blomqvist and Horn (1984); Rochet (1991); Cremer and Pestieau (1996) studied the joint design of optimal insurance and redistribution but in these papers the government is the sole provider of insurance. Another strand in the taxation literature studies government taxation in the presence of endogenous consumption insurance, understood either as informal exchanges in family networks or asset trades. Attanasio and Rıos-Rull (2000) and Krueger and Perri (2011) demonstrate a potentially large crowding-out of private insurance in response to increased public insurance. Park (2014); Ábrahám et al. (2016b); Heathcote et al. (2017); Chang and Park (2017); Raj (2019), characterize the optimal tax systems in such economies. In contrast to these papers, it is *pre-tax* earnings risk – rather than *consumption* risk – that is endogenous to policy in our model. Finally, several papers in the optimal taxation literature allow wages to be determined on private labor markets, for instance Hungerbühler, Lehmann, Parmentier, and Van der Linden (2006); Rothschild and Scheuer (2013, 2016, 2014); Stantcheva (2014); Piketty, Saez, and Stantcheva (2014); Scheuer and Werning (2017, 2016); Ales, Kurnaz, and Sleet (2015); Ales and Sleet (2016); Ales, Bellofatto, and Wang (2017); Sachs, Tsyvinski, and Werquin (2020). These papers do not account for wage-rate *risk* and performance-based earnings caused by moral hazard frictions.

Outline of the Paper. Our paper is organized as follows. We set up our baseline static environment in Section 1. In Section 2, we analyze the incidence of arbitrary nonlinear tax reforms on the structure of performance-based compensation and on the distribution of utilities. In Section 3, we derive the excess burden and the social welfare gains of tax reforms. We then focus on a special case of our model to derive sharper results, as well as the optimal rate of progressivity, in Section 4. We study our results quantitatively in Section 5. The proofs, extensions of our baseline model, and the dynamic analysis are gathered in Appendices A to H.

1 Environment

1.1 Labor Market

Individuals. There is a continuum of mass one of agents indexed by their exogenous innate ability $\theta \in \Theta \subset \mathbb{R}_+$ distributed according to the c.d.f. $F(\theta)$. Their preferences over consumption c and labor effort $a \geq 0$ are represented by a separable utility function $u(c) - h(a)$, where u and h are twice continuously differentiable, u is concave, and h is strictly convex. An agent with earnings[2] w pays a tax liability $T(w)$ and consumes $c = w - T(w)$. The tax schedule $T : \mathbb{R}_+ \to \mathbb{R}$ is twice continuously differentiable. We denote by $R(w) \equiv w - T(w)$ the retention function and by $r(w) \equiv R'(w) = 1 - T'(w)$ the retention (or net-of-tax) rate. We assume that the utility of earnings $w \mapsto v(w) \equiv u(R(w))$ is concave.[3]

Labor Contract. A worker with ability θ who provides effort a produces output

$$y = \theta(a + \eta), \tag{1}$$

where the "performance shock" $\eta \in \mathbb{R}$ is a random variable with mean 0, distributed on a (possibly unbounded) interval with interior $(\underline{\eta}, \bar{\eta})$. The firm observes both the agent's ability θ and her realized output y, but cannot disentangle her effort a from her performance shock η. A performance-based contract specifies an effort level and an earnings schedule as a function of realized output. Following Edmans and Gabaix (2011), we impose the following assumption in order to characterize the optimal contract analytically.

Assumption 1. *The agent chooses effort* after *observing the realization of her performance shock η. The firm recommends the same effort level $a(\theta)$ for all agents with ability θ.*

We discuss Assumption 1 in Section 1.3 below. We relax its second part and extend our analysis to arbitrary effort schedules $a(\theta, \eta)$ in Appendix C. Note that the firm

---

[2]Throughout the paper we denote a worker's earnings or income by w, while the term wage-rate stands for earnings per unit of effort w/a.

[3]This condition holds as long as the tax schedule T is not too regressive; see Appendix A for details. 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, the tax schedule which encourages such earnings randomization is Pareto inefficient.

can infer the worker's performance shock $\hat{\eta} = y/\theta - a(\theta)$ upon observing her output y, assuming that she has exerted the recommended effort level $a(\theta)$. Therefore, the earnings contract can be equivalently expressed as a function of the inferred performance shock $\hat{\eta}$ rather than the realized output y. Since recommended effort is incentive-compatible by construction, in equilibrium the firm infers the worker's true performance shock, that is, $\hat{\eta} = \eta$. Thus, throughout the paper we simply denote the earnings schedule by the map $\eta \mapsto w(\theta, \eta)$.

The firm chooses the contract $\{a(\theta), w(\theta, \cdot)\}$ that maximizes its expected profit given the tax schedule T and the worker's reservation utility $U(\theta)$, that is,[4]

$$\Pi(\theta) = \max_{a(\theta), w(\theta, \cdot)} \mathbb{E}\left[y - w(\theta, \eta)\right], \tag{2}$$

subject to the incentive-compatibility constraints:

$$a(\theta) = \arg\max_{a \geq 0} u\left(R\left(w(\theta, \eta)\right)\right) - h(a), \ \ \forall \eta, \tag{3}$$

and the participation constraint:

$$\mathbb{E}\left[u\left(R\left(w(\theta, \eta)\right)\right)\right] - h(a(\theta)) \geq U(\theta). \tag{4}$$

Since the participation constraint (4) binds at the optimum, the expected utility of workers with ability θ is equal to $U(\theta)$. The incentive-compatibility constraints (3) deserve some explanation. Since effort is chosen after the worker observes her performance shock η, it must maximize utility state-by-state rather than in expectation. Thus, equation (3) must hold for every performance shock realization η.

Labor Market Equilibrium. To close the model, we assume that there is free entry of firms in each labor market θ.[5] Thus, in equilibrium profits are equal to zero,

$$\Pi(\theta) = 0. \tag{5}$$

---

[4]Throughout the paper, the operator $\mathbb{E}$ denotes the expectation over performance shocks η, or equivalently output y, conditional on ability θ.

[5]We can easily generalize our analysis to environments where firms have market power and make positive profits. In particular, the optimal contract characterization (6, 7) holds for any reservation value $U(\theta)$, not necessarily determined by free-entry. For instance we can assume that the worker's reservation value is a convex combination of the reservation value under free entry and some exogenous outside option.

This condition pins down the workers' reservation value $U(\theta)$, which is just high enough so that no additional firm finds it profitable to enter the labor market.

1.2 Equilibrium Labor Contract

The following proposition characterizes the optimal contract between the firm and a worker with ability θ. It is an application of the results of Edmans and Gabaix (2011) to our environment with a nonlinear tax schedule.

Proposition 1. *The optimal contract $\{a(\theta), w(\theta, \cdot)\}$ and equilibrium expected utility $U(\theta)$ of agents with ability θ when $a(\theta) > 0$ are characterized by the following three equations.[6] The earnings schedule $w(\theta, \cdot)$ satisfies*

$$u(R(w(\theta, \eta))) - h(a(\theta)) = U(\theta) + h'(a(\theta))\eta, \quad \forall \eta. \tag{6}$$

The optimal effort level $a(\theta)$ satisfies

$$\mathbb{E}\left[\frac{h'(a(\theta))}{v'(w(\theta, \eta))}\right] + \mathbb{E}\left[\frac{h''(a(\theta))}{v'(w(\theta, \eta))}\eta\right] = \theta. \tag{7}$$

The equilibrium reservation utility $U(\theta)$ is determined by

$$\mathbb{E}\left[v^{-1}(U(\theta) + h(a(\theta)) + h'(a(\theta))\eta)\right] = \theta a(\theta). \tag{8}$$

Proof. See Appendix A. $\square$

Earnings Schedule. In order to motivate high-performing workers to provide as much effort $a(\theta)$ as those with lower performance shocks, the firm needs to reward them with higher earnings. Equation (6) shows that the agent's ex-post utility $u(R(w(\theta, \eta))) - h(a(\theta))$ is an affine function of the performance shock η that the firm infers. The linearity of the contract in the utility space is a consequence of our assumption of a separable utility function.[7] Since we assumed that the utility of earn-

---

[6]When the optimal effort level is zero, the worker optimally receives no compensation from the firm. Our analysis goes through without assuming $a(\theta) > 0$ if $h'(0) = 0$.

[7]Note that the earnings schedule (6) is non-trivial even if the utility of consumption $u(\cdot)$ is linear: in this case, a progressive income tax schedule implies that the utility of earnings $v(w) = R(w)$ is strictly concave, so that the worker is effectively averse to pre-tax earnings risk and values insurance against performance shocks.

ings $w \mapsto v(w) \equiv u(R(w))$ is concave, this translates into a convex earnings schedule $w(\theta, \cdot)$. Empirically, performance-pay contracts are indeed often convex, either due to nonlinear commission rates as in the case of stock and travel brokers (Levitt and Syverson (2008))[8] or to stock options (Edmans and Gabaix (2011, 2016)).

The two key features of the utility schedule (6) are its demogrant and its slope. Its demogrant in (6) is equal to $U(\theta)$: a higher reservation value leads the firm to raise the utility of workers *uniformly* regardless of their performance so as to preserve incentive-compatibility. Its slope is equal to $h'(a(\theta))$: inducing an agent with large unobservable performance shock to provide costly work effort requires a larger reward if the marginal disutility of labor is higher. Crucially, since the marginal disutility $h'(\cdot)$ is increasing, the sensitivity of utility to performance shocks is strictly increasing in labor effort $a(\theta)$. This observation captures the fundamental insight that eliciting higher effort from a worker in the presence of moral hazard requires a higher exposure to output risk.

Effort Level. Equation (7) pins down the value of effort that maximizes the firm's profit. The optimal level $a(\theta)$ is such that the expected gain in output $\theta\hat{a}$ due to a marginal increase $\hat{a} > 0$ in the workers' effort is exactly compensated by the pay raise necessary to elicit this higher effort. This cost has two components. First, to ensure that agents' participation constraint (4) remains satisfied despite their higher labor supply, their earnings must increase to compensate their utility loss $-\Delta h(a(\theta)) = -h'(a(\theta))\hat{a}$. In a frictionless economy, this would be the only effect and (7) would reduce to the familiar optimality condition $\frac{h'(a(\theta))}{v'(w(\theta,\eta))} = \theta$, according to which the marginal rate of substitution (MRS) between effort and earnings is equal to the marginal rate of transformation, or labor productivity θ.

In our setting with moral hazard, agency frictions create a wedge between labor productivity and the (expected) marginal rate of substitution, even in the absence of any distortionary taxes.[9] Providing incentives to work harder requires increasing

---

[8]While it is common for real-estate brokers to be compensated with a fixed commision rate, thus leading to a linear earnings schedule, Levitt and Syverson (2008) show that such contracts are suboptimal and could be improved by introducing convexity.

[9]The term $\frac{h''(a(\theta))}{v'(w(\theta,\eta))}$ can be rewritten as $\frac{1}{\varepsilon(\theta)a(\theta)} \frac{h'(a(\theta))}{v'(w(\theta,\eta))}$ where $\varepsilon(\theta)$ is the Frisch elasticity of labor supply. Thus, the wedge τ_{MCI} between the marginal rate of substitution and the marginal rate of transformation at performance shock realization η, defined by $(1 + \tau_{\mathrm{MCI}}) \frac{h'(a(\theta))}{v'(w(\theta,\eta))} = \theta$, is equal to $\frac{\eta}{\varepsilon(\theta)a(\theta)}$.

the sensitivity $h'(a(\theta))$ of utility to performance shocks by $\Delta h'(a(\theta)) = h''(a(\theta))\hat{a}$, and hence the slope of the earnings schedule by $\frac{h''(a(\theta))\hat{a}}{v'(w(\theta,\eta))}$. This mechanically changes the labor cost of an agent with performance shock $\eta \in \mathbb{R}$ by $\frac{h''(a(\theta))\hat{a}}{v'(w(\theta,\eta))}\eta$. This leads to the second expectation in the left-hand side of (7) which we call the *marginal cost of incentives* (MCI). In particular, eliciting a higher effort level requires raising (respectively, lowering) the earnings of high- (resp., low-) performers. Yet, since the marginal utility $v'(w) = r(w)u'(R(w))$ is decreasing the profit generated by a smaller wage bill for unlucky workers does not fully compensate the firm for the cost of raising the wages of lucky workers. As a result, the expected cost of providing incentives is positive.

Expected Utility. Finally, equation (8) is simply a rewriting of the free-entry condition (5). It implies that the average income $\mathbb{E}[w(\theta,\eta)]$ of agents with ability θ is equal to their expected output $\mathbb{E}[y] = \theta a(\theta)$. Using formula (6), this equilibrium condition pins down the workers' reservation value $U(\theta)$.

1.3   Discussion of Assumptions

To obtain the tractable characterization of the contract described in Proposition 1, our analysis relied on several key assumptions.

Utility Function. The first restriction is the separability of the utility function between consumption and effort. This assumption is not essential and is only made for clarity of exposition. As in Edmans and Gabaix (2011), it is straightforward to extend our analysis to a larger class of utility functions, namely, $\phi(u(c) - h(a))$ where ϕ exhibits non-increasing absolute risk aversion (NIARA).[10] In particular, this would allow us to nest the functional form assumed by Holmstrom and Milgrom (1987). The fact that the slope of the contract is equal to $h'(a(\theta))$, which is crucial for our main results, is robust to this more general specification. The only difference that this more general specification would make is that the distribution of a rent by the firm would no longer lead to a *uniform* shift in ex-post utilities via the demogrant $U(\theta)$. Our arguments can however be straightforwardly extended to alternative distributions of rents.

---

[10]Most common utility functions, in particular those with constant absolute risk aversion (CARA) or constant relative risk aversion (CRRA), belong to the NIARA class.

Timing. The first part of Assumption 1 imposes that the worker chooses effort a after observing the performance shock η. This timing assumption was originally introduced by Laffont and Tirole (1986) and was subsequently used by, for instance, Edmans and Gabaix (2011); Garrett and Pavan (2015). It allows us to solve the firm's problem for a very general class of utility functions. Allowing for arbitrary utility functions is crucial for our analysis. Indeed, nonlinear taxes effectively modify the concavity of the utility that workers derive from their gross earnings. If we had to restrict the utility function to a specific functional form (for instance, CARA as in Holmstrom and Milgrom (1987)) we would only be able to study tax schedules that preserve this functional form (for instance, linear or affine). Instead, the tractability allowed by our timing assumption allows us to characterize the incidence of arbitrarily nonlinear taxes.

Effort. In the main body of the paper, we impose that the firm chooses to elicit the same level of effort regardless of the worker's performance shock – this is the second part of Assumption 1. This restriction is also imposed by Edmans and Gabaix (2011) in their main model, and by Edmans, Gabaix, Sadzik, and Sannikov (2012). It is an exogenous restriction on the set of contracts that is not without loss of generality. It substantially simplifies our analysis without restricting the shape of the earnings schedule, which is crucial for our investigation of private insurance and nonlinear taxation. Carroll and Meng (2016) provide a microfoundation of this restriction; they call this property *reliability* and show that it may be optimal when firms aim to design a contract that is robust to uncertainty about the distribution of the performance shock.[11] For completeness, we relax this "constant-effort" assumption and generalize our main result (Theorem 1) to fully optimal contracts in Appendix C – our theoretical analysis remains technically straightforward and carries qualitatively over to this case.

Performance Shocks. Finally, and importantly, note that we do not impose any restriction on the distribution of performance shocks η, other than it must take values in an interval (bounded or unbounded). We view this generality as an important feature of our analysis. As an example, this allows us to capture the structure of contracts that specify of a fixed baseline income and an additional bonus paid with

---

[11]In particular, in our environment such a contract leads to the same level of expected output $\mathbb{E}y = \theta a(\theta)$ regardless of the distribution of η. However, the firm's expected profit depends on the distribution of η as earnings $w(\theta, \eta)$ are not linear in η.

positive probability by letting the distribution of η have a mass point at the lower bound $\underline{\eta}$, and a smooth density on $(\underline{\eta}, \bar{\eta})$. Importantly, it also allows us to let the distribution of η, and hence the degree of performance-pay, depend explicitly on the ability level θ.

2 General Tax Incidence Analysis

This section is devoted to the positive analysis of nonlinear tax incidence. We derive the impact of tax reforms on earnings, first in Section 2.1 in a benchmark setting with exogenous risk, then in Section 2.2 in our general environment that takes into account the endogeneity of private insurance. In Section 2.3 we derive the impact of tax reforms on individual utilities. We finally introduce the relevant notions of earnings elasticities in Section 2.4 in order to express our tax incidence formulas in terms of empirically estimable variables.

Nonlinear Tax Reforms. We start by formally defining the concept of nonlinear reforms of an arbitrary initial tax system. Consider a given (potentially suboptimal) tax schedule T, say the U.S. tax code, and another function $\hat{T} : \mathbb{R}_+ \to \mathbb{R}$. Our goal is to evaluate the effects of perturbing the initial tax schedule T by $\delta\hat{T}$, where $\delta > 0$ is a scalar that parametrizes the size of the reform in the direction $\hat{T}$. Formally, consider an outcome variable Ψ, for instance individual earnings, utility, government revenue, or social welfare, that depends on the tax schedule T. The first-order change in the value of this functional $T \mapsto \Psi(T)$ following a marginal tax reform in the direction $\hat{T}$ is given by the Gateaux derivative

$$\hat{\Psi}(T, \hat{T}) \equiv \lim_{\delta \to 0} \frac{\Psi(T + \delta\hat{T}) - \Psi(T)}{\delta}. \tag{9}$$

We analyze several concrete examples of tax reforms in Section 4 and Appendix B.

2.1 Exogenous Risk Benchmark

As a preliminary step towards our general analysis, we derive in this section the incidence of tax reforms $\hat{T}$ on earnings $w(\theta, \eta)$ and utilities $U(\theta)$ that would arise in an environment with fully exogenous risk. In this benchmark model, as in our

framework, there are two sources of heterogeneity: innate ability θ and job-specific productivity shocks η. A worker with characteristics (θ, η) is offered a fixed wage rate $x(\theta, \eta)$ that reflects her exogenous labor productivity. To make this model comparable to ours we assume moreover that the worker's labor effort $a(\theta)$ is independent of η.[12] A worker's income $w(\theta, \eta)$ is then the product of her exogenous labor productivity $x(\theta, \eta)$ and her labor supply $a(\theta)$. The key difference with our general environment is that in this benchmark setting, wage rates $\frac{w(\theta, \eta)}{a(\theta)} = x(\theta, \eta)$ are policy-invariant.

Incidence of Tax Reforms on Earnings. In such an environment, tax reforms only affect earnings $w(\theta, \eta)$ by the endogenous change in effort $\hat{a}(\theta)$ caused by the reform, multiplied by the constant wage rate $x(\theta, \eta)$. Thus, the incidence of tax reforms is given by[13]

$$\hat{w}_{\text{ex}}(\theta, \eta) \quad = \quad x(\theta, \eta)\,\hat{a}(\theta) \quad = \quad w(\theta, \eta)\,\frac{\hat{a}(\theta)}{a(\theta)}. \tag{10}$$

This is the standard behavioral response to taxes through labor supply choices analyzed in most of the optimal taxation literature following Mirrlees (1971).[14] Importantly, note that the effort change $\hat{a}(\theta)$ in formula (10) depends on the particular reform that is implemented – formally, it is the Gateaux derivative of the effort functional $a(\theta)$ in the direction $\hat{T}$. Thus, at this stage $\frac{\hat{a}(\theta)}{a(\theta)}$ is a *policy elasticity* in the sense of Hendren (2015).[15] Section 2.4 below is devoted to expressing this labor supply response in terms of standard elasticities and income effect parameters that can be estimated empirically independently of a particular choice of tax reform.

Formula (10) implies that $\hat{w}_{\text{ex}}(\theta, \eta) > 0$ iff $\hat{a}(\theta) > 0$. That is, the earnings schedule is shifted up (resp., down) if effort increases (resp., decreases) following the

---

[12]This can be justified by assuming that in this model effort is chosen before observing the realization of η.

[13]For notational simplicity, whenever there is no ambiguity we ignore the dependence of the Gateaux derivative $\hat{a}(\theta)$ on the initial tax schedule T and the tax reform $\hat{T}$.

[14]Note that we would obtain exactly the same expression in the Mirrlees model without within-group inequality, that is, $\sigma_\eta^2 = 0$. In this case, all earnings differences are due to innate ability (or labor productivity) θ and effort $a(\theta)$, so that the compensation schedule $w(\theta, \cdot)$ conditional on ability is degenerate. Equation (10) then reduces to $\hat{w}_{\text{ex}}(\theta, \eta) = \theta \hat{a}(\theta)$.

[15]This is the concept of elasticity used in several papers in the taxation literature, for instance, Chetty and Saez (2010).

reform. Earnings adjust on average by

$$\mathbb{E}\left[w\left(\theta,\eta\right)\frac{\hat{a}\left(\theta\right)}{a\left(\theta\right)}\right] \;=\; \theta\hat{a}\left(\theta\right). \tag{11}$$

Now, consider the impact of the reform on earnings *risk* around this mean adjustment. We measure earnings risk by the variance of log-earnings conditional on ability θ. Since effort $a\left(\theta\right)$ does not depend on η, equation (10) immediately implies that earnings risk after the reform is the same as before the reform, that is,

$$\mathrm{Var}\left[\log\left(w\left(\theta,\eta\right)+\delta\hat{w}_{\mathrm{ex}}\left(\theta,\eta\right)\right)\mid\theta\right] \;=\; \mathrm{Var}\left[\log\left(w\left(\theta,\eta\right)\right)\mid\theta\right] \tag{12}$$

for $\delta>0$ small enough. Therefore, in the benchmark model that ignores the endogeneity of private insurance, tax reforms affect the average level of earnings but do not modify the amount of risk to which workers are exposed.[16]

Incidence of Tax Reforms on Welfare. Finally, in the benchmark model with exogenous risk, the incidence of the tax reform on the average utility of agents with ability θ is given by

$$\hat{U}\left(\theta\right) \;=\; -\mathbb{E}\left[u'\left(R\left(w\left(\theta,\eta\right)\right)\right)\hat{T}\left(w\left(\theta,\eta\right)\right)\right]. \tag{13}$$

Intuitively, an increase in the tax payment of an agent by $\hat{T}\left(w\left(\theta,\eta\right)\right)$ lowers her ex-post utility by the marginal utility of consumption $u'\left(R\left(w\left(\theta,\eta\right)\right)\right)$. This is a simple consequence of the envelope theorem: since labor effort is chosen optimally by equation (3), the endogenous change in effort $\hat{a}\left(\theta\right)$ triggered by the reform has no first-order impact on welfare.[17] Taking expectations leads to the change in expected utility (13).

---

[16]We can also define earnings risk at a disaggregated level by the pass-through function $\frac{\partial\log w(\theta,\eta)}{\partial\eta}$, that is, the sensitivity of log-earnings to performance shocks. Equation (10) implies that the pass-through is unaffected by the reform for every value of η. If we define instead earnings risk as the sensitivity of earnings (rather than log-earnings) with respect to performance shocks, that is, $\frac{\partial w(\theta,\eta)}{\partial\eta}$, tax reforms would raise earnings risk if and only if $\frac{\partial\hat{w}(\theta,\eta)}{\partial\eta}>0$. In the setting analyzed in this section, formula (10) implies that this is the case whenever the reform raises effort, that is, $\hat{a}\left(\theta\right)>0$.

[17]In particular, in this environment a change in marginal tax rates that keeps the total tax payment unchanged has no first-order impact on individual welfare.

2.2 Incidence of Tax Reforms on Earnings

We now proceed to characterizing the incidence of an arbitrary tax reform $\hat{T}$ on the compensation schedule $w(\theta, \cdot)$ of workers with ability θ in our general environment. This is the first main result of our paper.

Theorem 1. *Suppose that $a(\theta) > 0$. Denote by $\hat{a}(\theta)$ the change in effort induced by the reform, which we study in Section 2.4 below. The first-order effect of the tax reform $\hat{T}$ on earnings $w(\theta, \cdot)$ is given by*

$$\hat{w}(\theta, \eta) = \hat{w}_{\mathrm{ex}}(\theta, \eta) + \hat{w}_{\mathrm{co}}(\theta, \eta) + \hat{w}_{\mathrm{pp}}(\theta, \eta), \tag{14}$$

where the crowding-out effect $\hat{w}_{\mathrm{co}}$ has mean zero and is given by

$$\hat{w}_{\mathrm{co}}(\theta, \eta) = \frac{\hat{T}(w(\theta, \eta))}{r(w(\theta, \eta))} - \frac{(v'(w(\theta, \eta)))^{-1}}{\mathbb{E}\left[(v'(w(\theta, \cdot)))^{-1}\right]} \mathbb{E}\left[\frac{\hat{T}(w(\theta, \cdot))}{r(w(\theta, \cdot))}\right], \tag{15}$$

and the performance-pay effect $\hat{w}_{\mathrm{pp}}$ has mean zero and is given by

$$\hat{w}_{\mathrm{pp}}(\theta, \eta) = \left[\frac{h'(a(\theta)) + h''(a(\theta))\eta}{v'(w(\theta, \eta))} - \frac{w(\theta, \eta)}{a(\theta)}\right] \hat{a}(\theta). \tag{16}$$

Proof. See Appendix B. □

Equation (14) gives the adjustment of the earnings schedule following an arbitrary tax reform $\hat{T}$ as a function of the tax rates, earnings distribution, and labor supply responses in the *initial* (pre-reform) economy. In practice, this formula only requires choosing a functional form for the utility function u and the disutility of effort h in order to evaluate the incidence of any potential reform of the current tax code.

Theorem 1 shows that the earnings adjustment in response to the tax reform, $\hat{w}(\theta, \eta)$, is in general different than in the standard model with exogenous risk, $\hat{w}_{\mathrm{ex}}(\theta, \eta)$, analyzed in Section 2.1. Specifically, the tax reform modifies the earnings schedule by the same *average* amount as in the benchmark model (see equation (11)), since $\mathbb{E}[\hat{w}_{\mathrm{co}}(\theta, \eta)] = \mathbb{E}[\hat{w}_{\mathrm{pp}}(\theta, \eta)] = 0$. However, it also introduces two adjustments to earnings *risk* around this mean shift. The first, $\hat{w}_{\mathrm{co}}(\theta, \eta)$, captures the *crowding-out* of the private insurance contract by the tax change, keeping effort constant. The second, $\hat{w}_{\mathrm{pp}}(\theta, \eta)$, is the *performance-pay* effect due to the endogenous change in labor effort. We analyze them in turn.

16

Crowding-Out of Private Insurance. Equation (15) gives the adjustment to the compensation schedule that the firm must implement in order to keep the worker's incentive and participation constraints both satisfied following the reform. First, consider the adjustment $\frac{\hat{T}(w(\theta,\eta))}{r(w(\theta,\eta))}$ of the earnings schedule (first term in (15)). This term implies that the agent's consumption $c(\theta,\eta) = w(\theta,\eta) - T(w(\theta,\eta))$ changes by

$$
\begin{aligned}
\hat{c}(\theta,\eta) &= -\hat{T}(w(\theta,\eta)) + (1 - T'(w(\theta,\eta)))\,\hat{w}(\theta,\eta) \\
&= -\hat{T}(w(\theta,\eta)) + (1 - T'(w(\theta,\eta)))\frac{\hat{T}(w(\theta,\eta))}{1 - T'(w(\theta,\eta))} = 0.
\end{aligned}
$$

Thus, absent any other forces – in particular, if effort were kept constant – the firm would adjust the contract such that, for every performance shock realization η, the agent's disposable income $c(\theta,\eta)$, and hence her realized utility, remain fixed. In other words, any attempt by the government to affect consumption insurance would be fully absorbed by the firm so as to keep the worker's payoffs unchanged.

Second, suppose that the tax reform is such that the tax liabilities of workers with ability θ are reduced, that is, $\hat{T}(w(\theta,\eta)) < 0$ for all η. Per our discussion in the previous paragraph, this reform generates a rent for the firm equal to $-\mathbb{E}[\frac{\hat{T}(w(\theta,\cdot))}{r(w(\theta,\cdot))}] > 0$: intuitively, the firm compensates the reduction in tax payments by an equivalent reduction in wages. Now, by the free-entry condition, this rent must be shared with workers, whose expected utility rises as a result.[18] Recall that, by equation (6), this increase in utility must be distributed uniformly among all agents – regardless of their performance η – in order to preserve their incentive compatibility condition for effort. But this implies that the salary of high-performers must increase by a larger amount, since their marginal utility of earnings $v'(w(\theta,\eta))$ is lower. As a result, the share of the rent assigned to workers with performance shock η is inversely proportional to their marginal utility, that is, equal to $\frac{(v'(w(\theta,\eta)))^{-1}}{\mathbb{E}[(v'(w(\theta,\cdot)))^{-1}]}$. This leads to the second term in (15).

Performance-Pay Effect. Now, suppose that in response to the tax reform $\hat{T}$, the firm finds it optimal to elicit a higher effort level, so that $\hat{a}(\theta) > 0$. In order to do so, we showed in Section 1.2 that it must both compensate workers for their utility loss, and increase the pass-through of performance shocks to earnings. These two adjustments are captured by the term $\frac{h'(a(\theta)) + h''(a(\theta))\eta}{v'(w(\theta,\eta))}$ in equation (16). Since

---

[18] We analyze this change in reservation value $\hat{U}(\theta)$ in Section 2.3 below.

$\frac{h'(a(\theta))+h''(a(\theta))\eta}{v'(w(\theta,\eta))}\hat{a}(\theta)$ is an increasing function of η, eliciting a higher effort level requires increasing the sensitivity of earnings to performance. Now, $\hat{w}_{\mathrm{pp}}$ is defined by subtracting the income change $\hat{w}_{\mathrm{ex}}(\theta,\eta)$ that would arise in the benchmark model with exogenous risk in response to the same change in effort (equation (10)). As a result, the performance-pay effect has mean zero and is the pure contribution of moral hazard to the change in earnings risk via labor supply decisions.

Generalization to an Effort Schedule. In Appendix C we extend this result to the environment where the firm can elicit an arbitrary (non-constant) effort schedule $a(\theta,\eta)$. We show that the crowding-out effect is identical to (15). The performance-pay effect is analogous to (16) except that it depends on the change in the entire effort schedule rather than in the single effort level.

2.3 Incidence of Tax Reforms on Utilities

We now proceed to analyzing the incidence of tax reforms on the expected utility $U(\theta)$ of workers with ability θ in our general environment.

Proposition 2. *The first-order effect of the tax reform $\hat{T}$ on expected utility $\hat{U}(\theta)$ is given by*

$$\hat{U}(\theta) = -\frac{1}{\mathbb{E}\left[\frac{1}{v'(w(\theta,\eta))}\right]}\mathbb{E}\left[\frac{\hat{T}(w(\theta,\eta))}{r(w(\theta,\eta))}\right]. \qquad (17)$$

Proof. See Appendix B. □

To understand Proposition 2, recall that a decrease in the tax payment of an agent by $\hat{T}(w(\theta,\eta)) < 0$ allows the firm to decrease her earnings by $\frac{\hat{T}(w(\theta,\eta))}{r(w(\theta,\eta))}$ in order to keep her consumption (and, hence, incentives) unchanged. Moreover, by the envelope theorem, any change in pay that operates via labor supply $(\hat{w}_{\mathrm{ex}},\hat{w}_{\mathrm{pp}})$ generates only second-order changes in total labor costs. Therefore, the tax reform creates an expected rent for the firm equal to $-\mathbb{E}[\frac{\hat{T}(w(\theta,\eta))}{r(w(\theta,\eta))}] > 0$, which is then shared with the workers and leads to an increase in their reservation value by $\hat{U}(\theta) > 0$.

Now, this increase in expected utility $\hat{U}(\theta) > 0$ must be distributed across workers with different performance shocks η. As explained in the previous sections, every worker's utility must increase uniformly. Therefore, realized earnings $w(\theta,\eta)$ must

increase in proportion to the inverse marginal utility $1/v'(w(\theta, \eta))$. Hence, this sharing rule costs the firm $\mathbb{E}[\frac{\hat{U}(\theta)}{v'(w(\theta,\eta))}]$. As a result, the value of $\hat{U}(\theta)$ which ensures that profits remain equal to zero (free-entry condition) satisfies:

$$\mathbb{E}\left[\frac{\hat{T}(w(\theta, \eta))}{r(w(\theta, \eta))}\right] + \mathbb{E}\left[\frac{\hat{U}(\theta)}{v'(w(\theta, \eta))}\right] = 0.$$

Solving for $\hat{U}(\theta)$ easily leads to equation (14).

Analogous to the standard model where tax changes affect individual consumption directly rather than being intermediated by firms, equation (17) implies that workers' expected utility increases when their expected tax payments (weighted by retention rates) are reduced. Conversely, an increase in their expected tax bill lowers their utility. However, the level of change in utility $\hat{U}(\theta)$ differs from that obtained in the benchmark model with exogenous risk (equation (13)) unless $\sigma_\eta^2 = 0$. As a simple example, suppose that the initial tax schedule is affine, so that the retention rate $r(w(\theta, \eta))$ is constant. Consider a tax reform that consists of a uniform lump-sum transfer for all agents. We show in the Appendix that this reform is represented by $\hat{T}(w) = -1$ for all w. In the benchmark model with exogenous risk, equation (13) shows that individual welfare would increase on average by the expected marginal utility, $\mathbb{E}[u'(R(w(\theta, \eta)))]$. Now, in the general model with agency frictions, applying Jensen's inequality to equation (17) yields $0 < \hat{U}(\theta) < \mathbb{E}[u'(R(w(\theta, \eta)))]$. Therefore, a lump-sum transfer leads to a strictly smaller rise in utility when tax cuts are distributed by firms than when they are directly targeted to workers.

2.4 Elasticities of Average Earnings

The last step of our tax incidence analysis is to characterize the impact of a tax reform $\hat{T}$ on the optimal effort level $a(\theta)$. We tackle this in two (complementary) ways. First, we use a structural approach and derive analytically the impact of tax reforms on labor effort in terms of primitives. Second, we express these labor supply responses in terms of sufficient statistics that can be estimated empirically regardless of the values of the underlying primitives.

Structural Approach. When the structure of the model is simple enough, it is worthwhile to derive explicitly the elasticity of effort with respect to the particular

tax reform under consideration, that is, $\frac{\hat{a}(\theta)}{a(\theta)}$. This allows us to compare the incidence of taxes across different contractual environments. The next result and Lemma 2 below illustrate this approach through the lens of simple examples.

Lemma 1. *Suppose that the utility function is linear in consumption, that is $u(c) = c$, and that earnings $w(\theta, \eta)$ are located in a bracket with constant marginal tax rate τ for all performance shocks η. The response of labor effort to an arbitrary tax reform $\hat{T}$ is then given by*

$$\frac{\hat{a}(\theta)}{a(\theta)} = -\frac{1}{1-\tau}\varepsilon(\theta)\,\mathbb{E}\left[\hat{T}'(w(\theta, \eta))\right] - \frac{1}{1-\tau}\mathrm{Cov}\left(\hat{T}'(w(\theta, \eta)); \frac{\eta}{a(\theta)}\right), \quad (18)$$

where $\varepsilon(\theta) \equiv \frac{h'(a(\theta))}{a(\theta)h''(a(\theta))}$ is the Frisch elasticity of labor supply.

Proof. See Appendix B. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Equation (18) shows that the response of labor effort to the tax reform $\hat{T}$ is the sum of two terms, which reflect both elements (MRS and MCI) of the first-order condition (7). First, the marginal rate of substitution (first expectation in (7)) implies that the change in expected marginal tax rates, $\mathbb{E}[\hat{T}'(w(\theta, \eta))]$, reduces labor supply by the Frisch elasticity $\varepsilon(\theta)$. This is the standard response one would obtain in models with exogenous risk. Second, recall that in our moral hazard environment the optimal effort level $a(\theta)$ is also determined by the marginal cost of incentives (second expectation in (7)). But the MCI is positively related to the progressivity of the tax schedule: with quasilinear utility we have $\mathrm{MCI} \propto \mathrm{Cov}(\frac{1}{r(w(\theta, \eta))}; \eta)$, which is equal to zero (respectively, positive) when the marginal tax rates are constant (resp., increasing with income). Consequently, starting from an affine tax code, we expect a progressive tax reform – for which the marginal tax rate adjustments $\hat{T}'(\cdot)$ increase with income – to raise the cost of incentive provision, and hence trigger an additional downward adjustment in effort. Formally, this is indeed implied by the negative covariance term in equation (18). Therefore, taking into account the endogeneity of private insurance against output risk magnifies the negative impact of raising tax progressivity on labor effort. In Section 4, we generalize this result to the case of a utility function with income effects and a nonlinear baseline tax schedule and show that the same insight carries over.

Sufficient-Statistic Approach. In our most general environment, the analytical expressions for the policy elasticities $\frac{\hat{a}(\theta)}{a(\theta)}$ are technically straightforward to derive, but they may fail to deliver sharp comparative statics with respect to the variance of performance shocks σ_η^2 or the strength of moral hazard frictions. Instead, it is standard since Saez (2001) to express these labor supply responses in terms of substitution and income effects that can be estimated in the data, and treat the resulting elasticities as *sufficient statistics* in our tax incidence analysis (Chetty (2009)). Namely, our tax formulas depend on the empirical values of these parameters, regardless of the underlying structure of the model that generates them – that is, in our case, regardless of whether private insurance against performance shocks is exogenous or endogenous. Our goal is therefore to express the labor supply response $\frac{\hat{a}(\theta)}{a(\theta)}$ to any potential tax reform $\hat{T}$ in terms of standard elasticity parameters that can be estimated independently of the particular reform.

To do so, recall that, by the free-entry condition (5), average earnings conditional on ability θ, $\mathbb{E}[w(\theta, \cdot)]$, are equal to $\theta a(\theta)$. As a consequence, the elasticities of effort $a(\theta)$ with respect to tax changes are equal to the corresponding elasticities of average earnings $\mathbb{E}[w(\theta, \cdot)]$. Note that to evaluate average earnings $\mathbb{E}[w(\theta, \cdot)]$ the econometrician does not need to observe the actual value of ability θ – it is enough to group workers into ordinal ability groups proxied by education, experience, etc. We can thus define the (compensated) elasticity of average earnings of agents with ability θ with respect to the retention rate at income level $w(\theta, \eta)$ by

$$\varepsilon_{\mathbb{E}w,r}(\theta, \eta) \equiv \frac{r(w(\theta, \eta))}{\mathbb{E}[w(\theta, \cdot)]} \frac{\partial \mathbb{E}[w(\theta, \cdot)]}{\partial r(w(\theta, \eta))}. \tag{19}$$

We also define the income effect parameter as the semi-elasticity of average earnings of agents with ability θ with respect to a lump-sum transfer at income level $w(\theta, \eta)$, that is,

$$\varepsilon_{\mathbb{E}w,R}(\theta, \eta) \equiv \frac{1}{\mathbb{E}[w(\theta, \cdot)]} \frac{\partial \mathbb{E}[w(\theta, \cdot)]}{\partial R(w(\theta, \eta))}\bigg|_U. \tag{20}$$

These elasticities can be estimated empirically, and their explicit analytical expressions in terms of primitives are given in Appendix B.

Proposition 3. *The first-order effect of the tax reform $\hat{T}$ on labor effort $a(\theta)$ can be*

21

expressed as

$$\frac{\hat{a}\left(\theta\right)}{a\left(\theta\right)} \;\; = \;\; -\mathbb{E}\left[\varepsilon_{\mathbb{E}w,r}\left(\theta,\eta\right)\frac{\hat{T}'\left(w\left(\theta,\eta\right)\right)}{r\left(w\left(\theta,\eta\right)\right)}\right] + \mathbb{E}\left[\varepsilon_{\mathbb{E}w,R}\left(\theta,\eta\right)\frac{\hat{T}\left(w\left(\theta,\eta\right)\right)}{w\left(\theta,\eta\right)}\right] \quad (21)$$

where $\varepsilon_{\mathbb{E}w,r}\left(\theta,\eta\right)$ and $\varepsilon_{\mathbb{E}w,R}\left(\theta,\eta\right)$ are defined in (19) and (20), respectively.

Proof. See Appendix B. $\hspace{1cm}\square$

The interpretation of Lemma 3 is standard. An increase in the marginal tax rate by $\hat{T}'\left(w\left(\theta,\eta\right)\right)$ (resp., an increase in the average tax rate by $\frac{\hat{T}(w(\theta,\eta))}{w(\theta,\eta)}$) at the income level $w\left(\theta,\eta\right)$ affects the optimal effort level $a\left(\theta\right)$ in proportion to the compensated elasticity $\varepsilon_{\mathbb{E}w,r}\left(\theta,\eta\right)$ (resp., the income effect parameter $\varepsilon_{\mathbb{E}w,R}\left(\theta,\eta\right)$). The intuition underlying these substitution and income effects is the same as in the standard model of nonlinear income taxation. Namely, an increase in marginal tax rates (respectively, in lump-sum liabilities) lowers (resp., raises) the worker's optimal effort level by creating a wedge between the marginal rate of substitution and the marginal rate of transformation in the optimality condition (7). The only difference is that in our framework, it is the firm rather than the worker that chooses how much effort should optimally be provided, and it achieves this by spreading or compressing the earnings schedule. Nevertheless, standard methods of estimating taxable income elasticities would give the correct values for the parameters (19) and (20).

3 Aggregate Effects of Tax Reforms

In this section, we use our tax incidence results of Section 2 to characterize the aggregate costs and benefits of tax reforms. We introduce the government and define formally the concepts of excess burden and welfare gains of policies in Section 3.1. We derive the theoretical results in Section 3.2. Readers primarily interested in the incidence of tax reforms on individual performance-pay contracts can skip to Section 4.

3.1 Government

In our model, the government observes both between- and within-group inequality, that is, earnings differences due to ex-ante ability θ (proxied by education, experience,

etc.) and ex-post job-specific shocks η. However, taxes and transfers can only be conditioned on realized earnings w and not on ability θ. The labor income tax schedule is function $T \in \mathcal{C}^2(\mathbb{R}_+, \mathbb{R})$.

Government Revenue and Social Welfare. Given the tax schedule T, government revenue is given by

$$\mathcal{R}(T) = \int_\Theta \mathbb{E}\left[T\left(w\left(\theta, \eta\right)\right)\right] \mathrm{d}F\left(\theta\right). \tag{22}$$

Throughout the paper, we assume that the government faces an exogenous expenditure requirement $G \geq 0$. Any extra revenue is used for redistribution between workers with different (uninsurable) levels of ability θ. Social welfare is evaluated by a weighted-utilitarian functional

$$\mathcal{W}(T) = \int_\Theta \alpha\left(\theta\right) U\left(\theta\right) \mathrm{d}F\left(\theta\right), \tag{23}$$

where the map of Pareto weights $\theta \mapsto \alpha\left(\theta\right)$ is positive, decreasing, and satisfies $\int_\Theta \alpha\left(\theta\right) \mathrm{d}F\left(\theta\right) = 1$.

Mechanical Effect of Tax Reforms. Consider a tax reform $\hat{T}$ of the initial tax schedule T. The mechanical, or statutory, effect of this reform is equal to its impact on government revenue assuming that everyone's earnings remain fixed. It is given by

$$\mathcal{ME}(T, \hat{T}) = \int_\Theta \mathbb{E}[\hat{T}\left(w\left(\theta, \eta\right)\right)] \mathrm{d}F\left(\theta\right). \tag{24}$$

That is, in the absence of endogenous earnings responses, government revenue would simply change by the sum of (positive or negative) additional tax payments $\hat{T}\left(w\left(\theta, \eta\right)\right)$ of all agents.

Excess Burden of Tax Reforms. The excess burden, or deadweight loss, of a tax reform $\hat{T}$ is (minus) the change in government revenue caused by the endogenous earnings adjustments. Since the government retains a share $T'\left(w\left(\theta, \eta\right)\right)$ of the

workers' earnings gains or losses $\hat{w}(\theta, \eta)$, the excess burden is given by

$$\mathcal{EB}(T, \hat{T}) = -\int_{\Theta} \mathbb{E}[T'(w(\theta, \eta))\,\hat{w}(\theta, \eta)]\mathrm{d}F(\theta), \qquad (25)$$

where $\hat{w}(\theta, \eta)$ is given by (14). For instance, if a tax reform mechanically raises \$1 of revenue absent earnings adjustments, but causes distortions – say, reductions in labor supply – which lower government revenue by ¢20, then the marginal excess burden is equal to a fraction 20% of the mechanical effect. The total impact of the tax reform on government budget (that is, the Gateaux derivative of the tax revenue functional $\mathcal{R}(T)$) is therefore equal to $\hat{\mathcal{R}}(T, \hat{T}) = \mathcal{ME}(T, \hat{T}) - \mathcal{EB}(T, \hat{T})$.

Welfare Gains of Tax Reforms.   The welfare gains of a tax reform $\hat{T}$ is the change in social welfare $\mathcal{W}(T)$ that it causes, expressed in monetary units. The change in social welfare is equal to $\hat{\mathcal{W}}(T, \hat{T}) = \int \alpha(\theta)\,\hat{U}(\theta)\,\mathrm{d}F(\theta)$, where $\hat{U}(\theta)$ is the change in expected utility incurred by agents with ability θ and $\alpha(\theta)$ measures their weight in the social objective. To convert this welfare measure into units of revenue, consider another, benchmark reform $\hat{T}^*$ within the available set of tax instruments, that costs one dollar of revenue.[19] Let the marginal value of public funds λ be the increase in social welfare brought about by this reform $\hat{T}^*$. We then define the welfare gain of the tax reform $\hat{T}$ by

$$\mathcal{WG}(T, \hat{T}) = \frac{1}{\lambda}\int_{\Theta} \alpha(\theta)\,\hat{U}(\theta)\,\mathrm{d}F(\theta). \qquad (26)$$

Optimum Tax Schedule.   The optimal tax schedule is such that no tax reform of the initial tax schedule that keeps the government budget constraint satisfied has a positive first-order impact on social welfare. We show in Appendix D that the optimal tax schedule (respectively, the optimum within a restricted class of tax instruments)

---

[19]If universal lump-sum taxes and transfers are available, as in Mirrlees (1971), we naturally choose $\hat{T}^*$ to be a uniform lump-sum transfer. We show in Appendix B that a lump-sum transfer of \$1 per worker is represented by the constant function -1 for all w. Denote by $\hat{\mathcal{R}}(T, -1) < 0$ the loss in government revenue from this transfer, once all behavioral responses have been taken into account. Then the reform $\hat{T}^*(w) = -1/\left|\hat{\mathcal{R}}(T, -1)\right|$ for all w is a uniform lump-sum transfer that reduces government budget by \$1 by construction. If instead the policy is restricted to the CRP tax schedules as in Section 4, the benchmark reform $\hat{T}^*$ consists of a decrease in the parameter τ, normalized analogously to yield \$1 of revenue. See Appendix D for details.

is characterized by

$$\mathcal{EB}(T,\hat{T}) \;=\; \mathcal{ME}(T,\hat{T}) + \mathcal{WG}(T,\hat{T}), \tag{27}$$

for all tax reforms $\hat{T}$ (resp., all tax reforms in the restricted class). In other words, the marginal cost and marginal benefit of any reform must be equal at the optimum.

3.2 Excess Burden and Welfare Gains

Substituting expression (14) for $\hat{w}(\theta,\eta)$ into equation (25) yields the following characterization. This is the second main result of our paper.

Theorem 2. *The excess burden of the tax reform $\hat{T}$ is given by*

$$\mathcal{EB}(T,\hat{T}) \;=\; -\int_{\Theta} \mathbb{E}\left[T'\left(w\left(\theta,\eta\right)\right)\hat{w}_{\mathrm{ex}}\left(\theta,\eta\right)\right]\mathrm{d}F\left(\theta\right) \tag{28}$$
$$- \sum_{i\in\{\mathrm{co,pp}\}}\int_{\Theta}\mathrm{Cov}\left(T'\left(w\left(\theta,\eta\right)\right),\hat{w}_i\left(\theta,\eta\right)\right)\mathrm{d}F\left(\theta\right)$$

The welfare gains of the tax reform $\hat{T}$ are given by

$$\mathcal{WG}(T,\hat{T}) \;=\; -\frac{1}{\lambda}\int_{\Theta}\mathbb{E}\left[\tilde{\alpha}\left(\theta;\eta\right)u'\left(R\left(w\left(\theta,\eta\right)\right)\right)\hat{T}\left(w\left(\theta,\eta\right)\right)\right]\mathrm{d}F\left(\theta\right), \tag{29}$$

where the modified social welfare weights are given by $\tilde{\alpha}\left(\theta;\eta\right) \equiv \frac{\left(v'(w(\theta,\eta))\right)^{-1}}{\mathbb{E}\left[\left(v'(w(\theta,\cdot))\right)^{-1}\right]}\alpha\left(\theta\right).$

Proof. See Appendix D. $\qquad\square$

Formulas (28) and (29) can be used in practice to evaluate whether a concrete tax reform proposal has a positive or negative effect on government revenue and social welfare, starting from any (not necessarily optimal) tax code.

Excess Burden of Tax Reforms. The first integral in (28) is equal to the deadweight loss one would obtain in standard models with exogenous risk – recall that in this case, the tax reform affects earnings via the standard labor supply channel by $\hat{w}_{\mathrm{ex}}\left(\theta,\eta\right) = w\left(\theta,\eta\right)\frac{\hat{a}(\theta)}{a(\theta)}$. This deadweight loss depends on the average earnings elasticities summarized in $\frac{\hat{a}(\theta)}{a(\theta)}$, as described in Lemma 3. This integral is analogous to those typically derived in the optimal taxation literature (see for instance Saez (2001)).

Now consider the case where the endogeneity of private insurance is taken into account. Recall that the full adjustment to earnings $\hat{w}(\theta, \eta)$ has the same mean as in the frictionless benchmark. Thus, any fiscal externalities due to moral hazard must come from the change in earnings *risk* due to the crowding-out and the performance-pay effects $\hat{w}_{\text{co}}(\theta, \eta)$, $\hat{w}_{\text{pp}}(\theta, \eta)$. The first implication of Corollary 2 is that if the marginal tax rates $T'(w(\theta, \eta))$ are initially constant – as in Chetty and Saez (2010) – both covariances in the second line of equation (28) are equal to zero. Therefore, the excess burden of the tax reform is the same as in the standard model, despite the presence of moral hazard frictions and endogenous risk. In other words, the performance-based nature of contracts and the endogenous crowding-out of private insurance do not give rise to additional fiscal externalities when the tax code T is initially affine, even if the tax reform $\hat{T}$ itself is highly nonlinear.[20]

Consider finally the case where the tax schedule T is initially nonlinear. In this case, the covariances in (28) are no longer equal to zero and capture the impact on government budget of the novel sources of earnings risk highlighted in formula (14). Suppose for concreteness that the tax schedule is initially progressive, that is, the marginal tax rates $T'(\cdot)$ are increasing. In this case, $\text{Cov}(T'(w(\theta, \eta)), \hat{w}_i(\theta, \eta)) > 0$ whenever $\frac{\partial \hat{w}_i(\theta, \eta)}{\partial \eta} > 0$, that is, whenever the sensitivity of pre-tax earnings to performance shocks increases following the reform. Therefore, a spread (resp., contraction) of the earnings distribution causes a positive (resp., negative) fiscal externality, that is, a first-order gain (resp., loss) in government revenue. This is a consequence of Jensen's inequality: a progressive (concave) tax code generates more tax revenue for the government if earnings are more volatile, keeping their mean constant. For budget purposes, the government is therefore tempted to induce an increase in the dispersion of pre-tax earnings.

Welfare Gains of Tax Reforms. In the benchmark model with exogenous risk, the increase in social welfare achieved by giving one additional unit of consumption (say, via a tax break) to agents with ability θ and performance shock η is given by the marginal social welfare weight $\alpha(\theta) u'(R(w(\theta, \eta)))$, equal to their marginal

---

[20]In particular, consider the highest-income earners for whom performance-pay contracts are particularly prevalent, and suppose that their baseline income absent any bonus − that is, their earnings given the lowest realization $\underline{\eta}$ of the performance shock − is located in the highest tax bracket. Then the performance-sensitivity of their salary is irrelevant for their contribution to government revenue.

utility of consumption times their weight $\alpha(\theta)$ in the social objective. Now, in the environment with moral hazard frictions, the crowding-out of private insurance by tax policy highlighted in Theorem 1 has welfare consequences that must be taken into account.[21]

Indeed, we saw in Section 2.2 that the tax break raises the expected utility of the workers (Proposition 2), as in a standard model with exogenous risk. Crucially, however, recall that this utility gain must be shared uniformly across agents in order to preserve their effort incentives. But since the marginal utility is decreasing, this implies that workers with a higher output realization y end up getting a higher increase in consumption. As a result, expression (29) implies that the marginal social welfare weights that would arise in the benchmark model are now weighted by the share $\frac{(v'(w(\theta,\eta)))^{-1}}{\mathbb{E}\left[(v'(w(\theta,\cdot)))^{-1}\right]}$ of the tax cut that workers actually receive. These weights are regressive – richer agents end up with higher effective welfare weights in the social objective. Intuitively, tax cuts accrue mostly to the highest-performing agents of a given ability group. They are thus less efficiently targeted than in the standard model without firm intermediation, in which the government could directly alter workers' consumption. This regressive distribution of rents in turn reduces the welfare benefits of providing social insurance compared to the exogenous-risk environment.

4   The Loglinear Framework

In this section we introduce a special case of our general model that allows us to derive sharp consequences of our results of Sections 2 and 3. In particular, under the following functional form restrictions the equilibrium labor contract is loglinear and our tax incidence formulas become particularly transparent.

Assumption 2. *The utility of consumption is logarithmic, $u(c) = \log c$. The Frisch elasticity of labor supply $\varepsilon > 0$ is constant, that is $h(a) = (1 + \frac{1}{\varepsilon})^{-1} a^{1+\frac{1}{\varepsilon}}$. The performance shocks are normally distributed, $\eta \sim \mathcal{N}(0, \sigma_\eta^2)$. The tax schedule has a constant rate of progressivity (CRP),[22] that is, there exist $\tau \in \mathbb{R}$ and $p < 1$ such that*

---

[21]Because of the envelope theorem, the performance-pay effect (equation (16)) induces only second-order welfare gains or losses. The first part of crowding-out (first term in equation (15)) also keeps welfare constant since it ensures that workers' consumption remains fixed.

[22]The CRP tax code is a good approximation of the U.S. tax system, see for instance Heathcote, Storesletten, and Violante (2017). The rate of progressivity p is equal to (minus) the elasticity of the retention rate $1 - T'(w)$ with respect to income w. Alternatively, $1 - p$ is equal to the ratio of

$T(w) = w - \frac{1-\tau}{1-p} w^{1-p}.$

We characterize the equilibrium labor contract in Section 4.1. We then focus on a particular tax reform, namely, an increase in the (constant) rate of progressivity of the initial tax schedule. We derive the incidence of this reform on earnings and utilities in Section 4.2, and its excess burden and welfare gains in Section 4.3. We conclude in Section 4.4 by characterizing the optimal rate of progressivity in this economy. In Appendix G and H, we generalize our analysis of the optimal rate of progressivity to the dynamic environment of Edmans, Gabaix, Sadzik, and Sannikov (2012).

4.1 Equilibrium Labor Contract

Under these assumptions, the labor contract characterized in Proposition 1 can be simplified as follows.

Corollary 1. *Suppose that Assumption 2 holds. Denote by* $\psi \equiv \frac{\partial \log w(\theta,\eta)}{\partial \eta}$ *the pass-through of performance shocks to log-earnings. The earnings schedule is log-linear and given by*

$$\log w(\theta, \eta) = \log(\theta a) + \psi \eta - \frac{1}{2}\psi^2 \sigma_\eta^2 \quad with \quad \psi = \frac{a^{1/\varepsilon}}{1-p}. \tag{30}$$

Effort a is independent of θ *and satisfies*

$$a = \left[(1-p)\left(1 - \varepsilon_{\psi,a}\psi^2\sigma_\eta^2\right)\right]^{\frac{\varepsilon}{1+\varepsilon}}, \tag{31}$$

where $\varepsilon_{\psi,a} \equiv \frac{\partial \log \psi}{\partial \log a} = \frac{1}{\varepsilon}$. *Expected utility is given by*

$$U(\theta) = \log(R(\theta a)) - h(a) - \frac{1}{2}(1-p)\psi^2\sigma_\eta^2. \tag{32}$$

Proof. See Appendix E. □

In the standard Mirrlees (1971) model, the worker's wage rate is equal to her marginal productivity θ, and her earnings are θa. In our more general model, equation (30) implies that the firm designs an incentive-based compensation contract that has mean θa, but is also dispersed around this mean. The amount of risk to which the

marginal retained income $1 - T'(w)$ to average retained income $1 - T(w)/w$.

firm exposes the worker is summarized by the constant (that is, independent of η) *pass-through* ψ of performance shocks to log-earnings.

Crucially, this parameter $\psi \equiv \frac{\partial \log w(\theta,\eta)}{\partial \eta} = \frac{a^{1/\varepsilon}}{1-p}$ is endogenous to policy: it depends on the rate of progressivity p of the tax schedule both directly and indirectly through the optimal effort level a. If the elasticities

$$\varepsilon_{\psi,a} \;=\; \frac{\partial \log \psi}{\partial \log a}, \quad \text{and} \quad \varepsilon_{\psi,1-p} \;=\; \frac{\partial \log \psi}{\partial \log (1-p)} \tag{33}$$

were both equal to zero, the model would be equivalent to one with an exogenous and uninsurable shock η analogous to θ.

In general, however, the elasticity $\varepsilon_{\psi,a}$ is positive and measures the strength of the moral hazard friction: it determines how much more exposure to performance shocks is necessary to elicit a higher level of effort from the agent. Since $\varepsilon_{\psi,a} = \frac{1}{\varepsilon}$, our model implies that the sensitivity of earnings risk to the desired effort level is inversely proportional to the Frisch elasticity of labor supply. If in response to a tax reform the firm wants to reduce the effort provided by the worker, it implements it by reducing her exposure to risk, that is, by providing more insurance against performance shocks. In the sequel we refer to $\varepsilon_{\psi,a}$ as the *performance-pay elasticity*.

Second, the elasticity $\varepsilon_{\psi,1-p} = -1 < 0$ implies that higher tax progressivity leads to a steeper pre-tax earnings schedule. This means that ceteris paribus (that is, keeping effort constant), public insurance crowds out private insurance against output risk. Intuitively, this is because an increase in tax progressivity compresses the disposable income distribution and thus reduces the amount of risk that workers are effectively facing; as a response, the firm spreads out the pre-tax earnings schedule in order to preserve incentives for effort. In the sequel we refer to $\varepsilon_{\psi,1-p}$ as the *crowding-out elasticity*.

Finally, equations (31) and (32) imply that the worker's effort and expected utility are both strictly lower in the environment with moral hazard and endogenous private insurance, where $\varepsilon_{\psi,a} > 0$ and $\varepsilon_{\psi,1-p} < 0$, than in the exogenous-risk model where $\varepsilon_{\psi,a} = \varepsilon_{\psi,1-p} = 0$. Effort is a decreasing function of the rate of tax progressivity p.

4.2 Tax Incidence Analysis

Throughout this section we consider a tax reform that marginally raises the rate of progressivity p. This policy is represented by (see Appendix E for technical details)

$$\hat{T}(w) = \left(\log w - \frac{1}{1-p}\right) \frac{1-\tau}{1-p} w^{1-p}, \ \ \forall w > 0. \tag{34}$$

To derive the incidence of this tax reform, we can either directly differentiate with respect to p the equilibrium labor contract given by Corollary 1, or apply Theorem 1 to the corresponding function (34). We first derive the impact of this tax reform on labor effort before analyzing its effect on earnings and utility.

Lemma 2. *Suppose that Assumption 2 holds. The elasticity of effort with respect to progressivity $\varepsilon_{a,1-p} = \frac{\partial \log a}{\partial \log(1-p)}$ is given by*

$$\varepsilon_{a,1-p} = \frac{\varepsilon}{1+\varepsilon} \cdot \frac{1 + \varepsilon_{\psi,a}\,\psi^2\sigma_\eta^2}{1 + \frac{1-\varepsilon}{1+\varepsilon}\,\varepsilon_{\psi,a}\,\psi^2\sigma_\eta^2}, \tag{35}$$

where $\varepsilon_{\psi,a} = \frac{1}{\varepsilon}$ denotes the performance-pay elasticity (33). Thus, the labor supply elasticity $\varepsilon_{a,1-p}$ is strictly larger in the presence of moral hazard ($\varepsilon_{\psi,a} > 0$) than in the benchmark model with exogenous risk ($\varepsilon_{\psi,a} = 0$).

Proof. See Appendix E. □

Equation (35) gives an analytical expression for the labor supply elasticity that drives the response of labor supply to the tax reform (34), $\frac{\hat{a}(\theta)}{a(\theta)} = -\frac{1}{1-p}\varepsilon_{a,1-p}$. This elasticity is strictly larger in an economy with moral hazard and endogenous private insurance than in the benchmark setting with exogenous risk. Specifically, in the polar case where $\varepsilon_{\psi,a} = 0$, we obtain $\varepsilon_{a,1-p} = \frac{\varepsilon}{1+\varepsilon}$, which is an increasing function of the Frisch elasticity ε. Instead, when the exposure to risk varies endogenously with effort so that $\varepsilon_{\psi,a} > 0$, we have $\varepsilon_{a,1-p} > \frac{\varepsilon}{1+\varepsilon}$. The intuition underlying this result is analogous to the case of the utility without income effects analyzed in Lemma 1. Recall that the marginal cost of eliciting higher effort is equal to the expected marginal rate of substitution (MRS, first term in (7)) plus, in the presence of moral hazard, the expected marginal cost of incentive provision (MCI, second term in (7)). But the latter increases when the tax code becomes more progressive, since it is proportional to $\frac{1}{v'(w(\theta,\eta))} = \frac{w(\theta,\eta)}{1-p}$ if the utility is logarithmic and the tax schedule is CRP. Note

that the greater the Frisch elasticity, the more the standard model underestimates the true distortionary cost of raising tax progressivity.

Corollary 2. *The impact of an increase in progressivity on earnings is given by formula (14), where the standard adjustment in a model with exogenous risk is given by*

$$\frac{\hat{w}_{\mathrm{ex}}(\theta,\eta)}{w(\theta,\eta)} = -\frac{1}{1-p}\,\varepsilon_{a,1-p}, \tag{36}$$

the crowding-out effect is given by[23]

$$\frac{\hat{w}_{\mathrm{co}}(\theta,\eta)}{w(\theta,\eta)} = -\frac{1}{1-p}\,\varepsilon_{\psi,1-p}\left(\psi\eta - \psi^2\sigma_\eta^2\right), \tag{37}$$

and the performance-pay effect is given by

$$\frac{\hat{w}_{\mathrm{pp}}(\theta,\eta)}{w(\theta,\eta)} = -\frac{1}{1-p}\,\varepsilon_{\psi,a}\,\varepsilon_{a,1-p}\left(\psi\eta - \psi^2\sigma_\eta^2\right), \tag{38}$$

where $\varepsilon_{\psi,a} = \frac{1}{\varepsilon}$ and $\varepsilon_{\psi,1-p} = -1$ denote the pass-through elasticities (33). Overall, pre-tax earnings are strictly more exposed to output risk after the reform.

Proof. See Appendix E. ☐

Equations (36) to (38) give closed-form expressions for the three sources of earnings adjustments caused by the tax reform: the standard labor supply effect, the crowding-out effect, and the performance-pay effect. The first, $\hat{w}_{\mathrm{ex}}(\theta,\eta)$, is straightforward: it simply states that if wage rates are exogenous, the percentage earnings response to the reform, $\frac{\hat{w}_{\mathrm{ex}}(\theta,\eta)}{w(\theta,\eta)}$, is equal to the percentage effort response, $\frac{\hat{a}(\theta)}{a(\theta)} = -\frac{1}{1-p}\varepsilon_{a,1-p}$. Taking into account the endogeneity of private insurance yields the other two effects, $\hat{w}_{\mathrm{co}}(\theta,\eta)$ and $\hat{w}_{\mathrm{pp}}(\theta,\eta)$. As explained in Section 4.1 above, the crowding-out effect strictly *raises* the amount of risk to which workers are exposed, measured by the pass-through of performance shocks to log-earnings, via the crowding-out elasticity $\varepsilon_{\psi,1-p} < 0$. This pre-tax earnings adjustment ensures that their incentives are preserved. On the other hand, the reform lowers the optimal effort that firms would

---

[23]The proof in the Appendix provides the decomposition of $\hat{w}_{\mathrm{co}}(\theta,\eta)$ into its two effects highlighted in equation (15).

like workers to exert by $\varepsilon_{a,1-p} > 0$. This reduction in effort is elicited by improving the worker's insurance against output shocks. Thus, the performance-pay effect strictly *reduces* exposure to risk via the performance-pay elasticity $\varepsilon_{\psi,a} > 0$. This indirect increase in private insurance counteracts the direct crowding-out response to the policy change.

Overall, using the structural expression (35) for the labor supply elasticity $\varepsilon_{a,1-p}$, we can easily show that

$$\varepsilon_{\psi,1-p} \;+\; \varepsilon_{\psi,a}\,\varepsilon_{a,1-p} \;\; < \;\; 0. \tag{39}$$

As a consequence, we obtain that the total earnings adjustment due to moral hazard frictions, $\hat{w}_{\mathrm{co}}(\theta,\eta) + \hat{w}_{\mathrm{pp}}(\theta,\eta)$, unambiguously leads to an increase in the sensitivity of pre-tax log-earnings to performance shocks. That is, the crowding-out effect outweighs the performance-pay effect and private insurance is reduced on net.

The Crowding-Out and Performance-Pay Effects Almost Offset Each Other.
We now study the relative magnitude of the crowding-out and the performance-pay effects. Our conclusion is that, while each of them taken separately has a large impact on the structure of compensation, on net they almost fully offset each other so that the earnings schedule is only barely riskier following an increase in tax progressivity.

Recall that the crowding-out elasticity $\varepsilon_{\psi,1-p}$ is equal to 1 in absolute value: this is equivalent to saying that keeping effort constant, the variance of log-consumption remains constant after the tax reform and the endogenous earnings adjustment. The performance-pay effect, on the other hand, is driven by the optimal change in effort given by the labor supply elasticity $\varepsilon_{a,1-p}$. The firm implements this change in effort by adjusting the sensitivity of the contract via the pass-through elasticity $\varepsilon_{\psi,a}$. Why does this performance-pay effect $\varepsilon_{\psi,a}\varepsilon_{a,1-p}$ have the same order of magnitude as the direct crowding-out adjustment $\varepsilon_{\psi,1-p}$?

The key insight is that the performance-pay elasticity $\varepsilon_{\psi,a}$ is proportional to the *inverse* of the (Frisch) elasticity of labor supply ε. This is an immediate consequence of our key result that the slope of the contract (6) (or (30) in the loglinear model) is equal to the marginal disutility of labor $h'(a(\theta))$. Thus, to raise the worker's effort by $\hat{a}$, the firm must raise the slope of the contract in percentage terms by $\frac{h''(a(\theta))\hat{a}}{h'(a(\theta))} = \frac{1}{\varepsilon(\theta)}\frac{\hat{a}}{a}$, where $\varepsilon(\theta)$ is the Frisch elasticity of labor supply.

As a result, to the extent that the labor supply elasticity $\varepsilon_{a,1-p}$ has a similar order of magnitude as the Frisch elasticity of labor supply ε, the performance-pay effect is approximately equal to $\frac{1}{\varepsilon} \times \varepsilon_{a,1-p} \approx 1$, that is, about the same as the crowding-out effect. Crucially, *this result is robust to the value of the labor supply elasticity.* Indeed, if the labor supply elasticity is small, so that effort moves only a little in response to a tax change, then the pass-through must mechanically increase by a large amount in order to elicit this change in effort, so that the product of the two elasticities is always approximately equal to 1. Intuitively, if labor supply is very inelastic, effort will barely change in response to tax reforms; but precisely because of this inelastic behavior, a very large change in performance-sensitivity will then be necessary to convince workers to adjust their effort by this small amount.

The previous discussion is correct if the labor supply elasticity $\varepsilon_{a,1-p}$ (or, more generally, $\frac{\hat{a}}{a}$ in our general model) is indeed approximately equal to the Frisch elasticity ε. In practice, this need not be exactly the case. Formula (35) gives the structural expression for $\varepsilon_{a,1-p}$ as a function of ε and the variance of performance shocks σ_η^2. We showed that the endogeneity of private insurance raises the labor supply elasticity with respect to an increase in tax progressivity, relative to the benchmark setting with exogenous risk. Therefore, we know that $\varepsilon_{a,1-p}$ must be at least as large as its value in this environment, namely, $\frac{\varepsilon}{1+\varepsilon}$.[24] We therefore have

$$\varepsilon_{\psi,a}\,\varepsilon_{a,1-p} \;>\; \frac{1}{\varepsilon} \times \frac{\varepsilon}{1+\varepsilon} \;=\; \frac{1}{1+\varepsilon}.$$

For a Frisch elasticity $\varepsilon \approx \frac{1}{2}$, this means that the performance pay effect offsets at least $\frac{1}{1+0.5} = \frac{2}{3}$ of the crowding-out effect $\varepsilon_{\psi,1-p} = -1$. To refine this estimate, we use the structural expression for the labor supply elasticity $\varepsilon_{a,1-p}$ derived in Lemma 2. A Taylor approximation in $\psi^2\sigma_\eta^2$ yields

$$-\frac{\hat{w}_{\text{co}}\left(\theta,\eta\right)}{\hat{w}_{\text{pp}}\left(\theta,\eta\right)} \;=\; \frac{\varepsilon_{a,1-p}}{\varepsilon} \;\approx\; 1 + \left(\varepsilon - 2\psi^2\sigma_\eta^2\right). \tag{40}$$

Our calibration in Section 5.4 implies that the variance of earnings conditional on ability that best matches the data is equal to $\psi^2\sigma_\eta^2 \approx 0.2$. We thus get $\left(\varepsilon - 2\psi^2\sigma_\eta^2\right) \approx 0.1$, so that the earnings adjustment due to crowding-out amounts to about 110% (in absolute value) of the adjustment caused by the performance-pay effect. In other

---

[24]This value is lower than the Frisch elasticity ε because of the income effects on labor supply.

words, the labor supply responses offset about 90% of the crowding-out of private insurance by tax progressivity.

Discussion. The analysis of our quantitative model in Section 5 confirms that the crowding out effect (15) and the performance-pay effect (16) are both significant but offset each other almost entirely, so that the overall effect of tax progressivity on earnings risk is small. In Section 5.6 and Appendix B, we analyze both theoretically and numerically the incidence of several other tax reforms: lump-sum tax and marginal tax rate increases on high incomes, and a constant percentage increase in retention rates. For each of these reforms, as in the case of an increase in tax progressivity, the performance-pay effect counteracts, and sometimes even dominates, the direct crowding-out of the private insurance contract. Intuitively, raising marginal tax rates leads to a spread of the pre-tax earnings distribution, but the reduction in labor supply that it causes tends to contract it. Conversely, raising lump-sum tax payments leads to a crowding-in of private pre-tax insurance, but the income effect on labor supply again runs in the opposite direction. These results are consistent with the empirical literature. In particular, Frydman and Molloy (2011) exploit the relative tax advantage of different forms of CEO pay from 1946 to 2005 and find that the structure of compensation responds little to changes in tax rates on labor income. By highlighting the two counteracting forces at play – crowding-out versus performance-pay adjustments – our analysis provides an explanation for these findings.

4.3 Excess Burden and Welfare Gains

We now derive expressions for the excess burden and the welfare gains of raising the rate of progressivity of the tax schedule.

Corollary 3. *Suppose that Assumption 2 holds. Suppose moreover that ability types are lognormally distributed, $\log \theta \sim \mathcal{N}(\mu_\theta, \sigma_\theta^2)$. The excess burden of an increase in the rate of progressivity p of the CRP tax schedule is given by*

$$\mathcal{EB} = \left(\frac{1}{(1-g)(1-p)} - 1 \right) \varepsilon_{a,1-p} C + (\varepsilon_{\psi,1-p} + \varepsilon_{\psi,a} \varepsilon_{a,1-p}) p \psi^2 \sigma_\eta^2 C, \quad (41)$$

where C is the economy's aggregate private consumption, g is the ratio of government expenditures G to aggregate output $Y = C + G$, $\varepsilon_{a,1-p}$ is the labor effort elasticity

given by (35), *and* $\varepsilon_{\psi,a} = \frac{1}{\varepsilon}$, $\varepsilon_{\psi,1-p} = -1$ *are the pass-through elasticities.*

Suppose moreover that the planner is utilitarian, that is, $\alpha(\theta) = 1$ *for all* θ.[25] *The welfare gains (including the mechanical effect) of an increase in the rate of progressivity are given by*

$$\mathcal{ME} + \mathcal{WG} = (1 - p)\left(\sigma_\theta^2 + \psi^2 \sigma_\eta^2\right) C + \varepsilon_{\psi,1-p}\, \psi^2\, \sigma_\eta^2\, C. \tag{42}$$

Proof. See Appendix E. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Excess Burden of Raising Progressivity. The first term in the right-hand side of (41), $[((1 - g)(1 - p))^{-1} - 1]\varepsilon_{a,1-p}C$, is the standard deadweight loss from distorting labor effort, that is, the behavioral effect of taxation that would arise in a model with exogenous risk. This effect, equal to the first integral in equation (28), is increasing in the elasticity of effort with respect to progressivity (35) that measures the disincentive effects of raising tax rates, and in the rate of progressivity of the tax code that captures the share of income losses borne by the government as reduced revenue. Moreover, government expenditures raise the excess burden of tax progressivity. Intuitively, this is because a given marginal increase in tax progressivity implies a larger deadweight loss if the tax burden is already large due to high spending needs.

The second part of equation (41) captures the fiscal externalities that arise when private insurance against output risk is endogenous, that is, the two covariance terms in equation (28). The term $\varepsilon_{\psi,1-p}p(\psi\sigma_\eta)^2 C$ is the value of $\int \mathrm{Cov}(T', \hat{w}_{\mathrm{co}})\mathrm{d}F$, and the term $\varepsilon_{\psi,a}\varepsilon_{a,1-p}p(\psi\sigma_\eta)^2 C$ is the value of $\int \mathrm{Cov}(T', \hat{w}_{\mathrm{pp}})\mathrm{d}F$. Since $\varepsilon_{\psi,1-p} < 0$, the crowding-out effect contributes to reducing the excess burden of the reform, because it increases the sensitivity of earnings to output risk – by Jensen's inequality, this generates more tax revenue when the tax schedule is initially progressive ($p > 0$). Conversely, since $\varepsilon_{\psi,a}, \varepsilon_{a,1-p} > 0$ the performance-pay effect contributes to raising the excess burden (lowering government revenue) via a reduction in effort and hence risk exposure. Now recall that, by Corollary 2, the crowding-out effect dominates the performance-pay effect so that the earnings schedule becomes more risky overall. Therefore, conditional on the value of the labor effort elasticity, the deadweight loss of raising taxes is strictly *smaller* than in the standard model with exogenous risk.

---

[25]In the Appendix we consider the more general case where the social welfare weights are given by $\alpha(\theta) = \frac{e^{-\alpha \log \theta}}{\int e^{-\alpha \log \theta'}\, dF(\theta')}$ for all θ, where $\alpha \geq 0$.

However, we showed in Section 4.2 that the crowding-out and performance-pay effects almost fully offset each other. We therefore expect the net positive fiscal externality to be small in magnitude. We confirm this intuition in Section 5.

Welfare Gains of Raising Progressivity.   Finally, equation (42) shows that the welfare gains (including the mechanical effect) $\mathcal{ME} + \mathcal{WG}$ of the tax reform is the sum of two terms. The first, $(1-p)\left(\sigma_\theta^2 + (\psi\sigma_\eta)^2\right)$, captures the insurance gains obtained by raising the rate of progressivity of the tax schedule, as in a standard optimal taxation model. Note that tax progressivity insures both the initial ability differences θ and the performance shock η passed-through to earnings, that is, both between- and within-group heterogeneity. The larger their respective variances σ_θ^2 and $\psi^2\sigma_\eta^2$ and the lower the initial rate of progressivity $p \in (-\infty, 1)$, the higher the gains of marginally raising progressivity.

However, recall that the private insurance contract adjusts endogenously to the policy reform. Since the pass-through elasticity with respect to progressivity is equal to $\varepsilon_{\psi,1-p} = -1$, the welfare effect of this crowding-out (last term in equation (42)) satisfies

$$\varepsilon_{\psi,1-p}\psi^2\sigma_\eta^2 \;\; < \;\; \varepsilon_{\psi,1-p}\left(1-p\right)\psi^2\sigma_\eta^2 \;\; = \;\; -\left(1-p\right)\psi^2\sigma_\eta^2.$$

As a result, the crowding-out *more* than fully offsets the additional insurance against performance shocks provided by public policy, $(1-p)\psi^2\sigma_\eta^2$. Intuitively, in response to increased public insurance through higher tax progressivity, the firm adjusts the pre-tax earnings contract so that total (public plus private) insurance remains unchanged – there is a one-for-one crowding-out. However, there is an additional force at play. Recall that in our model, tax changes are intermediated by firms rather than being directly distributed to individual workers. We saw that as a consequence, the benefits of a tax cut accrue primarily to the richest workers of a given ability group. This strictly reduces the welfare gains of raising progressivity relative to the benchmark model with exogenous risk.

Note finally that, while the earnings distribution and government revenue remain practically unchanged following a tax reform as the crowding-out and performance-pay effects almost offset each other, the welfare effects of raising progressivity, on the other hand, can be large in magnitude. Indeed, labor supply adjustments cause only

second-order changes in welfare by the envelope theorem. As a result, the welfare implications of crowding-out are *not* counteracted by those of the performance-pay effect. We evaluate quantitatively the welfare cost of ignoring the endogeneity of private insurance in Section 5.3.

4.4   Optimal Rate of Progressivity

We finally gather our results on the excess burden and the welfare gain of tax reforms to characterize the optimal CRP tax schedule.

Proposition 4. *Suppose that Assumption 2 holds, that ability types are lognormally distributed, and that the social welfare objective is utilitarian. The optimal rate of progressivity satisfies*

$$\frac{p^*}{(1-p^*)^2} \;=\; \frac{\sigma_\theta^2 + (1+\varepsilon_{\psi,1-p})\,\psi^2\sigma_\eta^2}{\left(1+\frac{g}{(1-g)p^*}\right)\varepsilon_{a,1-p} + (1-p^*)\,\varepsilon_{\psi,a}\varepsilon_{a,1-p}\psi^2\sigma_\eta^2}, \tag{43}$$

where $g = G/Y$ is the ratio of government spending to output, $\varepsilon_{a,1-p}$ is the elasticity of effort with respect to the rate of progressivity given by (35), and $\varepsilon_{\psi,a} = \frac{1}{\varepsilon}$ and $\varepsilon_{\psi,1-p} = -1$ are the pass-through elasticities. In particular, the optimal rate of progressivity is strictly smaller in the model with endogenous private insurance than in the benchmark environment with exogenous risk where $\varepsilon_{\psi,1-p} = \varepsilon_{\psi,a} = 0$.

Proof. See Appendix E. □

Formula (43) is obtained by equating the excess burden $\mathcal{EB}$ to the welfare gain (including the mechanical effect) $\mathcal{ME} + \mathcal{WG}$ of raising progressivity, both derived in Corollary 3. Consider first the polar case with exogenous risk, that is, where all earnings differences are attributed to exogenous labor productivity shocks θ and η. In this case, letting $\varepsilon_{\psi,1-p} = \varepsilon_{\psi,a} = 0$ in formula (43) leads to

$$\frac{p^*}{(1-p^*)^2} \;=\; \left(1+\frac{g}{(1-g)\,p^*}\right)^{-1}\frac{\sigma_\theta^2 + \psi^2\sigma_\eta^2}{\varepsilon_{a,1-p}}. \tag{44}$$

Thus, the optimal rate of progressivity in this model is increasing in the variances of the ability and performance shock distributions, and decreasing in the elasticity of effort $\varepsilon_{a,1-p} = \frac{\varepsilon}{1+\varepsilon}$. In this benchmark setting, the government trades-off the benefits

37

of insuring the entire earnings risk, which is determined by the variance of log-earnings $\text{Var}(\log w) = \sigma_\theta^2 + \psi^2 \sigma_\eta^2$, with the excess burden of raising progressivity.

Now consider the general model with endogenous partial insurance against performance shocks, so that $\varepsilon_{\psi,1-p} < 0$ and $\varepsilon_{\psi,a} > 0$. Equating the excess burden to the welfare gains of raising progressivity implies that p^* is the solution to

$$\left(\frac{1}{(1-g)(1-p^*)} - 1\right) \varepsilon_{a,1-p} + (\varepsilon_{\psi,1-p} + \varepsilon_{\psi,a}\varepsilon_{a,1-p}) p^* \psi^2 \sigma_\eta^2$$

$$= (1-p^*)\left[\sigma_\theta^2 + \psi^2 \sigma_\eta^2\right] + \varepsilon_{\psi,1-p}\psi^2\sigma_\eta^2, \tag{45}$$

from which (43) follows. This formula implies that p^* is strictly decreasing in σ_η^2, and hence that the optimal rate of progressivity is strictly lower than in the previous case. This is because the positive fiscal externality and the welfare loss of crowding-out exactly cancel each other out, as they are respectively equal to $\varepsilon_{\psi,1-p}p^*\text{Var}(\log w \mid \theta)$ and $[(1-p^*) + \varepsilon_{\psi,1-p}p^*]\text{Var}(\log w \mid \theta)$, where $\text{Var}(\log w \mid \theta) = \psi^2\sigma_\eta^2$ is the variance of log-earnings conditional on ability θ. The only remaining term is therefore the negative fiscal externality due to the performance-pay effect, $\varepsilon_{\psi,a}\varepsilon_{a,1-p}p^*\text{Var}(\log w \mid \theta)$. This fiscal externality is captured by the second term in the denominator of (43). Overall, we obtain that by ignoring these effects, a planner that would ignore the endogeneity of private insurance would overestimate the optimal rate of tax progressivity.

5 Quantitative Analysis

In Section 5.1 we extend the model of Section 4 to make it suitable for policy analysis. Specifically, we incorporate the coexistence of jobs with and without performance pay and a Pareto tail for productivity types. We calibrate the model to match several key moments of U.S. data in Section 5.2. We then analyze the impact of two tax reforms in Sections 5.3 and 5.4. First, we consider a small reform that increases the rate of progressivity by one percentage point around the current tax code. Second, we study a large reform that nearly doubles the current rate of progressivity and brings the economy to the utilitarian optimum. We finally compare in Section 5.5 the optimal rate of progressivity in our calibrated model with two important benchmarks: the optimum in the model without performance-pay jobs, and the progressivity rate

chosen by a government that would wrongly assume that wage risk is exogenous.

5.1 Quantitative Model

We extend the loglinear model of Section 4 by adding the following elements. A share π of workers have a *performance-pay job*, denoted with a subscript m, and the remaining share $1 - \pi$ of workers have a *normal job* (subscript n). The output of the worker with productivity θ and a job type $j \in \{m, n\}$ is $\theta(a_j + \eta)$, where a_j is the effort level and $\eta \sim N\left(0, \sigma_{\eta,j}^2\right)$ is the performance shock. Performance-pay jobs are subject to the agency frictions described in Section 1. At these jobs, the employer observes the output but not the effort of the worker nor the performance shock and, hence, offers a wage which depends on the stochastic output realization according to the pass-through coefficient ψ. Normal jobs, in contrast, are free from agency frictions and guarantee a risk-free wage.

We treat the job type of a worker as exogenous. In the data, the share of performance pay jobs increases with earnings (see Lemieux, MacLeod, and Parent (2009); Grigsby, Hurst, and Yildirmaz (2019)). We allow the share of job types to be correlated with productivity by assuming that productivity is drawn from a job-type-specific Pareto-lognormal distribution (Colombi (1990)). That is, conditional on the job type $j \in \{n, m\}$, the log productivity is the sum of independently drawn normal and exponential random variables: $\log \theta = \theta_1 + \theta_2$ where $\theta_1 \sim \mathcal{N}\left(\mu_{\theta,j}, \sigma_{\theta,j}^2\right)$ and $\theta_2 \sim \mathcal{E}xp\left(\lambda_{\theta,j}\right)$. We keep government expenditures G fixed when comparing different policy scenarii. We derive and analyze the theoretical formula for the optimal rate of progressivity in this generalized environment in Appendix F.

5.2 Calibration

We calibrate to model to match the evidence on elasticities and wage distribution in the U.S. We choose the value $\varepsilon = 0.5$ for the Frisch elasticity, which implies a compensated elasticity of labor supply at normal jobs of approximately 0.3. Both values are consistent with empirical evidence (Keane (2011); Chetty (2012)).

We assume that log-productivity distributions for both job types have a common normal variance $\sigma_{\theta,m}^2 = \sigma_{\theta,n}^2 = \sigma_\theta^2$ and tail parameter $\lambda_{\theta,m} = \lambda_{\theta,n} = \lambda_\theta$. As a result,

our model implies that the variance of log earnings is given by

$$\text{Var}\left(\log w\right) = \left(\sigma_\theta^2 + \frac{1}{\lambda_\theta^2}\right) + \pi\psi^2\sigma_\eta^2 + \pi\left(1 - \pi\right)\left(\mu_m - \mu_n + \log\frac{a_m}{a_n} - \frac{\psi^2\sigma_\eta^2}{2}\right)^2,$$

where $\sigma_\theta^2 + \frac{1}{\lambda_\theta^2}$ is the variance of log-productivity, $\psi^2\sigma_\eta^2$ is the variance of log-earnings at the performance-pay jobs due to the performance shocks, and the last term captures the contribution of the difference between the mean log-earnings at normal and performance-pay jobs.

Lemieux, MacLeod, and Parent (2009) study performance-pay jobs using Panel Study of Income Dynamics (PSID) and find that their fraction π was 0.45 in 1998, the most recent year included in their analysis. They report that performance-pay jobs have mean hourly wages higher by 30%, and the variance of wages higher by 42%, relative to normal jobs.[26] The first statistic pins down $\mu_m - \mu_n = \log\left(1.3\right)$. The second will be crucial in determining $\sigma_{\eta,m}^2$, the variance of the performance shock at performance-pay jobs.

For the *levels* of the mean and variance of log-earnings in the entire economy, we turn to the Survey of Consumer Finances (SCF) which uses data from the Internal Revenue Service Statistics of Income program to accurately represent the distribution of high income households. Based on the SCF, Heathcote and Tsujiyama (2019) report a mean household labor income of $77,325$ and an overall variance of log labor income of 0.618 in 2007. They also estimate the tail parameter of the log earnings distribution λ_θ at 2.2.

Regarding the government policy, Heathcote, Storesletten, and Violante (2017) estimate the empirical rate of tax progressivity at 0.181 and Heathcote and Tsujiyama (2019) report a ratio of government purchases to output of 18.8 percent.

Given these estimates, we choose $\sigma_\theta^2 = 0.31$ and $\sigma_{\eta,m}^2 = 0.4$ to match the overall variance of log-earnings as well as the relative variance between performance-pay and normal jobs. Matching mean labor income as well as the ratio of mean wage rates at the two job types implies $\mu_{\theta,m} = 3.88$ and $\mu_{\theta,n} = 3.62$. The implied distribution of wage rates and job types is depicted in Figure 1. The share of performance-pay jobs is largest in the top quartile of the wage distribution. This is consistent with the

---

[26]These values are based on Table 1, Figure IV and Figure V in Lemieux, MacLeod, and Parent (2009). When statistics are available for multiple years, the last year available is used (either 1996 or 1998).

Figure 1: Joint distribution of wage rates and job types

(a)

(b)
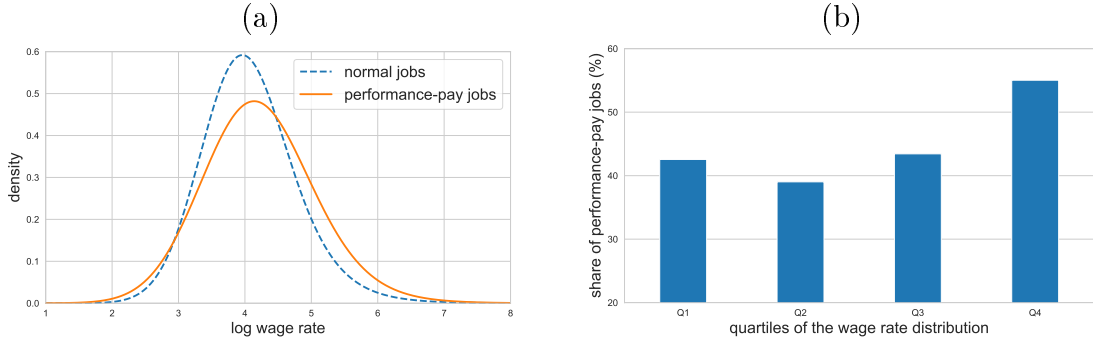
empirical evidence that bonuses and other forms of performance-related pay are more prevalent at higher income levels (Lemieux, MacLeod, and Parent (2009); Grigsby, Hurst, and Yildirmaz (2019)).

5.3   Marginal Reform of Tax Progressivity

Table 1 shows the impact of a small increase of the rate of progressivity by one percentage point, from 0.181 to 0.191, on the performance-pay jobs, the normal jobs, and all jobs. Note that the effects on performance-pay jobs are generally larger in absolute value, since performance-pay workers have higher average output and earnings than those in normal jobs. An increase in progressivity leads to a large redistributive gain for both types of jobs. For the performance-pay workers, an increase of progressivity would also lead to a substantial gain from better insurance against the earning risk if this risk was policy-invariant. However, an increase in progressivity generates a large crowding-out which, as we saw in Section 4, fully offsets the gains from insurance and, in addition, somewhat reduces the gains from redistribution. To understand the latter effect, note that workers who on average gained from the reform will see their earnings structure adjusted to keep incentives intact: their consumption will increase disproportionally in high-output contingencies, which reduces their expected utility gain. Hence, the crowding-out effect makes redistribution less potent. We find that, quantitatively, the redistributive gain for the performance-pay jobs is reduced by 6%.

The excess burden of the reform, $\mathcal{EB}$, is substantially larger for performance-pay jobs, because the elasticity of effort $\varepsilon_{a,1-p}$ is 25% greater for these workers. This difference in elasticities is only slightly mitigated by the combined impact of the fiscal

externalities caused by the crowding-out and performance-pay effects. These effects have a non-negligible impact on the excess burden when considered separately, but roughly cancel each other out and lead to a very modest positive fiscal externality. To understand this result, recall the excess burden formula obtained in Corollary 3. Since $\varepsilon_{\psi,1-p} = -1$, the positive fiscal externality due to crowding-out relative to the standard deadweight loss induced by labor supply responses is proportional to the inverse labor effort elasticity $1/\varepsilon_{a,1-p}$ and given by

$$\frac{1}{\varepsilon_{a,1-p}} \times \left(\frac{1}{(1-g)(1-p)} - 1\right)^{-1} p\psi^2\sigma_\eta^2 \approx 17.85\%.$$

Since $\varepsilon_{\psi,a} = \frac{1}{\varepsilon}$, the additional negative fiscal externality due to performance-pay relative to the standard deadweight loss is proportional to the inverse Frisch elasticity $1/\varepsilon$ and given by

$$-\frac{1}{\varepsilon} \times \left(\frac{1}{(1-g)(1-p)} - 1\right)^{-1} p\psi^2\sigma_\eta^2 \approx -14.9\%.$$

Therefore, each of these effects significantly alters the standard calculation of the excess burden of raising tax progressivity. However, the sum of these effects is proportional to the difference between the labor supply and the Frisch elasticities, $1/\varepsilon_{a,1-p} - 1/\varepsilon$. Since this difference is small, the overall fiscal externality caused by moral hazard frictions is equal to a mere 2.94% of the standard deadweight loss of raising tax progressivity. Therefore, this reform is only slightly less costly for the government budget than one would estimate by ignoring the endogeneity of private insurance.

5.4 Large Reform: From Status Quo to Optimum

We extend the theoretical optimal progressivity formula to our quantitative model with two types of jobs in Appendix F. We find that the utilitarian optimum progressivity rate is given by $p^* = 0.356$. This rate is almost double the current progressivity rate in the U.S., and the implied social welfare increase is equivalent to a 3% increase in consumption. To get a sense of the magnitude of this reform, note that the average tax rate, including transfers, of a worker with labor income $\$33,000$ would decrease from -0.7% to -14.2%. The tax rate at the mean household income $\$77,325$ would

42

Table 1: Impact of a small increase in the rate of progressivity

	Perf.-pay jobs	Normal jobs	All jobs
Welfare gain $\mathcal{ME} + \mathcal{WG}$	354	292	320
due to redistribution	376	292	330
due to insurance	114	0	51
due to crowding-out	-136	0	-61
Excess burden $\mathcal{EB}$	137	99	116
due to standard effect	141	99	118
due to crowding-out	-25	0	-11
due to performance-pay	21	0	9
Total: $\mathcal{ME} + \mathcal{WG}$ - $\mathcal{EB}$	217	192	203

increase from 13.6% to 15.6%. The tax rate at $500,000, which roughly corresponds to the top 1% threshold, would increase from 38.4% to 56.6%. In this section we analyze the impact of a large reform of the current tax code that implements the optimal progressivity and adjusts the other tax parameter to keep government revenue unchanged.

The impact of this reform is depicted in Figure 2 and analyzed in Table 2. Following a large increase in tax progressivity, the earnings schedule is barely altered: the pass-through of performance shocks to log-earnings increases only modestly from 0.731 to 0.76. As a result, the variance of log-earnings conditional on productivity increases by 8%, while the overall dispersion of log-earnings among performance pay workers increases by 2.3%. Given that the pre-tax earnings schedule hardly moves, this progressivity-increasing reform substantially flattens the consumption schedule, leading to a much better consumption insurance. Indeed, both the individual consumption risk and overall consumption dispersion among performance-pay workers fall by more than 30%. Better insured workers have weaker incentives to exert effort which, for our calibrated labor supply elasticity, falls by a substantial 9.6% due the large magnitude of the tax reform.

Underlying the weak response of the earnings schedule are two countervailing forces: the crowding-out of private insurance and the performance-pay effect. If firms attempted to motivate workers to maintain their original level of effort, better private

Figure 2: Earnings and consumption schedules of performance-pay workers

(a) Earnings schedule

(b) Consumption schedule
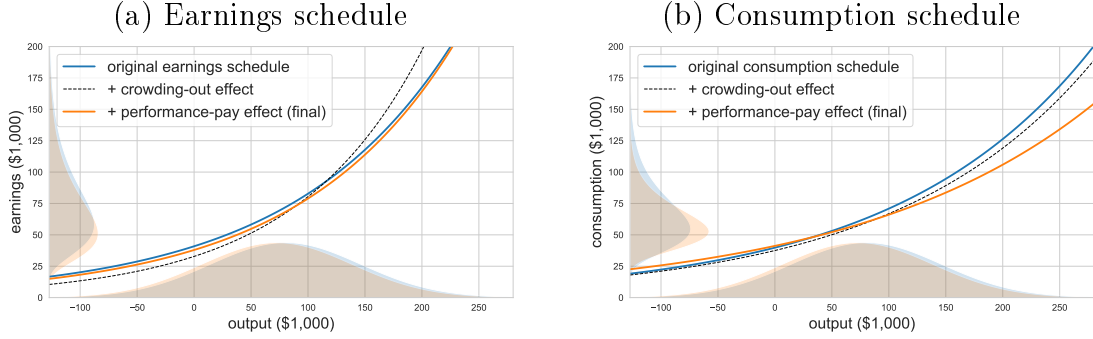
Note: The adjustment of the earnings and consumption schedules for a performance-pay worker with a mean productivity following an increase of progressivity rate from the current (0.181) to the optimal level (0.356).

Table 2: Earnings and consumption distribution statistics following the large reform

	Performance pay jobs			Normal jobs		
	$p = 0.181$	$p = 0.356$	% change	$p = 0.181$	$p = 0.356$	% change
effort	0.77	0.7	-9.59	0.94	0.86	-7.68
pass-through	0.73	0.76	3.89	0.	0.	0.
Individual wage and consumption risk						
$Var(\log(w \mid \theta))$	0.22	0.23	7.93	0.	0.	0.
$Var(\log(c \mid \theta))$	0.15	0.1	-33.2	0.	0.	0.
Overall wage and consumption dispersion						
$Var(\log(w))$	0.74	0.75	2.33	0.52	0.52	0.
$Var(\log(c))$	0.49	0.31	-36.66	0.35	0.22	-38.1

insurance via the income tax would crowd-out private insurance so as to leave the variance of log-earnings unchanged, since $\varepsilon_{\psi,1-p} = -1$. For that to happen, the pass-through would need to increase all the way to 0.93, raising the log-earnings risk of each performance-pay worker by 62%. However, firms in equilibrium choose a lower effort level and reduce the power of incentive-pay accordingly. This force – the performance-pay effect – counteracts the effect of crowding-out and brings the pass-through back to the vicinity of its original level. The combination of these two effects implies that, strikingly, the relative fall of log-consumption variance in the aftermath of the tax reform is nearly identical for the workers at jobs with and without agency frictions.
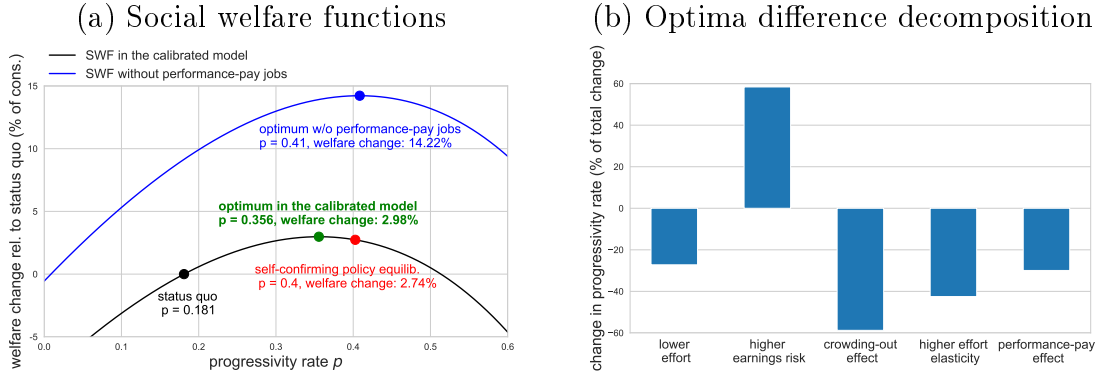
5.5 Performance-Pay Jobs and Optimal Progressivity

We now study the importance of performance-pay considerations for the optimal tax progressivity by comparing the optimal progressivity rate arising in the calibrated model to two important benchmarks. The first is the optimal rate of progressivity in the counterfactual economy without performance-pay jobs, obtained by setting the variance of the performance shock $\sigma^2_{\eta,m}$ to zero. The second is the rate of progressivity that would be chosen by the government who would erroneously assume that the entire wage risk is exogenous. This rate is found by applying the formula for the optimal rate of progressivity from the model with exogenous risk to our calibrated model economy, where wage-rate risk is actually endogenous. Following Rothschild and Scheuer (2016), we call the resulting progressivity rate a self-confirming policy equilibrium. The results are depicted in panel (a) of Figure 3.

First, in an economy without performance-pay jobs, the optimal rate of progressivity would increase from 0.356 to 0.41. To understand the discrepancy between the true optimum and this benchmark, we gradually switch on the various channels related to performance-pay jobs in the optimum formula (43), starting from an economy devoid of agency frictions (see panel (b) of Figure 3). First, workers at performance-pay jobs exert lower effort than those at normal jobs. This leads to a lower output and hence a higher share of government spending in GDP, G/Y. This in turn contributes to lower progressivity and explains approximately 30% of the overall progressivity change. Second, workers at performance-pay jobs face higher wage risk. That raises the gains of providing social insurance via tax progressivity. However, this additional benefit of insurance is fully canceled by the crowding-out of private insurance. Third, the labor supply elasticity of performance-pay workers is higher, increasing the excess burden of raising progressivity and explaining 40% of the progressivity difference. Finally, the performance-pay effect contributes to a reduction in government revenue via a compression of the earnings distribution, which explains the remaining 30%.

Second, we compare the optimal rate of progressivity with the SCPE. In this equilibrium concept, the government can correctly estimate the elasticity of effort, but treats wage risk as fully exogenous. Such a policymaker would mistakenly attempt to insure the endogenous part of the wage risk and choose too high a progressivity rate, equal to 0.4. However, log-earnings risk at performance-pay jobs increases by a mere 2.3% relative to its value in the true optimum. Once again, this is due to the counteracting forces of the crowding-out and the performance pay effects. Finally,

Figure 3: Optimal progressivity and performance pay jobs

(a) Social welfare functions	(b) Optima difference decomposition

Note: Panel (b) shows the contributions of various channels through which performance-pay jobs affect the optimal progressivity rate. These contributions are obtained by successively switching on the respective channels in the optimal progressivity formula. All the channels combined sum up to the difference in progressivity rate between the optimum without performance pay jobs and the optimum in the calibrated model.

ignoring the endogeneity of wage risk does not lead to a large miscalculation of optimal tax policy: the social welfare cost of choosing the SCPE is equivalent to a 0.24% drop in consumption. This value implies that increasing the U.S. rate of progressivity from the status quo to the SCPE reaps 93% of the welfare gains of moving from the status quo to the full optimum. Recall that this small welfare cost of sub-optimizing is *not* a necessary consequence of our theoretical analysis.[27] In fact, if there were only performance-pay jobs in the economy, the difference in progressivity between the SCPE and the full optimum would be 0.09 with a welfare difference equivalent to 1% change of consumption. Since in our calibration only roughly half of the jobs are performance-based, the impact on the progressivity rate is half of that. Furthermore, as the social welfare function is concave in p, half of the change in progressivity translates into a quarter of the change in social welfare.

5.6 Incidence of Other Tax Reforms

Recall that our theoretical tax incidence analysis of Section 2 gives us in closed-form the incidence of arbitrary tax reforms. Specifically, we can apply the theoretical formulas of Theorem 1 to reforms that do *not* keep the CRP structure of the tax

---

[27]Indeed, by the envelope theorem, the performance-pay effect is (at least locally) only second-order relative to the crowding-out effect from a welfare point of view. Thus, the two effects do not offset each other as they did when we studied the incidence on earnings and government revenue.
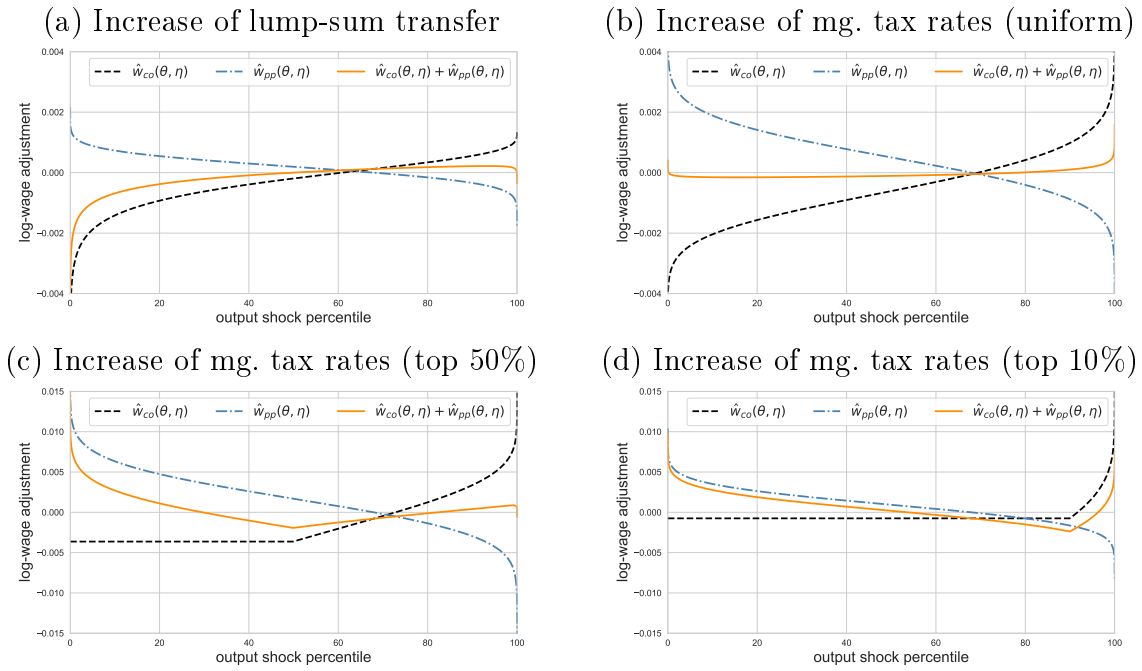
code.[28] We specialize our theoretical analysis to such reforms and study the direction of the crowding-out and the performance-pay effects in Appendix E. Here we propose two quantitative experiments.

First, in Figure 4 we depict the incidence of canonical tax reforms on the earnings schedule of a worker with mean ability and a performance-pay job. Specifically, we consider a $100 increase in lump-sum transfer as well as a 1 percentage point increase in marginal tax rates over a varying range of earnings, namely, for all earnings, for the top 50% earnings, and for the top 10% earnings. Our robust finding is that although the crowding-out effect (dashed black curves) contributes to higher earnings risk, it is mostly offset by the performance-pay effect (dashed-dotted blue curves). In the case of an additional lump-sum transfer, the performance-pay effect offsets more than 50% of the impact of crowding-out on the variance of log-earnings. For a uniform increase in marginal tax rates, the offset is more than 90%. When marginal tax rates are increased only for highest incomes, the offset rate even exceeds 100%: the performance-pay effect dominates the crowding-out effect and the earnings risk falls on net. To understand why the offset rate can be so high, recall that tax reforms which increase progressivity generate larger effort responses of performance-pay workers than reforms which spread the same tax burden in a more uniform manner – see Lemmas 1 and 2 and the subsequent discussions. Increasing marginal tax rates over a smaller range of high potential earnings – a progressive tax reform – generates a relatively larger effort response and, hence, a more substantial performance-pay effect in comparison to the crowding-out effect.

Second, in Figure 5, rather than focusing on a single earnings contract as in the previous paragraph, we study the impact of an increase in the top marginal tax rate in our calibrated economy with workers that are heterogeneous in ex-ante ability. Specifically, we consider a 1 percentage point increase in the marginal tax rate faced by the top 1% income earners, that is, above $441,000 in 2007 dollars. This hypothetical top tax bracket is depicted by the vertical dotted line in the left panel. The left (resp., right) panel gives the results as a function of mean earnings (resp., mean earnings percentile). This reform leads to a direct crowd-out which, ceteris paribus, increases the earnings risk for all workers in the top 5% of mean earnings, particularly so in the top 1%. This crowding-out is represented by the dashed black

---

[28]The only difficulty consists of computing the effort change $\hat{a}(\theta)$ in response to these reforms. We do so by solving the first-order condition (7) numerically.

Figure 4: Incidence of tax reforms: crowding-out and performance-pay effects

(a) Increase of lump-sum transfer

(b) Increase of mg. tax rates (uniform)

(c) Increase of mg. tax rates (top 50%)

(d) Increase of mg. tax rates (top 10%)

Note: Tax incidence computed for a worker with mean productivity. Panel (a) depicts an increase of the lump-sum transfer by \$100 (in 2007 dollars). Panels (b-d) depict an increase of the marginal tax rate by 1 pp for all earnings, for the top 50% earnings and for the top 10% earnings, respectively. The performance-pay effect offsets the impact of the crowding-out effect on the variance of log earnings by 52% (a), 92% (b), 136% (c) and 314% (d).

Figure 5: Increase of the top tax rate: impact on earnings risk

(a)   (b)
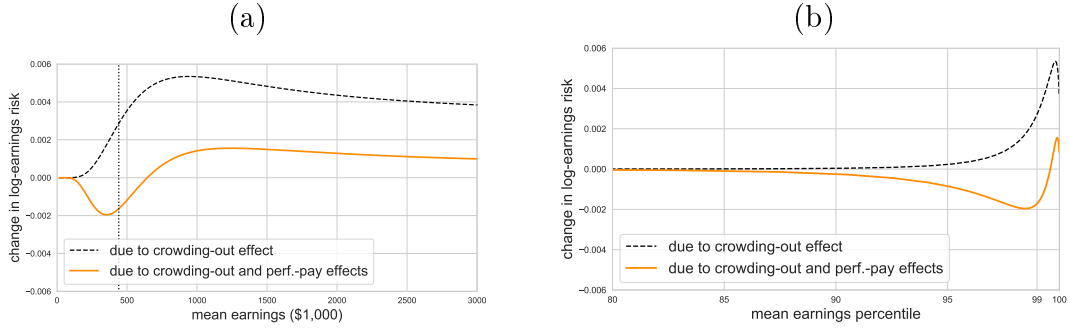
Note: Log-earnings risk is measured by $Var(\log(w(\theta, \eta) \mid \theta)$. The vertical dashed line in the left panel indicates the hypothetical top tax bracket threshold.

curves in both panels. However, the results change dramatically when we take into account labor effort responses. The performance-pay effect more than offsets the crowding-out effect everywhere apart from the very top earners, leading to a *lower* earnings risk for all workers below 99.5 percentile. Only the very top 0.5% experience any net crowding-out, but even for them the performance-pay effect offsets more than 70% of the additional pre-tax earnings risk.

Conclusion

We have set up and analyzed a tractable environment in which firms provide workers with endogenous private insurance against stochastic performance shocks in the face of moral hazard frictions. The government uses the tax-and-transfer system to redistribute income across workers who differ in uninsurable innate ability. The key feature of our model is that earnings risk is endogenous and has a productive role. The main and surprising conclusion of our analysis is that standard models that ignore the endogeneity of wage-rate risk actually come very close to evaluating the incidence of taxes on earnings contract, as well as the optimal level of tax progressivity. Underlying this result are two countervailing forces at play – a crowding-out and a performance-pay effect – which prevent taxes from having a large impact on the structure of performance-based compensation.

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 for firms to create such performance-pay

jobs (rather than "normal" jobs) in the first place, and study the incidence of tax reforms on the extensive margin of switching from one type of job to another. Second, in our model, private markets are constrained efficient and perfectly competitive. In other words, we gave private markets their "best chance" in making government policy redundant. Introducing frictions such as adverse selection in private markets – whereby firms cannot perfectly observe a worker's innate ability – and market power 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 via counteracting crowding-out and performance-pay effects. Testing these predictions empirically should be particularly fruitful.

References

ÁBRAHÁM, Á., F. ALVAREZ-PARRA, AND S. FORSTNER (2016a): "The effects of moral hazard on wage inequality in a frictional labor market," *Unpublished working paper*.

ÁBRAHÁM, Á., S. KOEHNE, AND N. PAVONI (2016b): "Optimal income taxation when asset taxation is limited," *Journal of Public Economics*, 136, 14–29.

ALES, L., A. A. BELLOFATTO, AND J. J. WANG (2017): "Taxing Atlas: Executive compensation, firm size, and their impact on optimal top income tax rates," *Review of Economic Dynamics*, 26, 62–90.

ALES, L., M. KURNAZ, AND C. SLEET (2015): "Technical change, wage inequality, and taxes," *The American Economic Review*, 105, 3061–3101.

ALES, L. AND C. SLEET (2016): "Taxing top CEO incomes," *American Economic Review*, 106, 3331–66.

ATTANASIO, O. AND J.-V. RIOS-RULL (2000): "Consumption smoothing in island economies: Can public insurance reduce welfare?" *European Economic Review*, 44, 1225–1258.

BANDIERA, O., I. BARANKAY, AND I. RASUL (2011): "Field experiments with firms," *Journal of Economic Perspectives*, 25, 63–82.

BELL, B. AND J. VAN REENEN (2014): "Bankers and their bonuses," *The Economic Journal*, 124, F1–F21.

BIRD, A. (2018): "Taxation and executive compensation: Evidence from stock options," *Journal of Financial Economics*, 127, 285–302.

BLOMQVIST, Å. AND H. HORN (1984): "Public health insurance and optimal income taxation," *Journal of Public Economics*, 24, 353–371.

BLOOM, N. AND J. VAN REENEN (2010): "Why do management practices differ across firms and countries?" *Journal of Economic Perspectives*, 24, 203–24.

CARROLL, G. AND D. MENG (2016): "Robust contracting with additive noise," *Journal of Economic Theory*, 166, 586–604.

CHANG, Y. AND Y. PARK (2017): "Optimal Taxation with Private Insurance," *Rochester Center for Economic Research Working Paper*.

CHETTY, R. (2009): "Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods," *Annual Review of Economics*, 1, 451–488.

——— (2012): "Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply," *Econometrica*, 80, 969–1018.

CHETTY, R. AND E. SAEZ (2010): "Optimal taxation and social insurance with endogenous private insurance," *American Economic Journal: Economic Policy*, 2, 85–114.

COLOMBI, R. (1990): "A new model of income distribution: The pareto-lognormal distribution," in *Income and wealth distribution, inequality and poverty*, Springer, 18–32.

CREMER, H. AND P. PESTIEAU (1996): "Redistributive taxation and social insurance," *International Tax and Public Finance*, 3, 281–295.

CULLEN, J. B. AND J. GRUBER (2000): "Does unemployment insurance crowd out spousal labor supply?" *Journal of Labor Economics*, 18, 546–572.

CUTLER, D. M. AND J. GRUBER (1996a): "Does public insurance crowd out private insurance?" *The Quarterly Journal of Economics*, 111, 391–430.

——— (1996b): "The effect of Medicaid expansions on public insurance, private insurance, and redistribution," *The American Economic Review*, 86, 378–383.

DALE-OLSEN, H. (2012): "Do Tax Reforms Affect Firm Performance and Executive Remuneration? Evidence from a Compressed Wage Environment," *Economica*, 79, 493–515.

DOLIGALSKI, P. (2019): "Optimal Income Taxation and Commitment on the Labor Market," *Working Paper*.

EDMANS, A. AND X. GABAIX (2011): "Tractability in incentive contracting," *The Review of Financial Studies*, 24, 2865–2894.

———— (2016): "Executive compensation: A modern primer," *Journal of Economic Literature*, 54, 1232–87.

EDMANS, A., X. GABAIX, AND D. JENTER (2017): "Executive compensation: A survey of theory and evidence," in *The Handbook of the Economics of Corporate Governance*, Elsevier, vol. 1, 383–539.

EDMANS, A., X. GABAIX, T. SADZIK, AND Y. SANNIKOV (2012): "Dynamic CEO compensation," *The Journal of Finance*, 67, 1603–1647.

FARHI, E. AND I. WERNING (2013): "Insurance and taxation over the life cycle," *The Review of Economic Studies*, 80, 596–635.

FINDEISEN, S. AND D. SACHS (2016): "Education and optimal dynamic taxation: The role of income-contingent student loans," *Journal of Public Economics*, 138, 1–21.

FOSTER, A. D. AND M. R. ROSENZWEIG (1994): "A test for moral hazard in the labor market: Contractual arrangements, effort, and health," *The Review of Economics and Statistics*, 213–227.

FRIEDRICH, B., L. LAUN, C. MEGHIR, AND L. PISTAFERRI (2019): "Earnings dynamics and firm-level shocks," Tech. rep., National Bureau of Economic Research.

FRYDMAN, C. AND D. JENTER (2010): "CEO compensation," *Annual Review of Financial Economics*, 2, 75–102.

FRYDMAN, C. AND R. S. MOLLOY (2011): "Does tax policy affect executive compensation? Evidence from postwar tax reforms," *Journal of Public Economics*, 95, 1425–1437.

GARRETT, D. F. AND A. PAVAN (2015): "Dynamic managerial compensation: A variational approach," *Journal of Economic Theory*, 159, 775–818.

GOLOSOV, M., N. KOCHERLAKOTA, AND A. TSYVINSKI (2003): "Optimal Indirect and Capital Taxation," *Review of Economic Studies*, 70, 569–587.

GOLOSOV, M., M. TROSHKIN, AND A. TSYVINSKI (2016): "Redistribution and social insurance," *American Economic Review*, 106, 359–86.

GOLOSOV, M. AND A. TSYVINSKI (2007): "Optimal taxation with endogenous insurance markets," *The Quarterly Journal of Economics*, 122, 487–534.

GRIGSBY, J., E. HURST, AND A. YILDIRMAZ (2019): "Aggregate nominal wage adjustments: New evidence from administrative payroll data," Tech. rep., National Bureau of Economic Research.

GUISO, L., L. PISTAFERRI, AND F. SCHIVARDI (2005): "Insurance within the firm," *Journal of Political Economy*, 113, 1054–1087.

HEATHCOTE, J., K. STORESLETTEN, AND G. L. VIOLANTE (2017): "Optimal tax progressivity: An analytical framework," *The Quarterly Journal of Economics*, 132, 1693–1754.

HEATHCOTE, J. AND H. TSUJIYAMA (2019): "Optimal Income Taxation: Mirrlees Meets Ramsey," *Working Paper*.

HENDREN, N. (2015): "The Policy Elasticity," in *Tax Policy and the Economy, Volume 30*, University of Chicago Press.

HOLMSTROM, B. AND P. MILGROM (1987): "Aggregation and linearity in the provision of intertemporal incentives," *Econometrica: Journal of the Econometric Society*, 303–328.

HUNGERBÜHLER, M., E. LEHMANN, A. PARMENTIER, AND B. VAN DER LINDEN (2006): "Optimal redistributive taxation in a search equilibrium model," *The Review of Economic Studies*, 73, 743–767.

KAPICKA, M. AND J. NEIRA (2013): "Optimal taxation in a life-cycle economy with endogenous human capital formation," *University of California, Santa Barbara WP*.

KAPLOW, L. (1991): "Incentives and government relief for risk," *Journal of Risk and Uncertainty*, 4, 167–175.

KEANE, M. P. (2011): "Labor supply and taxes: A survey," *Journal of Economic Literature*, 49, 961–1075.

KRUEGER, D. AND F. PERRI (2011): "Public versus private risk sharing," *Journal of Economic Theory*, 146, 920–956.

LAFFONT, J.-J. AND J. TIROLE (1986): "Using cost observation to regulate firms," *Journal of Political Economy*, 94, 614–641.

LAMADON, T. (2016): "Productivity shocks, long-term contracts and earnings dynamics," *manuscript, University of Chicago*.

LAMADON, T., M. MOGSTAD, AND B. SETZLER (2019): "Imperfect competition, compensating differentials and rent sharing in the US labor market," Tech. rep., National Bureau of Economic Research.

LAZEAR, E. P. AND P. OYER (2010): "Personnel Economics," *Handbook of Organizational Economics*.

LEMIEUX, T., W. B. MACLEOD, AND D. PARENT (2009): "Performance pay and wage inequality," *The Quarterly Journal of Economics*, 124, 1–49.

LEVITT, S. D. AND C. SYVERSON (2008): "Market distortions when agents are better informed: The value of information in real estate transactions," *The Review of Economics and Statistics*, 90, 599–611.

LUENBERGER, D. G. (1997): *Optimization by vector space methods*, John Wiley & Sons.

MAKRIS, M. AND A. PAVAN (2017): "Taxation under learning-by-doing," Tech. rep., Working Paper.

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

PARK, Y. (2014): "Optimal taxation in a limited commitment economy," *Review of Economic Studies*, 81, 884–918.

PIKETTY, T., E. SAEZ, AND S. STANTCHEVA (2014): "Optimal taxation of top labor incomes: A tale of three elasticities," *American Economic Journal: Economic Policy*, 6, 230–271.

PRENDERGAST, C. (1999): "The provision of incentives in firms," *Journal of Economic Literature*, 37, 7–63.

RAJ, A. (2019): "Optimal income taxation in the presence of networks of altruism," *Working Paper*.

ROCHET, J.-C. (1991): "Incentives, redistribution and social insurance," *The Geneva Papers on Risk and Insurance Theory*, 16, 143–165.

ROGERSON, W. P. (1985): "Repeated Moral Hazard," *Econometrica*, 53, 69–76.

ROSE, N. L. AND C. WOLFRAM (2002): "Regulating executive pay: Using the tax code to influence chief executive officer compensation," *Journal of Labor Economics*, 20, S138–S175.

ROTHSCHILD, C. AND F. SCHEUER (2013): "Redistributive Taxation in the Roy Model," *The Quarterly Journal of Economics*, 128, 623–668.

——— (2014): "A theory of income taxation under multidimensional skill heterogeneity," *NBER Working Paper No. 19822*.

——— (2016): "Optimal taxation with rent-seeking," *The Review of Economic Studies*, 83, 1225–1262.

SACHS, D., A. TSYVINSKI, AND N. WERQUIN (2020): "Nonlinear tax incidence and optimal taxation in general equilibrium," *Econometrica*, 88, 469–493.

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

SCHEUER, F. AND I. WERNING (2016): "Mirrlees meets Diamond-Mirrlees," *NBER Working Paper No. 22076*.

——— (2017): "The taxation of superstars," *The Quarterly Journal of Economics*, 132, 211–270.

SCHOENI, R. F. (2002): "Does unemployment insurance displace familial assistance?" *Public Choice*, 110, 99–119.

SHEARER, B. (2004): "Piece rates, fixed wages and incentives: Evidence from a field experiment," *The Review of Economic Studies*, 71, 513–534.

SLEET, C. AND H. YAZICI (2017): "Taxation, redistribution and frictional labor supply," *Working Paper*.

STANTCHEVA, S. (2014): "Optimal income taxation with adverse selection in the labour market," *Review of Economic Studies*, 81, 1296–1329.

——— (2017): "Optimal taxation and human capital policies over the life cycle," *Journal of Political Economy*, 125, 1931–1990.

A Proofs of Section 1

Concavity of the Utility of Earnings. Our analysis requires that the utility of earnings $w \mapsto v(w) \equiv u(R(w))$ is concave. It is easy to show that this is equivalent to

$$\pi_1(w)\,\pi_2(w) \;>\; -\gamma(w) \tag{46}$$

where $\gamma(w) \equiv -\frac{R(w)u''(R(w))}{u'(R(w))}$ is the agent's coefficient of relative risk aversion, and $\pi_1(w) \equiv \frac{1-T(w)/w}{1-T'(w)}$, $\pi_2(w) \equiv \frac{wT''(w)}{1-T'(w)}$ are two measures of the local rate of progressivity of the tax schedule. Specifically, the parameter $\pi_1(w)$ is the ratio of the average and marginal retention rates, and $\pi_2(w)$ 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. Note that most of our analysis is concerned with the incidence of tax reforms around a given initial tax schedule T. In this case, (46) is a restriction on the initial tax code T and can be easily verified in the data for practical applications. Note moreover that the tax reform itself is not restricted. When we characterize the optimal tax schedule within the CRP class, we assume that $u(c) = \log c$ which implies that $\gamma(w) = -1$. It is easy to verify that in this case condition (46) is always satisfied regardless of the value of p. $\qquad\square$

Proof of Proposition 1. The proof of this proposition follows directly from the results of Edmans and Gabaix (2011) since the utility of earnings $v(\cdot)$ is concave. We give here a heuristic proof of the main arguments. Given the earnings contract $\{w(\theta,\eta) : y \in \mathbb{R}\}$, an agent with ability θ and performance shock η chooses effort $a(\theta)$ to maximize utility $u(R(w(\theta,\eta))) - h(a(\theta))$ (see equation (3)). Since $y = \theta(a+\eta)$, we have $\frac{\partial w(\theta,\eta)}{\partial \eta} = \frac{\partial w(\theta,\eta)}{\partial a}$ so that the first-order condition reads

$$r(w(\theta,\eta))\,u'(R(w(\theta,\eta)))\frac{\partial w(\theta,\eta)}{\partial \eta} \;=\; h'(a(\theta)). \tag{47}$$

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

$$u(R(w(\theta,\eta))) \;=\; h'(a(\theta))\,\eta + k, \tag{48}$$

for some constant $k \in \mathbb{R}$. Since in equilibrium the participation constraint (4) 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

$$k = U(\theta) + h(a(\theta)). \tag{49}$$

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

$$u(R(w(\theta, \eta))) = h'(a(\theta))\eta + [U(\theta) + h(a(\theta))]. \tag{50}$$

Next, equation (7) is obtained by taking the first-order condition with respect to $a(\theta)$ in the firm's problem (2), taking as given the earnings contract (6) required to satisfy the workers' incentive and participation constraints (3, 4). Finally, equation (8) is simply a rewriting of (5). $\qquad\square$

B   Proofs of Section 2

Proof of Theorem 1. Recall that, given the effort level a and the reservation value $U(\theta)$, earnings as a function of the noise realization η (or equivalently output $y = \theta(a + \eta)$) satisfies:

$$u(w(\theta, \eta) - T(w(\theta, \eta))) = U(\theta) + h(a(\theta)) + h'(a(\theta))\eta.$$

In response to the tax reform $\delta \hat{T}$, the perturbed wage contract satisfies

$$
\begin{aligned}
&u\left[w(\theta, \eta) + \delta\hat{w}(\theta, \eta) - T(w(\theta, \eta) + \delta\hat{w}(\theta, \eta)) - \delta\hat{T}(w(\theta, \eta))\right] \\
&= U(\theta) + \delta\hat{U}(\theta) + h(a(\theta) + \delta\hat{a}(\theta)) + h'(a(\theta) + \delta\hat{a}(\theta))\eta.
\end{aligned}
$$

Differentiating with respect to δ and evaluating at $\delta = 0$ leads to

$$\left[\left(1 - T'\left(w\left(\theta, \eta\right)\right)\right) \hat{w}\left(\theta, \eta\right) - \hat{T}\left(w\left(\theta, \eta\right)\right) \right] u'\left(w\left(\theta, \eta\right) - T\left(w\left(\theta, \eta\right)\right)\right)$$
$$= \hat{U}\left(\theta\right) + \left[h'\left(a\left(\theta\right)\right) + h''\left(a\left(\theta\right)\right)\eta\right]\hat{a}\left(\theta\right).$$

Solving for $\hat{w}\left(\theta, \eta\right)$ leads to

$$\hat{w}\left(\theta, \eta\right) = \frac{\hat{T}\left(w\left(\theta, \eta\right)\right)}{r\left(w\left(\theta, \eta\right)\right)} + \frac{\hat{U}\left(\theta\right)}{r\left(w\left(\theta, \eta\right)\right) u'\left(R\left(w\left(\theta, \eta\right)\right)\right)}$$
$$+ \frac{h'\left(a\left(\theta\right)\right) + h''\left(a\left(\theta\right)\right)\eta}{r\left(w\left(\theta, \eta\right)\right) u'\left(R\left(w\left(\theta, \eta\right)\right)\right)}\hat{a}\left(\theta\right).$$

Adding and subtracting $w\left(\theta, \eta\right)\frac{\hat{a}(\theta)}{a(\theta)}$, that is, the earnings adjustment obtained in the model with exogenous risk, and substituting expression (17) derived below for the impact of the reform on expected utility $\hat{U}\left(\theta\right)$, easily yields (14). $\qquad\square$

<u>Proof of Proposition 2</u>. Recall that the reservation value satisfies $\mathbb{E}\left[w\left(\theta, \eta\right)\right] = \theta a\left(\theta\right)$. Hence, in response to the tax reform, we get

$$\mathbb{E}\left[w\left(\theta, \eta\right) + \delta\hat{w}\left(\theta, \eta\right)\right] = \theta\left(a\left(\theta\right) + \delta\hat{a}\left(\theta\right)\right),$$

that is, $\mathbb{E}\left[\hat{w}\left(\theta, \eta\right)\right] = \theta\hat{a}\left(\theta\right)$. Substituting expression (14) for $\hat{w}\left(\theta, \eta\right)$ in this equation leads to

$$\theta\hat{a}\left(\theta\right) = \mathbb{E}\left[\frac{\hat{T}\left(w\left(\theta, \eta\right)\right)}{r\left(w\left(\theta, \eta\right)\right)}\right] + \mathbb{E}\left[\frac{1}{r\left(w\left(\theta, \eta\right)\right) u'\left(R\left(w\left(\theta, \eta\right)\right)\right)}\right]\hat{U}\left(\theta\right)$$
$$+ \mathbb{E}\left[\frac{h'\left(a\left(\theta\right)\right) + h''\left(a\left(\theta\right)\right)\eta}{r\left(w\left(\theta, \eta\right)\right) u'\left(R\left(w\left(\theta, \eta\right)\right)\right)}\right]\hat{a}\left(\theta\right).$$

Using equation (7) that defines optimal effort and solving for $\hat{U}\left(\theta\right)$ leads to (17). $\qquad\square$

<u>Proof of Lemma 1</u>. Recall the optimal effort condition

$$\mathbb{E}\left[\frac{h'\left(a\left(\theta\right)\right) + h''\left(a\left(\theta\right)\right)\eta}{\left(1 - T'\left(w\left(\theta, \eta\right)\right)\right) u'\left(w\left(\theta, \eta\right) - T\left(w\left(\theta, \eta\right)\right)\right)}\right] = \theta.$$

60

Following a tax reform $\hat{T}$, the perturbed level of effort satisfies

$$\mathbb{E}\left[\frac{h'\left(a\left(\theta\right)+\delta\hat{a}\left(\theta\right)\right)+h''\left(a\left(\theta\right)+\delta\hat{a}\left(\theta\right)\right)\eta}{\left\{1-T'\left(\tilde{w}\left(\theta,\eta\right)\right)-\delta\hat{T}'\left(\tilde{w}\left(\theta,\eta\right)\right)\right\}u'\left\{R\left(\tilde{w}\left(\theta,\eta\right)\right)-\delta\hat{T}\left(\tilde{w}\left(\theta,\eta\right)\right)\right\}}\right] = \theta,$$

where we denote $\tilde{w}\left(\theta,\eta\right)\equiv w\left(\theta,\eta\right)+\delta\hat{w}\left(\theta,\eta\right)$. Taking the derivative of this expression with respect to δ evaluated at $\delta=0$ gives

$$
\begin{aligned}
0 &= \mathbb{E}\left[\frac{h''\left(a\left(\theta\right)\right)+h'''\left(a\left(\theta\right)\right)\eta}{r\left(w\left(\theta,\eta\right)\right)u'\left(R\left(w\left(\theta,\eta\right)\right)\right)}\hat{a}\left(\theta\right)\right]-\mathbb{E}\left[\frac{h'\left(a\left(\theta\right)\right)+h''\left(a\left(\theta\right)\right)\eta}{\left\{r\left(w\left(\theta,\eta\right)\right)u'\left(R\left(w\left(\theta,\eta\right)\right)\right)\right\}^{2}}\right. \qquad (51)\\
&\quad \left.\times\left\{\left[-T''\left(w\right)\hat{w}-\hat{T}'\left(w\right)\right]u'\left(R\left(w\right)\right)+r\left(w\right)\left[r\left(w\right)\hat{w}-\hat{T}\left(w\right)\right]u''\left(R\left(w\right)\right)\right\}\right]
\end{aligned}
$$

where the arguments $\left(\theta,\eta\right)$ have been removed from the second line for notational conciseness. Suppose that the utility function is quasilinear in consumption, and the tax schedule is initially affine. Equation (51) can then be rewritten as

$$
\begin{aligned}
0 &= \mathbb{E}\left[\frac{h''\left(a\left(\theta\right)\right)+h'''\left(a\left(\theta\right)\right)\eta}{1-\tau}\hat{a}\left(\theta\right)\right]+\mathbb{E}\left[\frac{h'\left(a\left(\theta\right)\right)+h''\left(a\left(\theta\right)\right)\eta}{\left(1-\tau\right)^{2}}\hat{T}'\left(w\left(y\mid\theta\right)\right)\right]\\
&= \frac{h''\left(a\left(\theta\right)\right)}{1-\tau}\hat{a}\left(\theta\right)+\frac{h'\left(a\left(\theta\right)\right)}{\left(1-\tau\right)^{2}}\mathbb{E}\left[\hat{T}'\left(w\left(y\mid\theta\right)\right)\right]+\frac{h''\left(a\left(\theta\right)\right)}{\left(1-\tau\right)^{2}}\mathbb{E}\left[\eta\hat{T}'\left(w\left(y\mid\theta\right)\right)\right].
\end{aligned}
$$

Solving for $\hat{a}\left(\theta\right)$ and letting $\varepsilon\left(\theta\right)=\frac{h'\left(a\left(\theta\right)\right)}{a\left(\theta\right)h''\left(a\left(\theta\right)\right)}$ easily leads to (18). $\qquad\square$

<u>Proof of Proposition 3</u>. Solving for $\hat{a}\left(\theta\right)$ in equation (51) implies that the Gateaux derivative of effort is given by

$$
\begin{aligned}
\hat{a}\left(\theta\right) &= -\mathbb{E}\left[\bar{\varepsilon}_{a,R}\left(\theta,\eta\right)\frac{\hat{T}\left(w\left(\theta,\eta\right)\right)}{r\left(w\left(\theta,\eta\right)\right)w\left(\theta,\eta\right)}\right]-\mathbb{E}\left[\bar{\varepsilon}_{a,r}\left(\theta,\eta\right)\frac{\hat{T}'\left(w\left(\theta,\eta\right)\right)}{r\left(w\left(\theta,\eta\right)\right)}\right]\\
&\quad +\mathbb{E}\left[\left(\bar{\varepsilon}_{a,R}\left(\theta,\eta\right)-p\left(w\left(\theta,\eta\right)\right)\bar{\varepsilon}_{a,r}\left(\theta,\eta\right)\right)\frac{\hat{w}\left(\theta,\eta\right)}{w\left(\theta,\eta\right)}\right],
\end{aligned}
$$

where $p\left(w\right)\equiv\frac{wT''\left(w\right)}{1-T'\left(w\right)}$ is the local rate of progressivity of the tax schedule, and where $\bar{\varepsilon}_{a,R},\bar{\varepsilon}_{a,r}$ denote the income effect parameter and compensated elasticity along the

linearized budget constraint, equal to

$$\bar{\varepsilon}_{a,r}\left(\theta,\eta\right) = \frac{h'\left(a\left(\theta\right)\right)}{a\left(\theta\right)h''\left(a\left(\theta\right)\right)} \frac{\left(1+\frac{h''(a(\theta))}{h'(a(\theta))}\eta\right)\frac{1}{r(w(\theta,\eta))u'(R(w(\theta,\eta)))}}{\mathbb{E}\left[\left(1+\frac{h'''(a(\theta))}{h''(a(\theta))}\eta\right)\frac{1}{r(w(\theta,\eta))u'(R(w(\theta,\eta)))}\right]}$$

and

$$\bar{\varepsilon}_{a,R}\left(\theta,\eta\right) = \frac{h'\left(a\left(\theta\right)\right)}{a\left(\theta\right)h''\left(a\left(\theta\right)\right)} \frac{\left(1+\frac{h''(a(\theta))}{h'(a(\theta))}\eta\right)\frac{w(\theta,\eta)u''(R(w(\theta,\eta)))}{(u'(R(w(\theta,\eta))))^2}}{\mathbb{E}\left[\left(1+\frac{h'''(a(\theta))}{h''(a(\theta))}\eta\right)\frac{1}{r(w(\theta,\eta))u'(R(w(\theta,\eta)))}\right]}.$$

Now substitute equations (14, 17) for $\hat{w}\left(\theta,\eta\right)$ in the previous equation, and solve for $\hat{a}\left(\theta\right)$ to get

$$\left\{1-\mathbb{E}\left[\left(\bar{\varepsilon}_{a,R}\left(\theta,\eta\right)-p\left(w\left(\theta,\eta\right)\right)\bar{\varepsilon}_{a,r}\left(\theta,\eta\right)\right)\frac{h'\left(a\left(\theta\right)\right)+h''\left(a\left(\theta\right)\right)\eta}{v'\left(w\left(\theta,\eta\right)\right)w\left(\theta,\eta\right)}\right]\right\}\hat{a}\left(\theta\right)$$

$$= -\mathbb{E}\left[\bar{\varepsilon}_{a,R}\left(\theta,\eta\right)\frac{\hat{T}\left(w\left(\theta,\eta\right)\right)}{\left(1-T'\left(w\left(\theta,\eta\right)\right)\right)w\left(\theta,\eta\right)}\right] - \mathbb{E}\left[\bar{\varepsilon}_{a,r}\left(\theta,\eta\right)\frac{\hat{T}'\left(w\left(\theta,\eta\right)\right)}{1-T'\left(w\left(\theta,\eta\right)\right)}\right]$$

$$+\mathbb{E}\left[\left(\bar{\varepsilon}_{a,R}\left(\theta,\eta\right)-p\left(w\left(\theta,\eta\right)\right)\bar{\varepsilon}_{a,r}\left(\theta,\eta\right)\right)\frac{\hat{T}\left(w\left(\theta,\eta\right)\right)}{r\left(w\left(\theta,\eta\right)\right)w\left(\theta,\eta\right)}\right]$$

$$-\mathbb{E}\left[\left(\bar{\varepsilon}_{a,R}\left(\theta,\eta\right)-p\left(w\left(\theta,\eta\right)\right)\bar{\varepsilon}_{a,r}\left(\theta,\eta\right)\right)\frac{\frac{1}{v'(w(\theta,\eta))w(\theta,\eta)}}{\mathbb{E}\left[\left(v'\left(w\left(\theta,\cdot\right)\right)\right)^{-1}\right]}\right]\mathbb{E}\left[\frac{\hat{T}\left(w\left(\theta,\eta\right)\right)}{r\left(w\left(\theta,\eta\right)\right)}\right].$$

Collecting terms leads to

$$\frac{\hat{a}\left(\theta\right)}{a\left(\theta\right)} = -\mathbb{E}\left[\varepsilon_{\mathbb{E}w,R}\left(\theta,\eta\right)\frac{\hat{T}\left(w\left(\theta,\eta\right)\right)}{r\left(w\left(\theta,\eta\right)\right)w\left(\theta,\eta\right)}\right] - \mathbb{E}\left[\varepsilon_{\mathbb{E}w,r}\left(\theta,\eta\right)\frac{\hat{T}'\left(w\left(\theta,\eta\right)\right)}{r\left(w\left(\theta,\eta\right)\right)}\right]$$

where the income effect parameter and compensated elasticity now account for the nonlinearity of the budget constraint (due to the fact that $\hat{w}$ depends on $\hat{T}$) and the endogeneity of the reservation value (due to the fact that $\hat{w}$ depends on $\hat{U}$) and are given by

$$\varepsilon_{\mathbb{E}w,R}\left(\theta,\eta\right) = \frac{p\left(w\left(\theta,\eta\right)\right)\bar{\varepsilon}_{a,r}\left(\theta,\eta\right)+\frac{\mathbb{E}\left[\left(\bar{\varepsilon}_{a,R}(\theta,\cdot)-p(w(\theta,\cdot))\bar{\varepsilon}_{a,r}(\theta,\cdot)\right)\frac{1}{v'(w(\theta,\cdot))w(\theta,\cdot)}\right]}{\mathbb{E}\left[\frac{1}{v'(w(\theta,\cdot))}\right]}w\left(\theta,\eta\right)}{1+\mathbb{E}\left[\left(p\left(w\left(\theta,\eta'\right)\right)\bar{\varepsilon}_{a,r}\left(\theta,\eta'\right)-\bar{\varepsilon}_{a,R}\left(\theta,\eta'\right)\right)\left(1+\frac{h''(a(\theta))}{h'(a(\theta))}\eta'\right)\frac{a(\theta)h'(a(\theta))}{v'(w(\theta,\eta'))w(\theta,\eta')}\right]}$$

62

and

$$\varepsilon_{\mathbb{E}w,r}\left(\theta,\eta\right) = \frac{\bar{\varepsilon}_{a,r}\left(\theta,\eta\right)}{1+\mathbb{E}\left[\left(p\left(w\left(\theta,\eta'\right)\right)\bar{\varepsilon}_{a,r}\left(\theta,\eta'\right)-\bar{\varepsilon}_{a,R}\left(\theta,\eta'\right)\right)\left(1+\frac{h''(a(\theta))}{h'(a(\theta))}\eta'\right)\frac{a(\theta)h'(a(\theta))}{v'(w(\theta,\eta'))w(\theta,\eta')}\right]}.$$

This concludes the proof. $\qquad\square$

C   Allowing for Non-Constant Effort

In this section we extend our tax incidence analysis to the case where the firm can offer an effort schedule $a(\theta,\eta)$, rather than imposing a constant effort level $a(\theta)$ as in the main body of the paper. The firm which employs a worker with productivity θ solves

$$\max_{w(\theta,\cdot),a(\theta,\cdot)} \mathbb{E}[\theta a(\theta,\eta) - w(\theta,\eta)]$$

subject to incentive-compatibility constraints

$$a(\theta,\eta) \in \arg\max_{a} v(w(\theta,\eta)) - h(a(\theta,\eta)) \quad \text{for all } \eta,$$

and the participation constraint

$$\mathbb{E}\left[v(w(\theta,\eta)) - h(a(\theta,\eta))\right] \geq U(\theta).$$

From Edmans and Gabaix (2011) we know that the optimal contract satisfies

$$v(w(\theta,\eta)) = K + h(a(\theta,\eta)) + \int_{\underline{\eta}}^{\eta} h'(a(\theta,x))dx,$$

where $K \in \mathbb{R}$. Using the binding participation constraint to solve for K, we obtain

$$v(w(\theta,\eta)) = U(\theta) + h(a(\theta,\eta)) + \int_{\underline{\eta}}^{\eta} h'(a(\theta,x))dx - \mathbb{E}\left[\int_{\underline{\eta}}^{\eta} h'(a(\theta,x))dx\right]. \qquad (52)$$

Unlike in the model with a single effort level, the slope of the ex-post utility potentially varies with performance. In particular, when the effort schedule is differentiable, we have

$$\frac{\partial v(w(\theta,\eta))}{\partial \eta} = h'(a(\theta,\eta))\left(1 + \frac{\partial a(\theta,\eta)}{\partial \eta}\right).$$

Edmans and Gabaix (2011) show that all incentive-compatible effort schedules are such that

$a(\theta, \eta) + \eta$ is increasing with η. This implies that the above slope is non-negative and that earnings are increasing with performance.

The first-order condition with respect to effort level $a(\theta, \eta')$, assuming an interior solution, reads

$$\theta f_\eta(\eta') = \frac{\partial \mathbb{E}\left[w(\theta, \eta)\right]}{\partial a(\theta, \eta')}, \tag{53}$$

where f_η denotes the density of the performance shock η. To compute the derivative of the expected wage on the right-hand side, first consider the derivative of the wage $w(\theta, \eta)$ with respect to the effort level $a(\theta, \eta')$, which we can compute using equation (52):

$$\frac{\partial w(\theta, \eta)}{\partial a(\theta, \eta')} = \begin{cases} -(1 - F_\eta(\eta')) \frac{h''(a(\theta, \eta'))}{v'(w(\theta, \eta))} & \text{if } \eta < \eta', \\ \frac{h'(a(\theta, \eta'))}{v'(w(\theta, \eta'))} - (1 - F_\eta(\eta')) \frac{h''(a(\theta, \eta'))}{v'(w(\theta, \eta'))} & \text{if } \eta = \eta', \\ \frac{h''(a(\theta, \eta'))}{v'(w(\theta, \eta))} - (1 - F_\eta(\eta')) \frac{h''(a(\theta, \eta'))}{v'(w(\theta, \eta))} & \text{if } \eta > \eta'. \end{cases} \tag{54}$$

Notice that increasing the effort level conditional on the output shock η' requires lowering earnings for worse performance ($\eta < \eta'$) and increasing earnings for better performance ($\eta > \eta'$). Taking expectations over η yields

$$\begin{aligned} \frac{\partial \mathbb{E}\left[w(\theta, \eta)\right]}{\partial a(\theta, \eta')} &= \frac{h'(a(\theta, \eta'))}{v'(w(\theta, \eta'))} f_\eta(\eta') \\ &\quad + (1 - F_\eta(\eta')) \left(\mathbb{E}\left[\frac{h''(a(\theta, \eta'))}{v'(w(\theta, \eta))} \mid \eta \geq \eta' \right] - \mathbb{E}\left[\frac{h''(a(\theta, \eta'))}{v'(w(\theta, \eta))} \right] \right). \end{aligned}$$

Plugging this expression into the first-order condition (53) yields the following first-order condition with respect to effort:

$$\begin{aligned} \theta f_\eta(\eta') &= \frac{h'(a(\theta, \eta'))}{v'(w(\theta, \eta'))} f_\eta(\eta') \\ &\quad + (1 - F_\eta(\eta')) \left(\mathbb{E}\left[\frac{h''(a(\theta, \eta'))}{v'(w(\theta, \eta))} \mid \eta \geq \eta' \right] - \mathbb{E}\left[\frac{h''(a(\theta, \eta'))}{v'(w(\theta, \eta))} \right] \right). \end{aligned} \tag{55}$$

The left-hand side is the marginal benefit from providing higher effort, equal to the expected output gain. The right-hand side consists of two terms: the marginal rate of substitution (MRS) and the marginal cost of incentives (MCI). The MRS is the expected wage cost of compensating the agent for higher effort in contingency η'. The MCI, on the other hand, is the expected wage cost of making the adjusted effort schedule incentive-compatible. As we noted above, increasing effort level conditional on output η' requires increasing earnings for better performance and reducing them for worse performance. Since earnings are increasing with performance and v is concave, such earnings adjustments are costly for the firm:

$MCI \geq 0$.

The following theorem extends the results of Theorem 1 to the model where non-degenerate effort schedules are allowed.

Theorem 3. *Denote by $\hat{a}(\theta, \eta)$ the change in effort schedule induced by the reform. Suppose that the original effort schedule is such that $a(\theta, \eta) > 0$ for all η. The first-order effect of the tax reform $\hat{T}$ on earnings $w(\theta, \cdot)$ is given by*

$$\hat{w}(\theta, \eta) = \hat{w}_{\mathrm{ex}}(\theta, \eta) + \hat{w}_{\mathrm{co}}(\theta, \eta) + \hat{w}_{\mathrm{pp}}(\theta, \eta)$$

where $\hat{w}_{\mathrm{ex}}(\theta, \eta) = \theta \hat{a}(\theta, \eta)$, the crowding-out effect $\hat{w}_{\mathrm{co}}$ has mean zero and is given by

$$\hat{w}_{\mathrm{co}}(\theta, \eta) = \frac{\hat{T}(w(\theta, \eta))}{r(w(\theta, \eta))} - \frac{(v'(w(\theta, \eta)))^{-1}}{\mathbb{E}\left[(v'(w(\theta, \cdot)))^{-1}\right]} \mathbb{E}\left[\frac{\hat{T}(w(\theta, \cdot))}{r(w(\theta, \cdot))}\right]$$

and the performance-pay effect $\hat{w}_{\mathrm{pp}}$ has mean zero and is given by

$$\begin{aligned}
\hat{w}_{\mathrm{pp}}(\theta, \eta) &= \frac{h'(a(\theta, \eta))}{v'(w(\theta, \eta))} \hat{a}(\theta, \eta) + \int_{\underline{\eta}}^{\eta} \frac{h''(a(\theta, x))}{v'(w(\theta, \eta))} \hat{a}(\theta, x) dx \\
&\quad - \mathbb{E}\left[\int_{\underline{\eta}}^{\eta'} \frac{h''(a(\theta, x))}{v'(w(\theta, \eta))} \hat{a}(\theta, x) dx\right] - \theta \hat{a}(\theta, \eta).
\end{aligned}$$

Proof. Consider a reform $\hat{R} = -\hat{T}$. Following the same steps as in the proof of Theorem 1, we can show that the change in wages is equal to

$$\hat{w}(\theta, \eta) = -\frac{\hat{R}(w(\theta, \eta))}{r(w(\theta, \eta))} + \frac{\hat{U}(\theta)}{v'(w(\theta, \eta))} + \int_{\underline{\eta}}^{\overline{\eta}} \frac{\partial w(\eta)}{\partial a(\eta')} \hat{a}(\eta') d\eta'. \tag{56}$$

Using prior results on $\frac{\partial w(\eta)}{\partial a(\eta')}$ from equation (54), we obtain

$$\begin{aligned}
\int_{\underline{\eta}}^{\overline{\eta}} \frac{\partial w(\eta)}{\partial a(\eta')} \hat{a}(\eta') d\eta' &= \frac{h'(a(\theta, \eta))}{v'(w(\theta, \eta))} \hat{a}(\theta, \eta) \\
&\quad + \int_{\underline{\eta}}^{\eta} \frac{h''(a(\theta, x))}{v'(w(\theta, \eta))} \hat{a}(\theta, x) dx - \mathbb{E}\left[\int_{\underline{\eta}}^{\eta'} \frac{h''(a(\theta, x))}{v'(w(\theta, \eta))} \hat{a}(\theta, x) dx\right].
\end{aligned}$$

Note that the expected value of the above expression is

$$\int_{\underline{\eta}}^{\overline{\eta}} \int_{\underline{\eta}}^{\overline{\eta}} \frac{\partial w(\theta,\eta)}{\partial a(\theta,\eta')} \hat{a}(\theta,\eta') d\eta' dF_\eta(\eta) = \int_{\underline{\eta}}^{\overline{\eta}} \int_{\underline{\eta}}^{\overline{\eta}} \frac{\partial w(\theta,\eta)}{\partial a(\theta,\eta')} dF_\eta(\eta) \hat{a}(\theta,\eta') d\eta'$$

$$= \int_{\underline{\eta}}^{\overline{\eta}} f_\eta(\eta') \theta \hat{a}(\theta,\eta') d\eta' = \mathbb{E}[\theta \hat{a}(\theta,\eta')]$$

where in the first equality we changed the order of integration and in the second equality we applied the first-order condition for effort (53). This implies that the performance-pay effect has mean zero. By the free-entry condition we have $\mathbb{E}[\theta \hat{a}(\eta')] = \mathbb{E}[\hat{w}(\eta')]$. Hence, the first two terms of (56) – the crowding-out effect – have mean zero. It follows that $\hat{U}(\theta) = \mathbb{E}\left[\frac{\hat{R}(w(\theta,\eta))}{r(w(\theta,\eta))}\right] \mathbb{E}\left[v'(w(\theta,\eta))^{-1}\right]^{-1}$.

$\square$

It follows from Theorem 3 that the crowding-out effect $\hat{w}_{\text{co}}(\theta,\eta)$ is exactly the same as in the simpler setting studied in the main body of the paper (equation (15)). The performance-pay effect $\hat{w}_{\text{pp}}(\theta,\eta)$ is more complex than in the simpler model (equation (16)). However, it is a natural extension of the expression obtained under the constant-effort assumption. Namely, the only substantial difference is that the term $\frac{h''(a(\theta))\hat{a}(\theta)}{v'(w(\theta,\eta))}\eta$ from (16), which measures the change in earnings necessary to elicit higher effort, is now replaced by the more general expression $\int_{\underline{\eta}}^{\eta} \frac{h''(a(\theta,\eta'))\hat{a}(\theta,\eta')}{v'(w(\theta,\eta))} d\eta'$. The interpretation of this term is analogous to its counterpart in the simpler model. The only added difficulty is that we now have to evaluate the change in the entire effort schedule $\hat{a}(\theta,\cdot)$ in response to the reform, rather than a scalar value $\hat{a}(\theta)$.

D   Proofs of Section 3

Proof of Theorem 2. Equation (14) implies that the excess burden of the reform $\hat{T}$ is equal to

$$\mathcal{EB}(T,\hat{T}) = -\int_\Theta \mathbb{E}\left[T'(w(\theta,\eta))\hat{w}_{\text{ex}}(\theta,\eta)\right] \mathrm{d}F(\theta)$$

$$-\int_\Theta \mathbb{E}\left[T'(w(\theta,\eta))\left(\frac{\hat{T}(w(\theta,\eta))}{r(w(\theta,\eta))} + \frac{\hat{U}(\theta)}{v'(w(\theta,\eta))}\right)\right] \mathrm{d}F(\theta)$$

$$-\int_\Theta \mathbb{E}\left[T'(w(\theta,\eta))\left(\frac{h'(a(\theta)) + h''(a(\theta))\eta}{v'(w(\theta,\eta))} - \frac{w(\theta,\eta)}{a(\theta)}\right)\right] \hat{a}(\theta) \mathrm{d}F(\theta),$$

where $\hat{U}(\theta)$ is given by (17). Since

$$\mathbb{E}\left[\frac{\hat{T}(w(\theta,\eta))}{r(w(\theta,\eta))} + \frac{\hat{U}(\theta)}{v'(w(\theta,\eta))}\right] = \mathbb{E}\left[\frac{h'(a(\theta)) + h''(a(\theta))\eta}{v'(w(\theta,\eta))} - \frac{w(\theta,\eta)}{a(\theta)}\right] = 0,$$

we can rewrite the excess burden $\mathcal{EB}(T,\hat{T})$ as

$$-\int_{\Theta} \mathbb{E}\left[T'(w(\theta,\eta))\,\hat{w}_{\text{ex}}(\theta,\eta)\right] dF(\theta)$$

$$-\int_{\Theta} \text{Cov}\left(T'(w(\theta,\eta)), \frac{\hat{T}(w(\theta,\eta))}{r(w(\theta,\eta))} + \frac{\hat{U}(\theta)}{v'(w(\theta,\eta))}\right) dF(\theta)$$

$$-\int_{\Theta} \text{Cov}\left(T'(w(\theta,\eta)), \left[\frac{h'(a(\theta)) + h''(a(\theta))\eta}{v'(w(\theta,\eta))} - \frac{w(\theta,\eta)}{a(\theta)}\right]\hat{a}(\theta)\right) dF(\theta).$$

This expression easily leads to equation (28).

Next, the welfare gain of the tax reform is given by

$$\mathcal{WG}(T,\hat{T}) = -\frac{1}{\lambda}\int_{\Theta} \alpha(\theta)\frac{1}{\mathbb{E}\left[(v'(w(\theta,\eta)))^{-1}\right]}\mathbb{E}\left[\frac{\hat{T}(w(\theta,\eta))}{r(w(\theta,\eta))}\right] dF(\theta)$$

$$= -\int_{\Theta} \mathbb{E}\left[\frac{1}{\lambda}\alpha(\theta)\frac{1}{\mathbb{E}\left[(v'(w(\theta,\cdot)))^{-1}\right]}\frac{\hat{T}(w(\theta,\eta))}{r(w(\theta,\eta))}\right] dF(\theta)$$

$$= -\int_{\Theta} \mathbb{E}\left[\frac{1}{\lambda}\alpha(\theta)\frac{\frac{1}{r(w(\theta,\eta))u'(R(w(\theta,\eta)))}}{\mathbb{E}\left[(v'(w(\theta,\cdot)))^{-1}\right]}u'(R(w(\theta,\eta)))\hat{T}(w(\theta,\eta))\right] dF(\theta).$$

This leads to equation (29).

When taxes are unrestricted, the marginal value of public funds λ can be obtained as follows. Consider a uniform lump-sum transfer, represented by the reform $\hat{T}^{**}(w) = -1$ for all w. Denoting by $\hat{a}^{**}(\theta)$ the effect of this reform on effort (via a pure income effect), it affects government revenue by $\hat{\mathcal{R}}(T,\hat{T}^{**})$ equal to

$$1 + \int_{\Theta} \mathbb{E}\left[T'(w(\theta,\eta))\left(-\frac{1}{r(w(\theta,\eta))} + \frac{\frac{1}{v'(w(\theta,\eta))}}{\mathbb{E}\left[\frac{1}{v'(w(\theta,\cdot))}\right]}\mathbb{E}\left[\frac{1}{r(w(\theta,\cdot))}\right]\right)\right] dF(\theta)$$

$$+ \int_{\Theta} \mathbb{E}\left[T'(w(\theta,\eta))\left(\frac{h'(a(\theta)) + h''(a(\theta))\eta}{v'(w(\theta,\eta))}\right)\right]\hat{a}^{**}(\theta)\,dF(\theta).$$

Consider now the reform in direction $\hat{T}^{**}(w) = -1$, normalized to reduce government

revenue by 1 dollar after all behavioral responses have been accounted for. This reform is represented by $\hat{T}^* = \frac{-1}{|\hat{\mathcal{R}}(T,\hat{T}^{**})|}$. Its effect on government revenue is $\hat{\mathcal{R}}(T,\hat{T}^*) = -1$ by construction, and its effect on social welfare is given by

$$\lambda \;\equiv\; \frac{\hat{\mathcal{W}}(T,\hat{T}^*)}{\left|\hat{\mathcal{R}}(T,\hat{T}^{**})\right|} \;=\; \frac{1}{\left|\hat{\mathcal{R}}(T,\hat{T}^{**})\right|} \int_\Theta \alpha\left(\theta\right) \frac{1}{\mathbb{E}\left[\frac{1}{v'(w(\theta,\eta))}\right]} \mathbb{E}\left[\frac{1}{r\left(w\left(\theta,\eta\right)\right)}\right] \mathrm{d}F\left(\theta\right).$$

Finally, we show that the optimal tax schedule must satisfy equation (27) for any tax reform $\hat{T}$. To do so, consider an arbitrary tax reform $\hat{T}$, normalized without loss of generality so that its mechanical effect is equal to 1 dollar, that is, $\int \mathbb{E}\left[\hat{T}\left(w\left(\theta,\eta\right)\right)\right]\mathrm{d}F\left(\theta\right) = 1$. Denote its effect on government revenue by $\hat{\mathcal{R}}(T,\hat{T})$ and its effect on social welfare by $\hat{\mathcal{W}}(T,\hat{T})$. Redistribute any tax revenue gain (or levy any tax revenue loss) from this reform via the reform $\hat{T}^*$ described in the previous paragraph, that is, a uniform lump-sum transfer that reduces government budget by 1 dollar. The tax reform

$$\hat{T} + \hat{\mathcal{R}}(T,\hat{T})\hat{T}^* \;\equiv\; \hat{T} + \hat{\mathcal{R}}(T,\hat{T})\frac{-1}{\left|\hat{\mathcal{R}}(T,-1)\right|}$$

is, by construction, budget-neutral. Its effect on social welfare is given by

$$\hat{\mathcal{W}}(T,\hat{T}) + \hat{\mathcal{R}}(T,\hat{T})\frac{\hat{\mathcal{W}}(T,\hat{T}^*)}{\left|\hat{\mathcal{R}}(T,-1)\right|} \;=\; \hat{\mathcal{W}}(T,\hat{T}) + \lambda\hat{\mathcal{R}}(T,\hat{T}).$$

Of course, the marginal value of public funds λ is the Lagrange multiplier on the government budget constraint. Now, the optimal tax schedule is such that every budget-neutral tax reform has a zero effect on social welfare (see, for instance, Luenberger (1997)), that is, for all $\hat{T}$,

$$\hat{\mathcal{W}}(T,\hat{T}) + \lambda\hat{\mathcal{R}}(T,\hat{T}) \;=\; 0.$$

But recall that $\frac{1}{\lambda}\hat{\mathcal{W}}(T,\hat{T}) = \mathcal{WG}(T,\hat{T})$ and $\hat{\mathcal{R}}(T,\hat{T}) = 1 - \mathcal{EB}(T,\hat{T})$, by definition of the welfare gains and the excess burden. This immediately implies the characterization (27). $\qquad\square$

E   Proofs of Section 4

Proof of Corollary 1. Suppose that the tax schedule is CRP, so that $R(w) = \frac{1-\tau}{1-p}w^{1-p}$. Equation (48) then implies that in order to induce agents with ability θ to choose the same effort a regardless of their noise realization η, the earnings contract must satisfy:

$$\log\left(w\left(\theta,\eta\right)\right) = \frac{a^{\frac{1}{\varepsilon}}}{1-p}\eta - \frac{1}{1-p}\log\left(\frac{1-\tau}{1-p}\right) + \frac{k}{1-p}, \tag{57}$$

for some $k \in \mathbb{R}$. Thus, log-earnings are linear in the performance shock $\eta = \frac{y}{\theta} - a$ that the firm infers upon observing realized output y. Imposing that the agent's participation constraint holds with equality pins down the value of k as a function of $U(\theta)$. Namely, equation (49) implies:

$$k = U\left(\theta\right) + \frac{1}{1+\frac{1}{\varepsilon}}a^{1+\frac{1}{\varepsilon}}$$

and hence

$$\log\left(w\left(\theta,\eta\right)\right) = \frac{a^{\frac{1}{\varepsilon}}}{1-p}\eta + \frac{1}{1-p}\frac{1}{1+\frac{1}{\varepsilon}}a^{1+\frac{1}{\varepsilon}} - \frac{1}{1-p}\log\left(\frac{1-\tau}{1-p}\right) + \frac{U\left(\theta\right)}{1-p}. \tag{58}$$

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

$$\log\left(w\left(\theta,\eta\right)\right) = \log\left(\theta a\right) + \frac{a^{\frac{1}{\varepsilon}}}{1-p}\eta - \frac{1}{2}\left(\frac{a^{\frac{1}{\varepsilon}}}{1-p}\right)^2\sigma_\eta^2. \tag{59}$$

Define the sensitivity of the before-tax and after-tax wages to output in the optimal contract by the semi-elasticities $\psi\left(\theta,\eta\right) \equiv \frac{1}{w(\theta,\eta)}\frac{\partial w(\theta,\eta)}{\partial\eta}$ and $\psi^c\left(\theta,\eta\right) \equiv \frac{1}{R(w(\theta,\eta))}\frac{\partial R(w(\theta,\eta))}{\partial\eta}$, respectively. We have $\psi\left(\theta,\eta\right) = \frac{a^{1/\varepsilon}}{1-p}$ and $\psi^c\left(\theta,\eta\right) = a^{1/\varepsilon}$. Both $\psi\left(\theta,\eta\right)$ and $\psi^c\left(\theta,\eta\right)$ 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 $v'(w) = \frac{r(w)}{R(w)} = \frac{1-p}{w}$, the firm's first-order condition reads

$$\theta = \mathbb{E}\left[\frac{h'(a)}{v'(w(\theta,\eta))} + \frac{h''(a)}{v'(w(\theta,\eta))}\eta\right]$$

$$= \frac{a^{\frac{1}{\varepsilon}}}{1-p}\mathbb{E}\left[w(\theta,\eta)\right] + \frac{1}{\varepsilon}\frac{a^{\frac{1}{\varepsilon}-1}}{1-p}\mathbb{E}\left[w(\theta,\eta)\eta\right].$$

We have

$$\mathbb{E}\left[w(\theta,\eta)\right] = \mathbb{E}\left[e^{\frac{a^{\frac{1}{\varepsilon}}}{1-p}\eta}\right] e^{\frac{1}{1-p}\frac{1}{1+\frac{1}{\varepsilon}}a^{1+\frac{1}{\varepsilon}} - \frac{1}{1-p}\log\left(\frac{1-\tau}{1-p}\right) + \frac{U(\theta)}{1-p}}$$

$$= e^{\frac{1}{2}\frac{a^{\frac{2}{\varepsilon}}}{(1-p)^2}\sigma_\eta^2} e^{\frac{1}{1-p}\frac{1}{1+\frac{1}{\varepsilon}}a^{1+\frac{1}{\varepsilon}} - \frac{1}{1-p}\log\left(\frac{1-\tau}{1-p}\right) + \frac{U(\theta)}{1-p}}.$$

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

$$\mathbb{E}\left[w(\theta,\eta)\eta\right] = \mathbb{E}\left[\eta e^{\frac{a^{\frac{1}{\varepsilon}}}{1-p}\eta}\right] e^{\frac{1}{1-p}\frac{1}{1+\frac{1}{\varepsilon}}a^{1+\frac{1}{\varepsilon}} - \frac{1}{1-p}\log\left(\frac{1-\tau}{1-p}\right) + \frac{U(\theta)}{1-p}}$$

$$= \frac{a^{\frac{1}{\varepsilon}}}{1-p}\sigma_\eta^2 e^{\frac{1}{2}\frac{a^{\frac{2}{\varepsilon}}}{(1-p)^2}\sigma_\eta^2} e^{\frac{1}{1-p}\frac{1}{1+\frac{1}{\varepsilon}}a^{1+\frac{1}{\varepsilon}} - \frac{1}{1-p}\log\left(\frac{1-\tau}{1-p}\right) + \frac{U(\theta)}{1-p}}.$$

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

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

and hence

$$\frac{a^{1+\frac{1}{\varepsilon}}}{1-p} + \frac{1}{\varepsilon}\frac{a^{\frac{2}{\varepsilon}}}{(1-p)^2}\sigma_\eta^2 = \theta a e^{-\frac{1}{1-p}\frac{1}{1+\frac{1}{\varepsilon}}a^{1+\frac{1}{\varepsilon}} - \frac{1}{2}\frac{a^{\frac{2}{\varepsilon}}}{(1-p)^2}\sigma_\eta^2 + \frac{1}{1-p}\log\left(\frac{1-\tau}{1-p}\right) - \frac{U(\theta)}{1-p}}$$

Now use the free-entry condition: equation (5) and the expression derived above for

$\mathbb{E}\left[w\left(y\mid\theta\right)\right]$ leads to

$$e^{\frac{1}{1-p}\frac{1}{1+\frac{1}{\varepsilon}}a^{1+\frac{1}{\varepsilon}}+\frac{1}{2}\frac{a^{\frac{2}{\varepsilon}}}{(1-p)^2}\sigma_\eta^2-\frac{1}{1-p}\log\left(\frac{1-\tau}{1-p}\right)+\frac{U(\theta)}{1-p}} = \theta a. \tag{60}$$

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

$$a^{1+\frac{1}{\varepsilon}} + \frac{1}{\varepsilon}\frac{a^{\frac{2}{\varepsilon}}}{1-p}\sigma_\eta^2 = 1-p. \tag{61}$$

Using the definition $\psi \equiv \frac{a^{\frac{1}{\varepsilon}}}{1-p}$ for the pass-through easily leads to (31). Note that if $\varepsilon = 1$, we obtain optimal effort in closed form:

$$a = \left(\frac{1}{1-p} + \frac{\sigma^2}{(1-p)^2}\right)^{-1/2}. \tag{62}$$

Finally, taking logs in equation (60) and defining $\psi \equiv \frac{a^{\frac{1}{\varepsilon}}}{1-p}$ easily leads to (32). $\qquad\square$

Proof of Corollary 2. Consider a tax reform that marginally raises the rate of progressivity p by a small amount $\delta \to 0$. The direction $\hat{T}$ of this tax reform satisfies

$$\left(w - \frac{1-\tau}{1-p-\delta}w^{1-p-\delta}\right) - \left(w - \frac{1-\tau}{1-p}w^{1-p}\right) = \delta\hat{T}(w) + o(\delta).$$

This leads to the representation (34).

Differentiating equation (61) with respect to $(1-p)$ leads to

$$\left[\left(1+\frac{1}{\varepsilon}\right)a^{\frac{1}{\varepsilon}} + \frac{2\sigma_\eta^2}{(1-p)\varepsilon^2}a^{\frac{2}{\varepsilon}-1}\right]\frac{\partial a}{\partial(1-p)} - \frac{\sigma_\eta^2}{(1-p)^2\varepsilon}a^{\frac{2}{\varepsilon}} = 1,$$

and hence

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

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

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

71

We conclude by expressing this elasticity in terms of the pass-through elasticities. We have $\psi = \frac{a^{\frac{1}{\varepsilon}}}{1-p}$ and $\varepsilon_{\psi,a} = \frac{1}{\varepsilon}$. We can thus write

$$\varepsilon_{a,1-p} = \frac{a^{\frac{1}{\varepsilon}+1} + 2\left(1-p\right)\varepsilon_{\psi,a}\psi^2\sigma_\eta^2}{\left(1+\frac{1}{\varepsilon}\right)a^{\frac{1}{\varepsilon}+1} + \frac{2}{\varepsilon}\left(1-p\right)\varepsilon_{\psi,a}\psi^2\sigma_\eta^2}.$$

But the first-order condition for labor effort reads

$$a^{1+1/\varepsilon} = \left(1-p\right)\left(1-\varepsilon_{\psi,a}\psi^2\sigma_\eta^2\right).$$

Substituting into the previous equation and rearranging terms leads to

$$\varepsilon_{a,1-p} = \frac{1 + \varepsilon_{\psi,a}\psi^2\sigma_\eta^2}{\left(1+\frac{1}{\varepsilon}\right) + \left(\frac{1}{\varepsilon}-1\right)\varepsilon_{\psi,a}\psi^2\sigma_\eta^2}.$$

This easily yields equation (35).

Expression (36) for $\hat{w}_{\text{ex}}\left(\theta,\eta\right)$ is immediate since $\frac{\hat{a}}{a} = \frac{1}{1-p}\varepsilon_{a,1-p}$ by definition. To obtain the performance-pay effect (38), we show that for any (not necessarily CRP) tax reform $\hat{T}$,

$$\hat{w}_{\text{pp}}\left(\theta,\eta\right) = \varepsilon_{\psi,a}\left(\psi\eta - \psi^2\sigma_\eta^2\right)\hat{w}_{\text{ex}}\left(y \mid \theta\right). \tag{63}$$

This equation implies that $\hat{w}_{\text{pp}}\left(\theta,\eta\right)$ has mean zero, but is dispersed around the mean whenever $\varepsilon_{\psi,a} > 0$, since the map $\eta \mapsto \psi\eta - \psi^2\sigma_\eta^2$ is strictly increasing. To prove this equation, note that

$$\frac{h'\left(a\left(\theta\right)\right) + h''\left(a\left(\theta\right)\right)\eta}{v'\left(w\left(\theta,\eta\right)\right)}\hat{a}\left(\theta\right) = \frac{R\left(w\left(\theta,\eta\right)\right)}{r\left(w\left(\theta,\eta\right)\right)}\left(a^{1+\frac{1}{\varepsilon}} + \frac{1}{\varepsilon}a^{\frac{1}{\varepsilon}}\eta\right)\frac{\hat{a}}{a}$$
$$= \frac{1}{1-p}w\left(\theta,\eta\right)\left[a^{1+\frac{1}{\varepsilon}} + \left(1-p\right)\varepsilon_{\psi,a}\psi\eta\right]\frac{\hat{a}}{a}$$
$$= \frac{1}{1-p}w\left(\theta,\eta\right)\left[\left(1-p\right)\left(1-\varepsilon_{\psi,a}\psi^2\sigma_\eta^2\right) + \left(1-p\right)\varepsilon_{\psi,a}\psi\eta\right]\frac{\hat{a}}{a}$$
$$= \left[1 + \varepsilon_{\psi,a}\left(\psi\eta - \psi^2\sigma_\eta^2\right)\right]w\left(\theta,\eta\right)\frac{\hat{a}}{a},$$

where the third equality uses the first-order condition for labor effort.

Next, we compute the crowding-out effect $\hat{w}_{co}(\theta, \eta)$. We have

$$
\begin{aligned}
\frac{\hat{T}(w(\theta,\eta))}{r(w(\theta,\eta))} &= \frac{1}{(1-\tau)(w(\theta,\eta))^{-p}}\left(\log(w(\theta,\eta)) - \frac{1}{1-p}\right)\frac{1-\tau}{1-p}(w(\theta,\eta))^{1-p} \\
&= \left(\ln(\theta a) + \psi\eta - \frac{1}{2}\psi^2\sigma_\eta^2 - \frac{1}{1-p}\right)\frac{1}{1-p}w(\theta,\eta),
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\hat{U}(\theta)}{v'(w(\theta,\eta))} &= -\frac{\frac{1}{v'(w(\theta,\eta))}}{\mathbb{E}\left[\frac{1}{v'(w(\theta,\cdot))}\right]}\mathbb{E}\left[\frac{\hat{T}(w(\theta,\cdot))}{r(w(\theta,\cdot))}\right] \\
&= -\frac{\frac{1}{1-p}w(\theta,\eta)}{\mathbb{E}\left[\frac{1}{1-p}w(\theta,\cdot)\right]}\mathbb{E}\left[\left(\log(w(\theta,\cdot)) - \frac{1}{1-p}\right)\frac{1}{1-p}w(\theta,\cdot)\right] \\
&= -\frac{1}{1-p}w(\theta,\eta)\left[\ln(\theta a) + \frac{1}{2}\psi^2\sigma_\eta^2 - \frac{1}{1-p}\right].
\end{aligned}
$$

Summing these expressions yields equation (37).

Summing all the effects, we can easily verify that the incidence of the reform is given by

$$
\frac{\partial \log w(\theta,\eta)}{\partial(1-p)} = \varepsilon_{a,1-p} + (\varepsilon_{\psi,a}\varepsilon_{a,1-p} + \varepsilon_{\psi,1-p})\left(\psi\eta - \psi^2\sigma_\eta^2\right),
$$

which is the expression as we would obtain by directly differentiating $\log w(\theta,\eta) = \log(\theta a) + \psi\eta - \frac{1}{2}\psi^2\sigma_\eta^2$.

Note that the earnings adjustment $\hat{w}_i(\theta,\eta)$ contributes to raising the sensitivity of log-earnings to performance shocks (pass-through function) iff

$$
\frac{\partial}{\partial\eta}\log(w(\theta,\eta) + \delta\hat{w}_i(\theta,\eta)) - \frac{\partial}{\partial\eta}\log(w(\theta,\eta)) > 0.
$$

For δ close enough to zero this inequality is equivalent to

$$
\frac{\partial\hat{w}(\theta,\eta)}{\partial\eta} > \hat{w}_i(\theta,\eta)\frac{\partial\log(w(\theta,\eta))}{\partial\eta}.
$$

73

The expressions derived above imply

$$
\begin{aligned}
\frac{\partial \hat{w}_{\mathrm{pp}}\left(\theta, \eta\right)}{\partial \eta} & = \frac{\hat{w}_{\mathrm{pp}}\left(\theta, \eta\right)}{\hat{w}_{\mathrm{ex}}\left(\theta, \eta\right)} \frac{\partial \hat{w}_{\mathrm{ex}}\left(\theta, \eta\right)}{\partial \eta} + \hat{w}_{\mathrm{pp}}\left(\theta, \eta\right) \frac{\psi}{\psi \eta - \psi^2 \sigma_\eta^2} \\
& = \hat{w}_{\mathrm{pp}}\left(\theta, \eta\right) \frac{\partial \log\left(w\left(\theta, \eta\right)\right)}{\partial \eta} + \psi \varepsilon_{\psi, a} w\left(\theta, \eta\right) \frac{\hat{a}}{a}.
\end{aligned}
$$

Thus, since $\hat{a} < 0$, we obtain $\frac{\partial \hat{w}_{\mathrm{pp}}(\theta, \eta)}{\partial \eta} < \hat{w}_{\mathrm{pp}}\left(y \mid \theta\right) \frac{\partial \log(w(\theta, \eta))}{\partial \eta}$ and the reform lowers the sensitivity of log-earnings to output. Analogously, the crowding-out effect lowers raises the sensitivity of log-earnings since $\varepsilon_{\psi, 1-p} < 0$.

Finally, we derive equation (40). We have

$$
\begin{aligned}
\frac{\hat{w}_{\mathrm{co}}\left(\theta, \eta\right)}{\hat{w}_{\mathrm{pp}}\left(\theta, \eta\right)} & = \frac{-\frac{1}{1-p} \varepsilon_{\psi, 1-p} \left(\psi \eta - \psi^2 \sigma_\eta^2\right) w\left(\theta, \eta\right)}{-\frac{1}{1-p} \varepsilon_{\psi, a} \varepsilon_{a, 1-p} \left(\psi \eta - \psi^2 \sigma_\eta^2\right) w\left(\theta, \eta\right)} \\
& = \frac{\varepsilon_{\psi, 1-p}}{\varepsilon_{\psi, a} \varepsilon_{a, 1-p}} = -\frac{\varepsilon}{\varepsilon_{a, 1-p}} \\
& \underset{o\left(\psi^2 \sigma_\eta^2\right)}{=} -\left(1 + \varepsilon\right) \left(1 + \frac{1-\varepsilon}{1+\varepsilon} \frac{1}{\varepsilon} \psi^2 \sigma_\eta^2 - \frac{1}{\varepsilon} \psi^2 \sigma_\eta^2\right) \\
& = 2\psi^2 \sigma_\eta^2 - \varepsilon - 1.
\end{aligned}
$$

where the second to last equality uses equation (35). $\qquad \square$

Proof of Corollary 3. Suppose that Assumption 2 holds, and that ability types are lognormally distributed, that is, $\log \theta \sim \mathcal{N}\left(\mu_\theta, \sigma_\theta^2\right)$. We substitute formula (34) for the tax reform $\hat{T}$ in equations (28) and (29) to compute each term of the excess burden and the welfare gains of marginally raising progressivity. The algebra is straightforward but tedious. It is available upon request and we only summarize our results here. The mechanical effect of the progressive tax reform is equal to

$$
\int_\Theta \mathbb{E}\left[\hat{T}\left(w\left(\theta, \eta\right)\right)\right] \mathrm{d}F\left(\theta\right) = \left[\mu_\theta + \ln a - \frac{1}{1-p} + \left(1-p\right) \sigma_\theta^2 + \left(\frac{1}{2} - p\right) \psi^2 \sigma^2\right] C
$$

where C denotes aggregate consumption and is given by

$$
C \equiv \int_\Theta \mathbb{E}\left[R\left(w\left(\theta, \eta\right)\right)\right] \mathrm{d}F\left(\theta\right) = \frac{1-\tau}{1-p} a^{1-p} e^{(1-p)\mu_\theta + \frac{1}{2}(1-p)^2 \sigma_\theta^2} e^{-\frac{1}{2} p(1-p)\psi^2 \sigma_\eta^2}.
$$

74

Let Y denotes aggregate output, given by

$$Y \equiv \int_\Theta (\theta a)\, dF(\theta) = a e^{\mu_\theta + \frac{\sigma_\theta^2}{2}}.$$

Note that in our baseline economy with government expenditures G, we have $Y = C + G$. The excess burden of the tax reform in the model with exogenous risk (that is, the first integral in (28)) is given by

$$\int_\Theta \mathbb{E}\left[T'(w(\theta,\eta))\, w(\theta,\eta)\, \frac{\hat{a}(\theta)}{a(\theta)}\right] dF(\theta) = -\frac{1}{1-p}(Y - (1-p)C)\varepsilon_{a,1-p}.$$

The fiscal externality due to the first component of crowding-out is given b

$$\int_\Theta \mathrm{Cov}\left(T'(w(\theta,\eta)), \frac{\hat{T}(w(\theta,\eta))}{r(w(\theta,\eta))}\right) dF(\theta)$$
$$= \left(e^{p\psi^2\sigma_\eta^2} - 1\right)\left[\mu_\theta + \ln a - \frac{1}{1-p} + (1-p)\sigma_\theta^2\right]C + \left(e^{p\psi^2\sigma_\eta^2} - 1 + 2p\right)\frac{1}{2}\psi^2\sigma_\eta^2 C.$$

The fiscal externality due to the second element of crowding-out is given by

$$\int_\Theta \mathrm{Cov}\left(T'(w(\theta,\eta)), \frac{\hat{U}(\theta)}{v'(w(\theta,\eta))}\right) dF(\theta)$$
$$= -\left(e^{p\psi^2\sigma_\eta^2} - 1\right)\left[\mu_\theta + \ln a - \frac{1}{1-p} + (1-p)\sigma_\theta^2\right]C - \left(e^{p\psi^2\sigma_\eta^2} - 1\right)\frac{1}{2}\psi^2\sigma_\eta^2 C.$$

Thus the total fiscal externality from the crowding-out effect is equal to $p\psi^2\sigma_\eta^2 C$. The fiscal externality due to the performance-pay effect is given by

$$\int_\Theta \mathrm{Cov}\left(T'(w(\theta,\eta)), \frac{h'(a(\theta)) + h''(a(\theta))\eta}{v'(w(\theta,\eta))}\hat{a}(\theta) - \hat{w}_{\mathrm{ex}}(\theta,\eta)\right) dF(\theta)$$
$$= -p\frac{1}{\varepsilon}\psi^2\sigma_\eta^2\varepsilon_{a,1-p}C.$$

Suppose that the social welfare weights are given by $\alpha(\theta) = \frac{e^{-\alpha\log\theta}}{\int e^{-\alpha\log\theta'}\,dF(\theta')}$ for all θ,

for some $\alpha \geq 0$. Then the effect of the tax reform on social welfare is given by

$$-\int_{\Theta} \mathbb{E}\left[\frac{\left(v'\left(w\left(\theta, \eta\right)\right)\right)^{-1}}{\mathbb{E}\left[\left(v'\left(w\left(\theta, \cdot\right)\right)\right)^{-1}\right]} \alpha\left(\theta\right) u'\left(R\left(w\left(\theta, \eta\right)\right)\right) \hat{T}\left(w\left(\theta, \eta\right)\right)\right] \mathrm{d}F\left(\theta\right)$$
$$= -\mu_{\theta} - \ln a + \frac{1}{1-p} - \frac{1}{2}\psi^2 \sigma_{\eta}^2 + \alpha \sigma_{\theta}^2.$$

Third, we compute the marginal value of public funds λ in the loglinear model, when the tax code is restricted to the CRP class. To do so, first consider a reform of the parameter τ, represented by formula (64). By definition, the parameter lambda is the effect on social welfare caused by a tax reform in this direction, normalized to raise government revenue by 1 dollar. The mechanical effect of the (non-normalized) reform (64) is equal to

$$\int_{\Theta} \mathbb{E}\left[\hat{T}\left(w\left(\theta, \eta\right)\right)\right] \mathrm{d}F\left(\theta\right) = \frac{C}{1-\tau}.$$

Since the elasticity of labor effort is $\varepsilon_{a,1-\tau} = 0$, the standard excess burden and the fiscal externality caused by the performance-pay effect are both equal to zero,

$$\int_{\Theta} \mathbb{E}\left[T'\left(w\left(\theta, \eta\right)\right) \hat{w}_{\mathrm{ex}}\left(\theta, \eta\right)\right] \mathrm{d}F\left(\theta\right)$$
$$= \int_{\Theta} \mathrm{Cov}\left(T'\left(w\left(\theta, \eta\right)\right), \hat{w}_{\mathrm{pp}}\left(\theta, \eta\right)\right) \mathrm{d}F\left(\theta\right) = 0.$$

The fiscal externalities caused by the two elements of crowding-out are given by

$$\int_{\Theta} \mathrm{Cov}\left(T'\left(w\left(\theta, \eta\right)\right), \frac{\hat{T}\left(w\left(\theta, \eta\right)\right)}{r\left(w\left(\theta, \eta\right)\right)}\right) \mathrm{d}F\left(\theta\right)$$
$$= -\int_{\Theta} \mathrm{Cov}\left(T'\left(w\left(\theta, \eta\right)\right), \frac{\hat{U}\left(\theta\right)}{v'\left(w\left(\theta, \eta\right)\right)}\right) \mathrm{d}F\left(\theta\right) = \left(e^{p\psi^2\sigma_{\eta}^2} - 1\right)\frac{C}{1-\tau}.$$

The welfare effect of the tax reform is given by

$$-\int_{\Theta} \mathbb{E}\left[\frac{\left(v'\left(w\left(\theta, \eta\right)\right)\right)^{-1}}{\mathbb{E}\left[\left(v'\left(w\left(\theta, \cdot\right)\right)\right)^{-1}\right]} \alpha\left(\theta\right) u'\left(R\left(w\left(\theta, \eta\right)\right)\right) \hat{T}\left(w\left(\theta, \eta\right)\right)\right] \mathrm{d}F\left(\theta\right) = -\frac{1}{1-\tau}.$$

Now, normalize the tax reform of the tax rate τ so that it delivers \$1 of revenue.

76

Since the sum of all the fiscal externalities is zero, the increase in government revenue of the tax reform $\hat{T} = \frac{1}{1-p} w^{1-p}$ is simply of mechanical effect $\frac{C}{1-\tau}$. Thus, we consider the normalized tax reform

$$\hat{T}^* = \frac{1-\tau}{C} \times \hat{T} = \frac{1}{C} \times \frac{1-\tau}{1-p} (w(\theta, \eta))^{1-p}.$$

The welfare impact of distributing an additional dollar of tax revenue via a reduction of the parameter τ is therefore equal to the welfare effect of the reform $-\hat{T}^*$, which is equal to

$$\lambda = \frac{1-\tau}{C} \times \frac{1}{1-\tau} = \frac{1}{C}.$$

This gives the marginal value of public funds λ in this setting and concludes the proof. $\qquad\square$

Proof of Proposition 4. We give two proofs of this result. First, we apply formula (27) using the explicit expressions of each term derived in the proof of Proposition 3 above. We must have, letting $\alpha(\theta) = 1$ for all θ,

$$
\begin{aligned}
0 &= \left[\mu_\theta + \ln a - \frac{1}{1-p} + (1-p)\sigma_\theta^2 + \left(\frac{1}{2} - p\right) \psi^2 \sigma_\eta^2 \right] C \\
&\quad - \left[\mu_\theta + \ln a - \frac{1}{1-p} + \frac{1}{2}\psi^2 \sigma_\eta^2 \right] C \\
&\quad - \frac{1}{1-p}(Y - (1-p)C)\varepsilon_{a,1-p} + p\psi^2\sigma_\eta^2 C - \left[p\frac{1}{\varepsilon}\psi^2\sigma_\eta^2 \right] \varepsilon_{a,1-p} C \\
&= \left[(1-p)\sigma_\theta^2 + \left(-\frac{p}{1-p}\frac{Y}{C} - p\frac{1}{\varepsilon}\psi^2\sigma_\eta^2 \right) \varepsilon_{a,1-p} \right] C.
\end{aligned}
$$

Since government expenditures are equal to G, we have

$$\frac{Y}{C} = \frac{Y}{Y-G} = \frac{1}{1-g}$$

77

where $g \equiv G/Y$. We thus obtain

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

which easily yields the result.

The second proof consists of directly calculating the optimal rate of progressivity in the loglinear model by equating to zero the derivative of social welfare in this environment. To do so, recall that the earnings schedule of agents with ability θ can be written as

$$
\log \left(w\left(\theta, \eta \right) \right) = \log \left(\theta a \right) + \psi \eta - \frac{1}{2} \left(\psi \sigma_\eta \right)^2
$$

and their expected utility as

$$
U\left(\theta \right) = \log \left(\frac{1-\tau}{1-p} \right) + (1-p) \log \left(\theta a \right) - \frac{1}{2} (1-p) \left(\psi \sigma_\eta \right)^2 - h\left(a \right).
$$

Utilitarian social welfare is therefore equal to

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

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

$$
0 = \frac{\partial U\left(\theta \right)}{\partial a} = (1-p) \frac{1}{a} - (1-p) \psi \sigma_\eta^2 \frac{\partial \psi}{\partial a} - h'\left(a \right).
$$

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

$$
\int_\Theta \mathbb{E}\left[R\left(w\left(\theta, \eta \right) \right) \right] f\left(\theta \right) d\theta = a e^{\mu_\theta + \frac{\sigma_\theta^2}{2}} - \frac{1-\tau}{1-p} e^{-\frac{p a^{\frac{2}{\varepsilon}} \sigma_\eta^2}{2(1-p)}} a^{1-p} e^{(1-p)\mu_\theta + (1-p)^2 \frac{\sigma_\theta^2}{2}}.
$$

Budget balance thus requires

$$\frac{1-\tau}{1-p} = \frac{ae^{\mu_\theta+\frac{\sigma_\theta^2}{2}} - G}{e^{-\frac{pa^{\frac{2}{\varepsilon}}\sigma_\eta^2}{2(1-p)}}a^{1-p}e^{(1-p)\mu_\theta+(1-p)^2\frac{\sigma_\theta^2}{2}}} = \frac{(1-g)\,ae^{\mu_\theta+\frac{\sigma_\theta^2}{2}}}{e^{-\frac{pa^{\frac{2}{\varepsilon}}\sigma_\eta^2}{2(1-p)}}a^{1-p}e^{(1-p)\mu_\theta+(1-p)^2\frac{\sigma_\theta^2}{2}}}.$$

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

$$0 = \mu_\theta + \log a + (1-p)\frac{1}{a}\frac{\partial a}{\partial(1-p)} - h'(a)\frac{\partial a}{\partial(1-p)} - \frac{\psi^2\sigma_\eta^2}{2}$$
$$- (1-p)\,\psi\sigma_\eta^2\left[\frac{\partial\psi}{\partial(1-p)} + \frac{\partial\psi}{\partial a}\frac{\partial a}{\partial(1-p)}\right] + \frac{\partial\log\left(\frac{1-\tau}{1-p}\right)}{\partial(1-p)},$$

with

$$\frac{\partial\log\left(\frac{1-\tau}{1-p}\right)}{\partial(1-p)} = \frac{g}{1-g}\frac{\partial\log a}{\partial(1-p)} - \mu_\theta - (1-p)\sigma_\theta^2 - \log a + p\frac{1}{a}\frac{\partial a}{\partial(1-p)}$$
$$- \left(\frac{1}{2}-p\right)\psi^2\sigma_\eta^2 + p(1-p)\,\psi\sigma_\eta^2\left[\frac{\partial\psi}{\partial(1-p)} + \frac{\partial\psi}{\partial a}\frac{\partial a}{\partial(1-p)}\right].$$

We therefore obtain

$$0 = \left[(1-p)\frac{1}{a} - h'(a) - (1-p)\,\psi\sigma_\eta^2\frac{\partial\psi}{\partial a}\right]\frac{\partial a}{\partial(1-p)} + p\frac{1}{a}\frac{\partial a}{\partial(1-p)} + \frac{g}{1-g}\frac{\partial\log a}{\partial(1-p)}$$
$$- (1-p)\sigma_\theta^2 - (1-p)\,\psi^2\sigma_\eta^2 - (1-p)^2\,\psi\sigma_\eta^2\frac{\partial\psi}{\partial(1-p)} + p(1-p)\,\psi\sigma_\eta^2\frac{\partial\psi}{\partial a}\frac{\partial a}{\partial(1-p)}.$$

Using the first-order condition for effort leads to

$$0 = \frac{1}{1-p}\left[p + \frac{g}{1-g}\right]\varepsilon_{a,1-p} + p\psi^2\sigma_\eta^2\varepsilon_{\psi,a}\varepsilon_{a,1-p}$$
$$- (1-p)\left[\sigma_\theta^2 + \psi^2\sigma_\eta^2\right] - (1-p)\,\psi^2\sigma_\eta^2\varepsilon_{\psi,1-p}.$$

Rearranging this equation leads to the result. $\qquad\square$

Further Examples of Tax Reforms. We now study the incidence of additional examples of tax reforms.

Lump-Sum Tax Increase on High Incomes. We focus on the top earners, whose incomes are located in the highest bracket characterized by a constant tax rate τ_{top}

and an income threshold w_{top}. Thus, the baseline tax schedule T is locally affine, that is, $T(w) = T(w_{\text{top}}) + \tau_{\text{top}}(w - w_{\text{top}})$ for all $w > w_{\text{top}}$. Moreover, we assume that $w(y \mid \theta) > w_{\text{top}}$ for all y. A uniform lump-sum increase of the income tax liabilities of these agents is represented by the tax reform $\hat{T}(w) = 1$ for all $w > w_{\text{top}}$. (Equivalently, $\hat{T}$ can be any positive constant.) Indeed, the perturbed tax schedule is then given by $T + \delta\hat{T} = T + \delta$. Thus, the tax function is shifted up by the constant $\delta < 0$. The value of the Gateaux derivative $\hat{\Psi}(T, \hat{T})$ then gives the first-order effect of this lump-sum transfer on the functional Ψ as its size δ becomes small.

Applying formulas (15) and (16) leads to the following results. The earnings adjustment caused by crowding-out is equal to

$$\hat{w}_{\text{co}}(\theta, \eta) = \frac{1}{1 - \tau}\left(1 - \frac{(v'(w(\theta, \eta)))^{-1}}{\mathbb{E}\left[(v'(w(\theta, \cdot)))^{-1}\right]}\right).$$

Intuitively, a uniform lump-sum tax increase can be fully absorbed by the firm via a counteracting lump-sum increase in earnings, without any change in private insurance. Thus, the first element of crowding-out is constant. On the other hand, the second element of crowding-out is decreasing in the performance shock. This is because the tax increase reduces expected utility $\hat{U}(\theta)$. To preserve incentive compatibility, this is achieved by reducing pre-tax earnings by larger amounts for higher-income workers, since their marginal utility is smaller. As a result, $\frac{\partial}{\partial\eta}\hat{w}_{\text{co}}(\theta, \eta) < 0$, so that the crowding-out effect reduces the sensitivity of pre-tax earnings to performance. On the other hand, the sum of the standard labor supply effect $\hat{w}_{\text{ex}}(\theta, \eta)$ and the performance-pay effect $\hat{w}_{\text{pp}}(\theta, \eta)$ is increasing, so that the labor supply responses raise the performance sensitivity of the contract. Indeed, a lump-sum tax increase creates a pure income effect and hence raises optimal effort, that is, $\hat{a}(\theta) > 0$. Implementing this higher effort level requires an increase in the sensitivity of earnings to performance shocks. Overall, a uniform tax increase can lead to either a spread or a contraction in the pre-tax earnings schedule, depending on the size of the income effect for high-income earners.

Marginal Tax Rate Increase on High Incomes. Focusing again on the highest income tax bracket, an increase in the top marginal tax rate is represented by the tax reform $\hat{T}(w) = w - w_{\text{top}}$ for all $w > w_{\text{top}}$. Indeed, the perturbed tax payments are then given by $T(w) + \delta\hat{T}(w) = T(w) + \delta(w - w_{\text{top}})$, and the marginal tax rates are

perturbed by the constant $\delta \hat{T}'(w) = \delta$ for $w \geq w_{\text{top}}$. We obtain the following results. Assuming that the utility function is logarithmic, we find that the crowding-out effect is equal to

$$\hat{w}_{\text{co}}(\theta, \eta) = \frac{1}{(1 - \tau) \Pi} (w(\theta, \eta) - \mathbb{E}[w(\theta, \eta)]),$$

where $\Pi = \mathbb{E}[w(\theta, \eta)] / w_{\text{top}}$. Thus, $\frac{\partial}{\partial \eta} \hat{w}_{\text{co}}(\theta, \eta) > 0$, so that the crowding-out effect raises the sensitivity of pre-tax earnings to performance shocks. Note that if the baseline tax code is linear and the tax rates increase uniformly for the whole population, the crowding-out effect is equal to zero. On the other hand, the sum of the standard labor supply effect $\hat{w}_{\text{ex}}(\theta, \eta)$ and the performance-pay effect $\hat{w}_{\text{pp}}(\theta, \eta)$ is increasing if the reform reduces optimal effort, $\hat{a}(\theta) < 0$, which raises the performance sensitivity of the contract. The larger the average uncompensated elasticity of labor supply, the stronger the performance-pay effect relative to the crowding-out effect, the more an increase in top tax rates reduces the dispersion of earnings at the top. Overall, the effect of this reform on the earnings distribution is ambiguous.

Proportional Decrease in Retention Rates. Suppose that Assumption 2 holds. Consider a tax reform that raises the parameter τ of the CRP tax schedule by a small amount δ. The direction $\hat{T}$ of this tax reform is such that

$$\left(w - \frac{1 - \tau - \delta}{1 - p} w^{1-p} \right) - \left(w - \frac{1 - \tau}{1 - p} w^{1-p} \right) = \delta \hat{T}(w) + o(\delta).$$

This easily implies that this tax reform $\hat{T}$ is defined by

$$\hat{T}(w) = \frac{1}{1 - p} w^{1-p}, \quad \forall w > 0. \tag{64}$$

Note that this reform changes the retention rates $r(w)$, in percentage terms, by a negative constant: $\frac{\hat{r}(w)}{r(w)} = -\frac{\hat{T}'(w)}{1 - T'(w)} = -\frac{1}{(1-\tau)(1-p)}$. Therefore, it amounts to a proportional reduction in retention rates.

Applying formula (14) to the tax reform (64) yields the following results. The two

components of crowding-out are equal to

$$\frac{\hat{T}\left(w\left(\theta,\eta\right)\right)}{r\left(w\left(\theta,\eta\right)\right)} = \frac{\left(v'\left(w\left(\theta,\eta\right)\right)\right)^{-1}}{\mathbb{E}\left[\left(v'\left(w\left(\theta,\cdot\right)\right)\right)^{-1}\right]}\mathbb{E}\left[\frac{\hat{T}\left(w\left(\theta,\cdot\right)\right)}{r\left(w\left(\theta,\cdot\right)\right)}\right] = \frac{w\left(\theta,\eta\right)}{\left(1-\tau\right)\left(1-p\right)}.$$

That is, in response to a tax reform that reduces all retention rates by the same percent amount (and hence raises marginal tax rates), the firm first increases all workers' salaries in proportion to their initial earnings in order to counteract their net income losses and keep their incentives unchanged. But the reform also reduces rents and hence leads firms to reduce salaries also in proportion to the workers' initial earnings. Indeed, since the utility function is logarithmic, this ensures that all agents' utilities decrease by the same amount. Therefore, we obtain that the total crowding out of private insurance by the reform, $\hat{w}_{\text{co}}\left(\theta,\eta\right)$, is equal to zero.

Now, the standard labor supply effect $\hat{w}_{\text{ex}}\left(\theta,\eta\right)$ and the performance-pay effect $\hat{w}_{\text{pp}}\left(\theta,\eta\right)$ are both also equal to zero because, by equation (31), the optimal effort level depends only on the rate of progressivity and not on the tax parameter τ. Intuitively, effort remains constant because the utility function is logarithmic, so that the substitution and income effects cancel out. As a result, the earnings schedule is completely unaffected by the reform. $\qquad\qquad\qquad\qquad\qquad\qquad\square$

F   Proofs of Section 5

In this section we derive the optimal progressivity formula in the quantitative model. The effort and the expected utility (conditional on ability θ) of a normal worker are

$$a_n = (1-p)^{\frac{\varepsilon}{1+\varepsilon}}$$

and

$$U_n(\theta) = \log\left(\frac{1-\tau}{1-p}\right) + (1-p)\log(\theta a_n) - h(a_n).$$

The effort and the expected utility (conditional on ability θ) of a performance-pay worker are

$$a_m = \left[(1-p)\left(1-\frac{1}{\varepsilon}\psi^2\sigma_{\eta,m}^2\right)\right]^{\frac{\varepsilon}{1+\varepsilon}}$$

82

and

$$U_m(\theta) = \log\left(\frac{1-\tau}{1-p}\right) + (1-p)\log(\theta a_m) - h(a_m) - \frac{\psi^2 \sigma_{\eta,m}^2}{2},$$

where the endogenous pass-through ψ is equal $\frac{a_m^{1/\varepsilon}}{1-p}$. The Utilitarian social welfare function is

$$
\begin{aligned}
\mathcal{W} &= \pi \int U_m(\theta) dF_m(\theta) + (1-\pi) \int U_n(\theta) dF_n(\theta) \qquad (65)\\
&= \log\left(\frac{1-\tau}{1-p}\right) + (1-p)\left(\pi\mu_{\theta,m} + (1-\pi)\mu_{\theta,n} + \frac{1}{\lambda_\theta}\right)\\
&\quad + \pi\left((1-p)\log(a_m) - h(a_m) - (1-p)\frac{\psi^2 \sigma_{\eta,m}^2}{2}\right)\\
&\quad + (1-\pi)\left((1-p)\log(a_n) - h(a_n)\right)
\end{aligned}
$$

where the distribution of productivities at performance-pay jobs $F_m(\theta)$ is Pareto-lognormal with parameters $(\mu_{\theta,m}, \sigma_\theta^2, \lambda_\theta)$, and the distribution of productivities at normal jobs $F_n(\theta)$ is Pareto-lognormal with parameters $(\mu_{\theta,n}, \sigma_\theta^2, \lambda_\theta)$. We derive the optimality condition by differentiating $\mathcal{W}$ with respect to $1-p$, applying the envelope theorem and equating the derivative to zero:

$$
\begin{aligned}
0 &= \frac{\partial \mathcal{W}}{\partial(1-p)} = \frac{\partial \log\left(\frac{1-\tau}{1-p}\right)}{\partial 1-p} + \pi\mu_{\theta,m} + (1-\pi)\mu_{\theta,n} + \frac{1}{\lambda_\theta}\\
&\quad + \pi\log(a_m) + (1-\pi)\log(a_n) - \pi\left(\frac{1}{2} + \varepsilon_{\psi,1-p}\right)\psi^2 \sigma_{\eta,m}^2.
\end{aligned}
$$

For each rate of progressivity, the other tax parameter τ is chosen to balance the government budget subject to fixed government spending G. Therefore, the resource constraint reads $Y = C + G$, where the aggregate output Y is given by

$$Y = \frac{\lambda_\theta}{\lambda_\theta - 1}\left(\pi e^{\mu_{\theta,m} + \frac{\sigma_\theta^2}{2}} a_m + (1-\pi)e^{\mu_{\theta,n} + \frac{\sigma_\theta^2}{2}} a_n\right)$$

and the aggregate consumption C is given by

$$C = \frac{1-\tau}{1-p}\frac{\lambda_\theta}{\lambda_\theta - (1-p)} \times$$

$$\left(\pi e^{(1-p)\mu_{\theta,m}+(1-p)^2\frac{\sigma_\theta^2}{2}}e^{-p(1-p)\frac{(\psi\sigma_{\eta,m})^2}{2}}a_m^{1-p} + (1-\pi)e^{(1-p)\mu_{\theta,n}+(1-p)^2\frac{\sigma_\theta^2}{2}}a_n^{1-p}\right).$$

Using the resource constraint, we have

$$\frac{1-\tau}{1-p} = \frac{1-\tau}{1-p}\frac{Y}{C}\left(1-\frac{G}{Y}\right).$$

Plug in the expression for Y and C and differentiate with respect to $1-p$ to obtain

$$\begin{aligned}
\frac{\partial \log\left(\frac{1-\tau}{1-p}\right)}{\partial (1-p)} &= -\frac{1}{\lambda_\theta - (1-p)} + \frac{1}{1-p}\varepsilon_{Y,1-p}\frac{g}{1-g} + \frac{1}{1-p}\varepsilon_{Y,1-p} \\
&\quad -c_m\left(\mu_{\theta,m} + (1-p)\sigma_\theta^2 + \varepsilon_{a_m,1-p} + \log(a_m)\right) \\
&\quad -c_m\left(1-p-\frac{1}{2} - p\varepsilon_{\psi,1-p} - p\varepsilon_{a_m,1-p}\varepsilon_{\psi,a_m}\right)\psi^2\sigma_{\eta,m}^2 \\
&\quad -c_n\left(\mu_{\theta,n} + (1-p)\sigma_\theta^2 + \varepsilon_{a_n,1-p} + \log(a_n)\right)
\end{aligned}$$

where $\varepsilon_{Y,1-p} \equiv y_m\varepsilon_{a_m,1-p} + y_n\varepsilon_{a_n,1-p}$ is the elasticity of the aggregate income with respect to $1-p$, $g = \frac{G}{Y}$ is the share of government spending in output and c_j is the share of jobs of type $j \in \{m,n\}$ in the aggregate consumption. Plugging this expression into the optimality condition and rearranging yields the final optimality condition

$$\begin{aligned}
&\left(p + \frac{g}{1-g}\right)\frac{\varepsilon_{Y,1-p}}{1-p} + (y_m - c_m)\left(\varepsilon_{a_m,1-p} - \varepsilon_{a_n,1-p}\right) + \pi p\varepsilon_{a_m,1-p}\varepsilon_{\psi,a_m}\psi^2\sigma_{\eta,m}^2 \\
&= \frac{1}{\lambda_\theta}\frac{1-p}{\lambda_\theta - (1-p)} + (1-p)\sigma_\theta^2 + \pi(1-p)(1+\varepsilon_{\psi,1-p})\psi^2\sigma_{\eta,m}^2 \\
&\quad -(\pi - c_m)\left(\mu_{\theta,m} + \log(a_m) - \mu_{\theta,n} - \log(a_n)\right) \\
&\quad -(\pi - c_m)\left(\frac{1}{2} - p\left(1 + \varepsilon_{\psi,1-p} + \varepsilon_{a_m,1-p}\varepsilon_{\psi,a_m}\right)\right)\psi^2\sigma_{\eta,m}^2
\end{aligned}$$

where y_j is the share of jobs of type $j \in \{m,n\}$ in the aggregate output. The above formula collapses to the formulas with only performance pay jobs when $\pi = 1$ and the standard progressivity formula when $\pi = 0$ (in both cases the blue terms disappear).

The first term is the standard deadweight loss from rising the progressivity rate adjusted by the government spending, evaluated using the elasticity of aggregate income $\varepsilon_{Y,1-p} \equiv y_m \varepsilon_{a_m,1-p} + y_n \varepsilon_{a_n,1-p}$. The second term is a correction to the deadweight loss due to both differences in elasticities between job types and discrepancies between income and consumption shares of job types. The intuition is as follows. Note that performance-pay workers, who are more elastic, have lower consumption share than income share when $p > 0$ due to higher wage-rate risk. Suppose we increase progressivity. Performance-pay workers will reduce their effort, which decreases both their income (negative effect on available resources) and their consumption (positive effect on available resources). Since their income share is higher than their consumption share, aggregating both effects across all performance-pay workers leads to a negative effect on available resources. The opposite is true for normal workers: their consumption share is greater than income share, which means that on aggregate, there are more available resources due to their responses. However, since performance-pay workers are more elastic than normal workers, the former effect dominates and increasing progressivity leads to an additional deadweight loss.

The next four terms are standard: the performance-pay effect, the redistribution gains due to the the Pareto tail and due to the normal variance of productivities, and the gain from insuring endogenous wage-rate risk at the performance-pay jobs net of the crowd-out.

The last two terms are novel. They are present because the consumption share of performance pay workers c_m is potentially different from their population share π. These terms correspond to various ways in which resources are redistributed between job types. The first of the two terms stands for the gain from insuring the "job type" risk, that is, the risk of having a performance-pay job vs. a normal job. Recall that this risk is exogenous. This term contributes to higher progressivity whenever there is a difference in mean consumption at the two job types.

The last term is related to the endogenous wage-rate risk. When taxes are progressive, higher wage-rate risk of performance-pay workers lowers their consumption relative to normal workers. Consequently, changes in progressivity as well as endogenous adjustments of the wage-rate risk will result in the transfers of resources across job types via the government budget constraint. Suppose that the term in the big brackets is positive, which is likely if p is not very high.[29] Then, when progressiv-

---

[29]Plugging in the values of elasticities, the term in the brackets becomes $\frac{1}{2} - p\frac{\varepsilon_{a_m,1-p}}{\varepsilon}$. Since

ity increases, performance pay workers end up contributing more to the government budget. This implies an additional redistribution from the performance-pay workers to the normal workers. This effect contributes to higher (lower) progressivity if the consumption share of performance-pay workers is higher (lower) than their population share.

G   Dynamic Model

In this section we extend our results to a dynamic model of the labor market. In our setting, individuals live for several periods and sign a long-term labor contract with a firm. We use the moral hazard model of Edmans, Gabaix, Sadzik, and Sannikov (2012) who extended Edmans and Gabaix (2011) to the multi-period environment. Workers' earnings can depend in an arbitrary way on their history of output realizations. The government levies a labor income tax in each period and has a redistributive social welfare objective.

G.1   Environment

Individuals are indexed by their exogenous and constant labor productivity $\theta \in \Theta$. They live for $S \geq 2$ periods, have time-separable preferences over consumption c_t and effort a_t with discount factor $\beta \in (0, 1)$. Throughout this section, we denote the history of a random variable x up to time $t \leq S$ by $x^t \equiv \{x_s\}_{1 \leq s \leq t}$ and let $x^0 = \emptyset$.

Flow output at time t is given by:

$$y_t = \theta (a_t + \eta_t), \tag{66}$$

where $\{\eta_t\}_{1 \leq t \leq S}$ are independent and identically distributed random variables. Throughout the analysis we assume that the utility of consumption is logarithmic with isoelastic disutility of labor, productivity θ is lognormally distributed with mean μ_θ and variance σ_θ^2, and the performance shocks η_t are normally distributed with mean 0 and variance σ_η^2. As in Section 1, we assume that the agent chooses period-t effort a_t *after* observing the realization of η_t. Therefore, the agent's strategy can be a function $a_t(\eta^t)$ of her history (including the current-period) of performance shocks.

$\varepsilon_{a_m, 1-p} > \frac{\varepsilon}{1+\varepsilon}$, this term is positive if $p < \frac{1+\varepsilon}{2}$.

Firms discount future profits at rate r. For simplicity we assume that $\beta (1 + r) = 1$. In each period t they observe the agent's productivity θ and her output history y^t up to that date, but not her effort levels a^t or performance shocks η^t. A labor contract specifies an effort level $a_t (\theta)$ in each period t, and an earnings function $w_t (\theta, \eta^t)$ that depends on the inferred history of performance shocks (given the recommended effort levels) up to and including time t.

Finally, in each period, the government levies an income tax. We suppose that the tax schedule has a constant and history-independent rate of progressivity p, so that for all $t \in \{1, \ldots, S\}$, the retention function in period t is given by

$$R_t (w) = \frac{1 - \tau_t}{1 - p} w^{1-p}.$$

The parameter τ_t ensures that the government balances its budget in each period. Finally, we rule out private savings so that an agent with earnings w_t in period t consumes $c_t = R_t (w_t)$.

G.2 Equilibrium Labor Contract

We start by setting up the contracting problem between the firm and a worker with productivity θ. The operator $\mathbb{E}_t$ denotes the expectation over all future performance shock realizations $\{\eta_s\}_{t+1 \leq s \leq S}$ conditional on ex-ante productivity θ and output history η^t.

Firm's problem. The firm's maximizes its expected profit

$$\Pi (\theta) = \max_{\{a_t(\theta), w_t(\theta, \eta^t)\}_{1 \leq t \leq S}} \mathbb{E}_0 \left[\sum_{t=1}^{S} \left(\frac{1}{1+r} \right)^{t-1} \left(y_t - w_t \left(\theta, \eta^t \right) \right) \right], \qquad (67)$$

subject to the incentive constraint which requires that for any alternative effort strategy $\{\tilde{a}_t (\eta^t)\}_{1 \leq t \leq S}$,

$$\mathbb{E}_1 \left[\sum_{t=1}^{S} \beta^{t-1} \left(u \left(R_t \left(w_t \left(\theta, \eta^t \right) \right) \right) - h \left(\tilde{a}_t \left(\eta^t \right) \right) \right) \right] \qquad (68)$$

$$\leq \mathbb{E}_1 \left[\sum_{t=1}^{S} \beta^{t-1} \left(u \left(R_t \left(w_t \left(\theta, \eta^t \right) \right) \right) - h \left(a_t \left(\theta \right) \right) \right) \right],$$

and the participation constraint:

$$\mathbb{E}_0 \left[\sum_{t=1}^{S} \beta^{t-1} \left(u \left(R \left(w_t \left(y^t \mid \theta \right) \right) \right) - h \left(a_t \left(\theta \right) \right) \right) \right] \geq U \left(\theta \right), \tag{69}$$

where y_t is given by (66) and $U \left(\theta \right)$ is the reservation value of workers with productivity θ.

Equilibrium. We assume that there is free entry of firms, so that in equilibrium profits are equal to zero:

$$\Pi \left(\theta \right) = 0. \tag{70}$$

This equation pins down the workers' reservation value $U \left(\theta \right)$.

Optimal Contract. We characterize the equilibrium contract in two steps: we first study its intertemporal, then its intratemporal, properties.

Lemma 3. *The earnings process $w_t \left(\theta, \eta^t \right)$ is a martingale. That is, expected period-t earnings are equal to realized period-$(t-1)$ earnings,*

$$\mathbb{E}_{t-1} \left[w_t \left(\theta, \eta^{t-1}, \eta_t \right) \mid \eta^t \right] = w_{t-1} \left(\theta, \eta^{t-1} \right). \tag{71}$$

Proof. See Appendix H. □

This lemma characterizes the intertemporal properties of the optimal contract between the firm and the worker. It is well known that the solution to dynamic contracting models under separable utility satisfies the Inverse Euler Equation (see, e.g., Rogerson (1985); Golosov, Kocherlakota, and Tsyvinski (2003)). Intuitively, the firm incurs a convex cost of providing effort incentives – giving x_t utils in period t requires paying a before-tax salary $R_t^{-1} \left(u^{-1} \left(x_t \right) \right)$, where the cost function $R_t^{-1} \circ u^{-1} \equiv C_t$ is convex. As a result, the optimal contract smooths out the cost of providing incentives over time, which requires $\mathbb{E}_t \left[C_t' \left(x_{t+1} \right) \right] = C_t' \left(x_t \right)$. Under the assumptions that the utility function u is logarithmic and the retention function R_t is CRP, this equation can be rewritten as (71).

Proposition 5. *Assume that effort is positive in each period, or that $h'(0) = 0$. Define the present value of effort by $A \equiv \sum_{s=1}^{S} \left(\frac{1}{1+r}\right)^{s-1} a_s$, and the sequences of sensitivity and pass-through parameters $\{\delta_t, \psi_t\}_{1 \leq t \leq S}$ by*

$$\delta_t = \frac{1}{\sum_{s=0}^{S-t} \beta^s}, \quad \text{and} \quad \psi_t = \delta_t \frac{h'(a_t)}{1-p}.$$

The earnings schedule satisfies

$$\log\left(w_t\left(\theta, \eta^t\right)\right) = \log\left(w_{t-1}\left(\theta, \eta^{t-1}\right)\right) + \psi_t \eta_t - \frac{1}{2}\psi_t^2 \sigma_\eta^2, \tag{72}$$

where initial earnings are given by $w_0 \equiv \delta_1 \theta A$. The optimal period-$t$ effort level a_t is independent of θ and satisfies

$$a_t = \left[(1-p)\left(\frac{a_t}{\delta_1 A} - \frac{1}{\delta_t}\varepsilon_{\psi_t, a_t}\psi_t^2 \sigma_\eta^2\right)\right]^{\frac{\varepsilon}{1+\varepsilon}}, \tag{73}$$

where $\varepsilon_{\psi_t, a_t} = \frac{1}{\varepsilon}$ is the elasticity of the pass-through parameter ψ_t with respect to effort a_t. Expected utility is given by

$$U(\theta) = \sum_{t=1}^{S} \beta^{t-1}\left[\log\left(R\left(\delta_1 \theta A\right)\right) - h(a_t) - \frac{1}{2\delta_t}\psi_t^2 \sigma_\eta^2\right]. \tag{74}$$

Proof. See Appendix H. $\square$

This proposition generalizes Corollary 1 to the dynamic setting and allows us to characterize the intratemporal properties of the optimal compensation contract. Equation (72) implies that earnings in each period t are a log-linear function of the performance shock η_t in that period. The pass-through of performance shocks η_t to log-earnings, $\psi_t = \partial \log w_t(\theta, \cdot)/\partial \eta_t$, is increasing in the rate of progressivity p and the optimal effort level a_t at time t. Note finally that the pass-throughs have the same form as in the static model – we therefore expect the insight that the performance-pay effect counteracts and offsets a large share of the crowding-out effect to carry over to the dynamic environment.

Formally, up to the optimal value of effort, the pass-through ψ_S in the terminal period S is the same as in the optimal static contract (see equation (30)) since $\delta_S = 1$. In earlier periods, on the other hand, the exposure to risk for a given effort level is

strictly smaller than in the static environment as the sensitivity parameters δ_t satisfy $\delta_t < 1$ for all $t \leq S - 1$. To understand the intuition for this result, note that equation (72) implies that an increase in the output realization y_t – either due to effort or to random shocks – boosts log-earnings in the current and 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.

G.3 Optimal Tax Progressivity

We finally characterize the optimal history-independent rate of progressivity p in the dynamic environment. The government chooses p to maximize a utilitarian social objective $\int_\Theta U(\theta) \, \mathrm{d}F(\theta)$ subject to period-by-period budget balance constraint that $\int_\Theta R_t(w_t(\theta, \eta^t)) \, \mathrm{d}F(\theta) \geq 0$.

Proposition 6. *The optimal rate of progressivity is given by*

$$\frac{p^*}{(1 - p^*)^2} = \frac{\sigma_\theta^2}{\varepsilon_{A,1-p} + (1-p) \sum_{s=1}^{S} \beta^{s-1} \frac{\delta_1}{\delta_s} \varepsilon_{\psi_s, a_s} \varepsilon_{a_s, 1-p} \psi_s^2 \sigma_\eta^2}, \tag{75}$$

where $\varepsilon_{A,1-p}$ is the elasticity of the present discounted value of effort A with respect to progressivity, and $\varepsilon_{\psi_s, a_s} = \frac{1}{\varepsilon}$.

Proof. See Appendix H. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad$ □

To compare the optimum rate of progressivity (75) to its static counterpart (43), first consider the benchmark environment with exogenous wage risk. That is, the planner observes ex-ante earnings heterogeneity due to productivity shocks θ, and ex-post heterogeneity due to performance shocks η_t passed through to earnings. In particular, it observes that the degree of performance-pay rises with age, as described in Proposition 5. However, it mistakenly believes that wage rates, and hence ψ_t, are exogenous. That is, it assumes that $\varepsilon_{\psi_s, a_s} = \varepsilon_{a_s, 1-p} = 0$ for all $s \geq 1$. In this case, the dynamic optimal tax formula (75) is identical to the static formula (43), except that the relevant labor supply elasticity is now the elasticity of the present-value of effort, $\varepsilon_{A,1-p}$.

Now consider the general model with ex-post earnings dispersion and endogenous wage risk, captured by the non-zero elasticities $\varepsilon_{\psi_s, a_s}$ and $\varepsilon_{a_s, 1-p}$. As in the static model, the fiscal externality and welfare effect induced by the crowding-out effect cancel each other out. The adjustment to the optimal rate of progressivity is given by the second term in the denominator of (75). Recall that this term accounts for the negative fiscal externality due to the performance-pay effect: a higher rate of progressivity reduces effort in period s, hence reduces the dispersion of earnings which in turn (by Jensen's inequality) negatively affects government revenue. This term resembles the present value of the corresponding terms in the static model, with one difference. Namely, the relevant discount factor is not β^{s-1} but $\beta^{s-1} \frac{\delta_1}{\delta_s}$. 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.

H Proofs of Section G

Proof of Lemma 3. Starting from an incentive compatible allocation, consider the following variations in retained wage/utility:

$$\hat{u}_{t-1} = u\left(R\left(w_{t-1}\left(\theta, \eta^{t-1}\right)\right)\right) - \beta\Delta$$

$$\hat{u}_t = u\left(R\left(w_t\left(\theta, \eta^{t-1}, \eta_t\right)\right)\right) + \Delta$$

and $\hat{u}_s = u\left(R\left(w_s\left(\theta, \eta^s\right)\right)\right)$ for all $s \notin \{t-1, t\}$. These perturbations preserve utility and incentive compatibility since for all a_{t-1}

$$\hat{u}_{t-1} - h\left(a_{t-1}\right) + \beta\mathbb{E}\left[\hat{u}_t | \eta^{t-1}\right]$$
$$= u\left(R\left(w_{t-1}\left(\theta, \eta^{t-1}\right)\right)\right) - h\left(a_{t-1}\right) + \beta\mathbb{E}\left[u\left(R\left(w_t\left(\theta, \eta^{t-1}, \eta_t\right)\right)\right) | \eta^{t-1}\right].$$

The optimal allocation must be unaffected by such deviations, so that

$$0 = \arg\min_{\Delta} \mathbb{E}\left[\sum_{s=1}^{S}\left(1+r\right)^{-t}\left(y_s - W\left(\hat{u}_s\right)\right)\right]$$

91

Where $W = (u \circ R)^{-1}$. The associated first-order condition evaluated at $\Delta = 0$ reads

$$W' \left(u \left(R \left(w_{t-1} \left(\theta, \eta^{t-1} \right) \right) \right) \right) = \frac{1}{\beta \left(1 + r \right)} \mathbb{E} \left[W' \left(u \left(\left(R \left(w_t \left(\theta, \eta^{t-1}, \eta_t \right) \right) \right) \right) \right) | y^{t-1} \right]$$

that is,

$$\mathbb{E} \left[\frac{1}{v' \left(w_t \left(\theta, \eta^{t-1}, \eta_t \right) \right)} \mid y^{t-1} \right] = \beta \left(1 + r \right) \frac{1}{v' \left(w_{t-1} \left(\theta, \eta^{t-1} \right) \right)}.$$

The inverse Euler equation (see Golosov, Kocherlakota, and Tsyvinski (2003)) holds in our setting. With log utility and a CRP tax schedule, this equation can be rewritten as

$$\left(1 - p \right) \mathbb{E} \left[w_t \left(\theta, \eta^{t-1}, \eta_t \right) \mid y^{t-1} \right] = \left(1 - p \right) \beta \left(1 + r \right) w_{t-1} \left(\theta, \eta^{t-1} \right),$$

which leads to equation (71) as $\beta \left(1 + r \right) = 1$. $\qquad\square$

Proof of Proposition 5. We provide a heuristic proof of this proposition, and the formal argument follows the same steps as in Edmans, Gabaix, Sadzik, and Sannikov (2012). Assume that a unique level of effort is implemented at each time t, that these effort levels are independent of previous output noise, and that local incentive constraints are sufficient conditions. Consider first the incentive compatibility constraint which ensures that the worker does not wish to choose a different level of effort than the one recommended by the firm. Consider a local deviation in effort a_t after history $\left(\eta^{t-1}, \eta_t \right)$. The effect of such a deviation on the worker's lifetime utility U should be zero,

$$\mathbb{E}_{t-1} \left[\frac{\partial U}{\partial y_t} \frac{\partial y_t}{\partial a_t} + \frac{\partial U}{\partial a_t} \right] = 0.$$

Since $\frac{\partial y_t}{\partial a_t} = \theta$, we obtain

$$\mathbb{E}_{t-1} \left[\frac{\partial U}{\partial y_t} \right] = -\frac{1}{\theta} \frac{\partial U}{\partial a_t} \tag{76}$$

Applying incentive compatibility for effort in the final period we obtain:

$$r \left(w_S \left(\theta, \eta^S \right) \right) u' \left(R \left(w_S \left(\theta, \eta^S \right) \right) \right) \frac{\partial w \left(\theta, \eta^{S-1}, \eta_S \right)}{\partial \eta_S} = h' \left(a_S \left(\theta \right) \right).$$

92

Fixing η^{S-1} and integrating this incentive constraint over η_S (meaning over realizations of η_S given $a(\theta)$) leads to

$$u\left(R\left(w_S\left(\theta,\eta^S\right)\right)\right) = h'(a_S(\theta))\eta_S + z^{S-1}\left(\eta^{S-1}\right)$$

for some function of past output $z^{S-1}\left(\eta^{S-1}\right)$. This implies in particular that

$$\frac{\partial u\left(R\left(w_S\left(\theta,\eta^S\right)\right)\right)}{\partial\eta_{S-1}} = \frac{\partial z^{S-1}\left(\eta^{S-1}\right)}{\partial\eta_{S-1}}.$$

Analogously, the incentive constraint for effort in the second to last period reads

$$r\left(w_{S-1}\left(\theta,\eta^{S-1}\right)\right)u'\left(R\left(w_{S-1}\left(\theta,\eta^{S-1}\right)\right)\right)\frac{\partial w\left(\theta,\eta^{S-1}\right)}{\partial\eta_{S-1}}$$

$$+\beta r\left(w_S\left(\theta,\eta^S\right)\right)u'\left(R\left(w_S\left(\theta,\eta^S\right)\right)\right)\frac{\partial w_S\left(\theta,\eta^S\right)}{\partial\eta_{S-1}} = h'(a_{S-1}(\theta)).$$

Integrating the previous expression over η_{S-1} and using the previous equation implies

$$u\left(R\left(w_{S-1}\left(\theta,\eta^{S-1}\right)\right)\right) + \beta z^{S-1}\left(\eta^{S-1}\right) = h'(a_{S-1}(\theta))\eta_{S-1} + z^{S-2}\left(\eta^{S-2}\right).$$

We now want to show that $z^{S-1}\left(\eta^{S-1}\right)$ is a linear function of η_{S-1}. Since the utility function is logarithmic and the tax schedule is CRP, we obtain

$$(1-p)\log\left(w_S\left(\theta,\eta^S\right)\right) = h'(a_S(\theta))\eta_S + z^{S-1}\left(\eta^{S-1}\right) - \log\left(\frac{1-\tau_S}{1-p}\right)$$

and

$$(1-p)\log\left(w_{S-1}\left(\theta,\eta^{S-1}\right)\right)$$
$$= h'(a_{S-1}(\theta))\eta_{S-1} - \beta z^{S-1}\left(\eta^{S-1}\right) + z^{S-2}\left(\eta^{S-2}\right) - \log\left(\frac{1-\tau_{S-1}}{1-p}\right).$$

Now recall that the inverse Euler equation reads

$$\mathbb{E}_{S-1}\left[w_S\left(\theta,\eta^S\right)\right] = w_{S-1}\left(\theta,\eta^{S-1}\right).$$

Using the previous expressions, this equality can be rewritten as

$$\mathbb{E}_{S-1}\left[e^{\frac{1}{1-p}h'(a_S(\theta))\eta_S}\right]e^{\frac{1}{1-p}z^{S-1}\left(\eta^{S-1}\right)}$$

$$= \left(\frac{1-\tau_S}{1-\tau_{S-1}}\right)^{\frac{1}{1-p}}e^{\frac{1}{1-p}h'(a_{S-1}(\theta))\eta_{S-1}}e^{-\beta\frac{1}{1-p}z^{S-1}\left(\eta^{S-1}\right)+\frac{1}{1-p}z^{S-2}\left(\eta^{S-2}\right)}.$$

This in turn implies

$$(1+\beta)z^{S-1}\left(\eta^{S-1}\right)$$

$$= h'(a_{S-1}(\theta))\eta_{S-1} + z^{S-2}\left(\eta^{S-2}\right) - \frac{1}{2}\frac{(h'(a_S(\theta)))^2}{1-p}\sigma_\eta^2 + \frac{1}{1-p}\log\left(\frac{1-\tau_S}{1-\tau_{S-1}}\right).$$

Therefore, $z^{S-1}\left(\eta^{S-1}\right)$, and in turn $u\left(R\left(w_{S-1}\left(\theta,\eta^{S-1}\right)\right)\right)$, 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 of exposition that $S = 2, \beta = 1, r = 0$, $\theta = 1$, so that $\delta_1 = \frac{1}{2}$ and $\delta_2 = 1$. From the arguments above we guess a log-linear specification for earnings:

$$\log w_1 = \psi_1\eta_1 + k_1$$
$$\log w_2 = \psi_{21}\eta_1 + \psi_2\eta_2 + k_1 + k_2.$$

The martingale property (71) requires $w_1 = \mathbb{E}_1[w_2]$, so that for all η_1, $e^{\psi_1\eta_1+k_1} = e^{\psi_{21}\eta_1+k_1}\mathbb{E}\left[e^{\psi_2\eta_2+k_2}\mid\eta_1\right]$. This requires $\psi_1 = \psi_{21}$ and $e^{-k_2} = \mathbb{E}\left[e^{\psi_2\eta_2}\mid\eta_1\right]$. Now, the total utility of the agent is given by

$$U = (1-p)\left[2\psi_1\eta_1 + \psi_2 y_2 + 2k_1 + k_2\right]$$
$$-h(a_1) - h(a_2) + \log\left(\frac{1-\tau_1}{1-p}\right) + \log\left(\frac{1-\tau_2}{1-p}\right).$$

The incentive constraint for effort 76 implies

$$\psi_1 = \frac{h'(a_1)}{2(1-p)}, \text{ and } \psi_2 = \frac{h'(a_2)}{1-p}$$

and therefore

$$k_2 = -\frac{h'(a_2)}{1-p} - \frac{\sigma_\eta^2}{2}\left(\frac{h'(a_2)}{1-p}\right)^2.$$

Replacing in the expression for log earnings leads to

$$\log w_1 = k_1' + \frac{h'(a_1)}{2(1-p)}\eta_1 - \frac{\sigma_\eta^2}{2}\left(\frac{h'(a_1)}{2(1-p)}\right)^2$$

and

$$\log w_2 = k_1' + \frac{h'(a_1)}{2(1-p)}\eta_1 - \frac{\sigma_\eta^2}{2}\left(\frac{h'(a_1)}{2(1-p)}\right)^2 + \left(\frac{h'(a_2)}{1-p}\right)\eta_2 - \frac{\sigma_\eta^2}{2}\left(\frac{h'(a_2)}{1-p}\right)^2,$$

where $k_1' \equiv k_1 + \psi_1 a_1 - \frac{\sigma_\eta^2}{2}\psi_1^2$. This constant is pinned down by the zero profit condition $\mathbb{E}\left[w_1 + w_2\right] = a_1 + a_2$, that is, $2e^{k_1'} = a_1 + a_2$. This implies

$$k_1' = \log\left(\frac{a_1 + a_2}{2}\right),$$

which concludes the proof of equation (72). Equations (73) and (74) are derived in the next proof. □

Proof of Proposition 6. Recall that the earnings schedule is given by

$$\log w_1 = \log(\delta_1 \theta A) + \psi_1 \eta_1 - \frac{\psi_1^2 \sigma_\eta^2}{2},$$

$$\log w_t = \log w_{t-1} + \psi_t \eta_t - \frac{\psi_t^2 \sigma_\eta^2}{2}.$$

The expected utility of workers with productivity θ is therefore equal to

$$U(\theta) = (1-p)\left[\frac{1}{\delta_1}\log(\delta_1 \theta A) - \sum_{s=1}^{S}\beta^{s-1}\frac{1}{\delta_s}\frac{\psi_s^2 \sigma_\eta^2}{2}\right]$$

$$- \sum_{s=1}^{S}\beta^{s-1}h(a_s) + \sum_{s=1}^{S}\beta^{s-1}\log\left(\frac{1-\tau_s}{1-p}\right),$$

from which (74) easily follows. Thus, utilitarian social welfare is

$$
\int_\Theta U(\theta) \, dF(\theta) = (1-p) \left[\frac{1}{\delta_1} \log(\delta_1 A) + \frac{1}{\delta_1} \mu_\theta - \sum_{s=1}^{S} \beta^{s-1} \frac{1}{\delta_s} \frac{\psi_s^2 \sigma_\eta^2}{2} \right]
$$
$$
- \sum_{s=1}^{S} \beta^{s-1} h(a_s) + \sum_{s=1}^{S} \beta^{s-1} \log \left(\frac{1-\tau_s}{1-p} \right).
$$

The first-order condition for optimal effort reads

$$
0 = \frac{\partial U(\theta)}{\partial a_t} = (1-p) \left[\frac{1}{\delta_1} \frac{1}{A} \frac{\partial A}{\partial a_t} - \beta^{t-1} \frac{1}{\delta_t} \psi_t \sigma_\eta^2 \frac{\partial \psi_t}{\partial a_t} \right] - \beta^{t-1} h'(a_t)
$$
$$
= (1-p) \left[\frac{1}{\delta_1} \frac{\left(\frac{1}{1+r}\right)^{t-1} a_t}{A} - \beta^{t-1} \frac{1}{\delta_t} \psi_t^2 \sigma_\eta^2 \varepsilon_{\psi_t, a_t} \right] \frac{1}{a_t} - \beta^{t-1} h'(a_t),
$$

which easily implies equation (73). Now, the expected present value of pre-tax and post-tax earnings in period t are given by $\mathbb{E}[w_t] = \delta_1 \theta A$ and

$$
\mathbb{E}\left[w_t^{1-p}\right]
$$
$$
= (\delta_1 \theta A)^{1-p} \mathbb{E}\left[e^{\sum_{s=1}^{t}(1-p)\psi_s \eta_s}\right] e^{-\sum_{s=1}^{t}(1-p)\frac{\psi_s^2 \sigma_\eta^2}{2}} = (\delta_1 \theta A)^{1-p} e^{-p(1-p)\sum_{s=1}^{t}\frac{\psi_s^2 \sigma_\eta^2}{2}}
$$

respectively, so that expected government revenue in period t is equal to

$$
\int_\Theta \mathbb{E}[T(w_t)] \, dF(\theta)
$$
$$
= \delta_1 A e^{\mu_\theta + \frac{\sigma_\theta^2}{2}} - \frac{1-\tau_t}{1-p} (\delta_1 A)^{1-p} e^{-p(1-p)\sum_{s=1}^{t}\frac{\psi_s^2 \sigma_\eta^2}{2}} e^{(1-p)\mu_\theta + (1-p)^2 \frac{\sigma_\theta^2}{2}}.
$$

Imposing period-by-period budget balance therefore requires

$$
\frac{1-\tau_t}{1-p} = \frac{(\delta_1 A)^p \, e^{\mu_\theta + \frac{\sigma_\theta^2}{2}}}{e^{-p(1-p)\left(\sum_{s=1}^{t}\psi_s^2\right)\frac{\sigma_\eta^2}{2}} e^{(1-p)\mu_\theta + (1-p)^2 \frac{\sigma_\theta^2}{2}}}.
$$

Substituting this expression into the social welfare function $\int_\Theta U(\theta) \, dF(\theta)$ implies

that social welfare is equal to

$$\frac{1}{\delta_1}\left[\log\left(\delta_1 A\right) + \mu_\theta + \left(1 - (1-p)^2\right)\frac{\sigma_\theta^2}{2}\right] - \sum_{s=1}^{S}\beta^{s-1}h\left(a_s\right)$$

$$+ p\left(1-p\right)\sum_{s=1}^{S}\beta^{s-1}\left(\sum_{i=1}^{s}\frac{\psi_i^2\sigma_\eta^2}{2}\right) - (1-p)\sum_{s=1}^{S}\beta^{s-1}\frac{1}{\delta_s}\frac{\psi_s^2\sigma_\eta^2}{2}$$

$$= \frac{1}{\delta_1}\left[\log\left(\delta_1 A\right) + \mu_\theta + \left(1 - (1-p)^2\right)\frac{\sigma_\theta^2}{2}\right] - \sum_{s=1}^{S}\beta^{s-1}h\left(a_s\right)$$

$$- (1-p)^2\sum_{s=1}^{S}\beta^{s-1}\frac{1}{\delta_s}\frac{\psi_s^2\sigma_\eta^2}{2}.$$

We can now maximize this expression with respect to $1 - p$ to get

$$\sum_{s=1}^{S}\left[\frac{1}{\delta_1}\frac{1}{A}\left(\frac{1}{1+r}\right)^{s-1} - \beta^{s-1}h'\left(a_s\right)\right]\frac{\partial a_s}{\partial\left(1-p\right)}$$

$$- (1-p)\sum_{s=1}^{S}\beta^{s-1}\frac{1}{\delta_s}\varepsilon_{\psi_s,a_s}\varepsilon_{a_s,1-p}\psi_s^2\sigma_\eta^2$$

$$= (1-p)\left[\frac{1}{\delta_1}\sigma_\theta^2 + \sum_{s=1}^{S}\beta^{s-1}\frac{1}{\delta_s}\psi_s^2\sigma_\eta^2\right] + (1-p)\sum_{s=1}^{S}\beta^{s-1}\frac{1}{\delta_s}\varepsilon_{\psi_s,1-p}\psi_s^2\sigma_\eta^2.$$

Using the first-order condition for effort derived above to simplify the left hand side of this expression implies

$$\frac{p}{1-p}\frac{1}{\delta_1 A}\sum_{s=1}^{S}\left(\frac{1}{1+r}\right)^{s-1}a_s\varepsilon_{a_s,1-p} + p\sum_{s=1}^{S}\beta^{s-1}\frac{1}{\delta_s}\varepsilon_{\psi_s,a_s}\varepsilon_{a_s,1-p}\psi_s^2\sigma_\eta^2$$

$$= (1-p)\left[\frac{1}{\delta_1}\sigma_\theta^2 + \sum_{s=1}^{S}\beta^{s-1}\frac{1}{\delta_s}\left(1 + \varepsilon_{\psi_s,1-p}\right)\psi_s^2\sigma_\eta^2\right].$$

But the elasticity of the present discounted value of effort is equal to

$$\varepsilon_{A,1-p} \equiv \frac{(1-p)}{A}\frac{\partial\sum_{s=1}^{S}\left(\frac{1}{1+r}\right)^{s-1}a_s}{\partial\left(1-p\right)} = \sum_{s=1}^{S}\left(\frac{1}{1+r}\right)^{s-1}\frac{a_s}{A}\varepsilon_{a_s,1-p}.$$

Moreover, we have $1 + \varepsilon_{\psi_s,1-p} = 0$. Substituting these two expressions into the

97

previous equation and rearranging terms leads to

$$\frac{p}{(1-p)^2}\left[\frac{1}{\delta_1}\varepsilon_{A,1-p}+(1-p)\sum_{s=1}^{S}\beta^{s-1}\frac{1}{\delta_s}\varepsilon_{\psi_s,a_s}\varepsilon_{a_s,1-p}\psi_s^2\sigma_\eta^2\right] = \frac{1}{\delta_1}\sigma_\theta^2.$$

This concludes the proof of equation (75). □