Optimal Income Taxation with Adverse Selection in the Labour Market

STEFANIE STANTCHEVA

Massachusetts Institute of Technology

This article studies optimal linear and non-linear redistributive income taxation when there is adverse selection in the labour market. Unlike in standard taxation models, firms do not know workers' abilities, and competitively screen them through non-linear compensation contracts, unobservable to the government, in a Miyazaki–Wilson–Spence equilibrium. Adverse selection leads to different optimal tax formulas than in the standard Mirrlees (1971) model because of the use of work hours as a screening tool by firms, which for higher talent workers results in a "rat race", and for lower talent workers in informational rents and cross-subsidies. The most surprising result is that, if the government has sufficiently strong redistributive goals, welfare is higher when there is adverse selection than when there is not. Policies that endogenously affect adverse selection are discussed. The model has practical implications for the interpretation, estimation, and use of taxable income elasticities, which are central to optimal tax design.

Key words: Adverse selection, Labour market, Optimal taxation, Rat race, Redistribution, Screening, Hidden types.

JEL Codes: D82, H21, H23, H24

1. INTRODUCTION

For many workers, the labour market may resemble a rat race, in which they have to compete for high-paying jobs by always working harder. Indeed, if talent and ability are difficult to recognize, hard work may be the only way for employees to favourably influence the perceptions of their employers and, hence, their pay. Understanding the informational structure of the labour market, and the mechanism through which hours of work and pay are set, is crucial for many policy questions. One of them is optimal income taxation, since labour supply is a key margin on which individuals may respond to taxation. What is the optimal income tax in a setting in which firms cannot directly observe workers' talents, but instead set non-linear compensation contracts to screen high ability from low ability ones? In this article, I attempt to answer this question by studying optimal linear and non-linear income taxes with adverse selection in the labour market.

The standard income taxation model, introduced in Mirrlees' (1971) seminal paper, assumes a frictionless labour market in which firms pay workers a wage equal to their ability, *i.e.* their marginal product per hour. The government, on the other hand, tries to redistribute from high to low ability workers, but does not observe abilities. It hence sets non-linear taxes subject to

incentive compatibility constraints to ensure that workers truthfully reveal their types. By contrast, in the current study, firms do not know workers' abilities and play an active role in determining hours of work and pay. When the government sets taxes, it must take into account the modified responses to them, due to the non-linear, screening wage schedules facing workers. Private market contracts are nested in and interacting with the government's contract. As an added challenge, the government does not observe those potentially complicated private labour market contracts, but only total income earned. Accordingly, it must not only anticipate which contracts workers will choose out of a fixed set, but also the set of labour contracts, that is the compensation structure itself, which will emerge endogenously to taxes.[1]

To explain the functioning of the labour market, I use a Miyazaki–Wilson–Spence (hereafter, MWS) equilibrium (Miyazaki, 1977; Wilson, 1977; Spence, 1978), which is always constrained efficient, thus *a priori* minimizing the scope for government intervention. I also discuss the Rothschild–Stiglitz (hereafter, RS) equilibrium notion (Rothschild and Stiglitz, 1976), which has its own peculiar challenges of potential non-existence and constrained inefficiency in the Supplementary Appendix. I derive new optimal linear tax formulas for a general discrete types model and characterize the full Pareto frontiers with nonlinear taxation.

The most surprising result is that, when the government has sufficiently strong redistributive goals, welfare is higher when there is adverse selection than when there is not. This result is due to the "rat race" in which high productivity workers are caught, which is engineered by firms to separate them from lower productivity ones. The use of work hours and pay as screening tools limits the flexibility of high types to react adversely to distortive taxation, and helps the government redistribute.

Second, since the usual envelope conditions on labour supply no longer hold, there are first-order welfare effects from affecting it through taxes, and the optimal linear tax formula is modified to include two new types of terms. The corrective "rat race" terms capture the cost of labour supply distortions on each type's welfare, and can make the optimal tax positive even absent any redistributive agenda, akin to a Pigouvian tax. In addition, firms are already performing some redistribution themselves by cross-subsidizing workers, which is captured in the "informational rent" terms of lower types. For a given elasticity of taxable income and a given income distribution, the optimal tax will be higher when there is adverse selection, provided the redistributive preferences of the government are sufficiently strong.[2]

Third, in the non-linear tax case, I compare the Pareto frontiers under three different informational regimes: the standard Mirrlees, the Second Best with Adverse Selection—in which neither the government nor firms know workers' types, but the government observes private labour contracts—and the Adverse Selection with unobservable private contracts. The main result carries over: whenever the government wants to redistribute from high to low ability workers, the Pareto frontiers with Adverse Selection—with either observable or unobservable private contracts—are strictly above the Mirrlees frontier. A sufficient condition on social preferences is that lower types are weighted cumulatively more than their cumulative proportions in the population. When private contracts are unobservable to the government, it can still implement any Second Best allocation with observable contracts using a mix of nonlinear income taxes levied on workers, and non-linear payroll taxes levied on firms.

I discuss the two main policy implications of these findings and draw the link to tax praxis. First, I outline how the interpretation, estimation, and use of taxable income elasticities is complicated

---

1. In the standard model, the pretax *income* distribution of the economy is endogenous to taxes, but the endogeneity is driven solely by hours worked, while wages are equal to the intrinsic productivities of workers.

2. In addition, in the RS setting, raising taxes can destroy an existing equilibrium, and hence tax policy is more constrained.

by the presence of adverse selection—an important cautionary tale given how central the latter are in the taxation literature. In particular, it is no longer straightforward to map measured elasticities into structural elasticities without knowledge of the underlying market structure. Estimation relying on reforms as natural experiments may be affected by the interconnections of different groups through their labour contracts. Even correctly estimated, these elasticities are no longer sufficient statistics for the deadweight loss of taxation, and strict reliance on them for optimal tax design may be misleading. Secondly, the result that welfare may be higher with adverse selection suggests that a government with highly redistributive preferences might find some degree of adverse selection useful, and naturally leads to question to what extent the information structure of the economy is endogenous to government policies. Some widely used labour market interventions, such as bans on discrimination, or regulations on firing and pay structures can affect the degree of adverse selection, and, by consequence, the optimal tax and welfare.

Empirical Literature on adverse selection: All results in this article are based on two empirically testable assumptions. First, there must be asymmetric information about worker quality between firms and workers, a friction that has been widely documented. Acemoglu and Pischke (1998) show that a worker's current employer has more information about his quality than other potential employers, suggesting that, at the time of hiring, quality is uncertain. Gibbons and Katz (1991) also test a model in which the incumbent employer has superior information, so that laid-off workers are perceived as lower ability.

The second assumption is that firms are screening their workers through the labour contracts offered, rather than through other direct means, such as ability tests. Although I focus on requirements on the hours of work, other productive actions which are costlier to lower ability workers, such as sophisticated training programs, or effort on specific tasks, could also serve as valid screening tools. Evidence that employers screen indirectly through training comes from Autor (2001) for Temporary Help Firms. Career concerns seem to make workers work harder in order to positively influence the perception of their employers about their talent (Holmstrom, 1978; Gibbons and Murphy, 1992; Baker *et al.*, 1994). Most closely related to this article is the empirical study of "rat races" at large law firms by Landers *et al.* (1996), who show that employees are required to work inefficiently long hours before being promoted to partners in order to distinguish those with a high propensity to work.

Related optimal taxation literature: This article contributes to the optimal taxation literature (as developed by Mirrlees, 1971; Diamond, 1998; Saez, 2001; Albanesi and Sleet, 2006; Golosov *et al.*, 2006; Weinzierl, 2011 among others), but mostly to a growing strand of it which considers the interplay of private markets and government-imposed taxation. The focus until now has generally been on private credit and insurance markets rather than on informational problems in the labour market itself. Golosov and Tsyvinski (2006) study optimal dynamic taxation when agents can secretly trade risk-free bonds, while Krueger and Perri (2010) examine the role of progressive income taxation in insuring agents when private risk sharing is imperfect. But unlike their private market equilibria, the one in this article is already constrained efficient.[3] Chetty and Saez (2010) highlight that the fiscal externality generated by private sector insurance that suffers from moral hazard or adverse selection needs to be taken into account in the optimal tax formulas. Unlike them, I focus on the labour supply contract, and deal more explicitly with the private market equilibrium. Scheuer (2013a,b) considers optimal income and profit taxes with incomplete credit markets for entrepreneurs. The link to the literature on contracts, imperfect information, and hidden trades is drawn in Section 4.1.

---

3. The potentially constrained inefficient RS equilibrium is available in the Supplementary Appendix.

The rest of the article is organized as follows. The next section describes the labour market, and solves for the optimal tax in the standard case with no adverse selection. Section 3 studies the optimal linear taxation problem with adverse selection, while Section 4 focuses on the optimal non-linear tax. Section 5 discusses the policy implications, and Section 6 concludes. Most proofs are in the Appendix, with some lengthier proofs in the Supplementary Appendix.

2. A MODEL OF THE LABOUR MARKET WITH ADVERSE SELECTION

2.1. *The labour market*

Consider a perfectly competitive labour market with workers of N different productivities, hired by risk-neutral competitive firms.[4] Type i has productivity $\theta_i \in \Theta = \{\theta_1, ..., \theta_N\}$, with $\theta_1 < ... < \theta_i < ... < \theta_N$, and produces $f(h) = \theta_i h$ units of output for h hours of work at a disutility cost of $\phi_i(h)$. The fraction of types i in the population is λ_i with $\sum_i \lambda_i = 1$. The assumptions on the cost functions required to permit screening are analogous to the ones in Spence (1978):

Assumption 1.   (i) $\phi_i'(h) > 0$, $\phi_i''(h) > 0 \ \forall h > 0$, and $\phi_i(0) = \phi_i'(0) = 0$, $\forall i$
   (ii) $\phi_i(h) < \phi_{i-1}(h) \ \ \forall h > 0, \ \forall i > 1$
   (iii) $\phi_i'(h) < \phi_{i-1}'(h) \ \forall h > 0, \forall i > 1$.

Hence, lower productivity workers not only have a higher cost of effort, but also a higher marginal cost. The utility of a worker of type i takes a simple quasilinear form:

$$U_i(c, h) = c - \phi_i(h)$$

where c is net consumption, equal to total pay y minus any taxes $T(y)$ paid to the government.

Firms cannot observe a worker's type, but can perfectly monitor hours of work. They hence post screening contracts specifying pairs of pay and hours $\{y_i, h_i\}_{i=1}^N$. There exist several equilibrium concepts for such hidden information settings, but no consensus about the best one. In this article, I focus mostly on an analytically tractable Miyazaki–Wilson–Spence foresight equilibrium (MWS), (Wilson, 1977; Miyazaki, 1977; Spence, 1978). The Supplementary Appendix contains the analysis of a Nash behaviour à la Rothschild and Stiglitz (1976).

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

In the MWS setting, each firm is only required to break even overall on its portfolio of contracts, allowing for cross-subsidization between contracts. Firms have foresight: they anticipate that if they offer a new contract, some existing contracts might become unprofitable and be withdrawn. An equilibrium always exists and is constrained efficient (Miyazaki, 1977), thus reducing the scope for government intervention.[5]

---

   4. The N-type model was introduced and solved by Spence (1978) in the context of insurance policies. I adapt it to the labour market and introduce taxes into the model.
   5. In a two types model ($N = 2$) if λ_1, the fraction of low types, is small, then the equilibrium involves cross-subsidization from high productivity to low productivity workers. High productivity workers are paid less than their product and low types are paid more than theirs. If λ_1 is sufficiently high, the MWS and RS equilibrium allocations coincide.

2.2. *The optimal linear tax without adverse selection*

Here and in Section 3, I suppose that the only two instruments available for redistribution are a linear income tax t, levied on total earned income y, and a lumpsum transfer T, which ensures budget balance.[6] As a benchmark, it is useful to solve for the standard Second Best Pareto frontier, in the case without adverse selection. A weighted sum of utilities is maximized, subject to the reaction functions of the private market. For any given tax, workers of type i choose a level of hours $h_i^*(t)$, referred to as the efficient level of hours for type i,[7] at which the marginal cost of effort just equals the net of tax return:

$$\phi_i'\left(h_i^*(t)\right) = \theta_i(1-t) \tag{1}$$

Earnings are $y_i(t) = \theta_i h_i^*(t)$. For a set of Pareto weights $\mu \equiv \{\mu_i\}_{i=1}^N$, the social welfare function is:

$$SWF(\mu) = \sum_{i=1}^N \mu_i(c_i(t) - \phi_i(h_i(t))) \tag{2}$$

Using that $c_i = y_i(1-t) + T$, the government's program is:

$$\left(P^{SB,N}(\mu)\right) : \max_t \left\{ \sum_{i=1}^N \mu_i\left(\theta_i h_i^*(t)(1-t) - \phi_i\left(h_i^*(t)\right) + T\right) \right\}$$

with

$$T = t \sum_{i=1}^N \lambda_i \theta_i h_i^*(t)$$

where $\left\{h_i^*(t)\right\}_i$ are the workers' reaction functions to taxes as defined in (1).

It is instructive to derive the optimal tax formula heuristically, using a perturbation argument as in Saez (2001). When the tax rate is raised by a marginal amount dt, there are three effects. The mechanical revenue effect, dM—the change in tax revenue if there were no behavioural responses—is simply equal to average income, denoted $y(t) \equiv \sum_i \lambda_i y_i(t)$:

$$dM = y(t)dt$$

The behavioural effect, dB, caused by changes in agents' labour supply, is:

$$dB = t\left(\sum_i \lambda_i \theta_i \frac{dh_i^*(t)}{dt}\right) dt$$

which, after some algebraic manipulations, can be rewritten as:

$$dB = -\frac{t}{1-t}\varepsilon_y y dt$$

---

6. Throughout the article, it is assumed that the government cannot observe abilities, or, equivalently, that no type-specific taxation is available. This case is called a Second Best case because the government needs to rely on distortive taxation in order to redistribute.

7. Note that this efficient level is conditional on taxes and hence different from the first-best level of hours, except for $t = 0$.

where ε_y is the usual aggregate elasticity of taxable income to the retention rate $(1-t)$, also equal to the income-share-weighted average of individual elasticities:

$$\varepsilon_y \equiv d\log y / d\log(1-t) = \sum_i \alpha_i(t)\varepsilon_{y_i}$$

where

$$\varepsilon_{y_i} \equiv d\log(y_i)/d\log(1-t)$$

is type i's taxable income elasticity, and $\alpha_i(t) \equiv \lambda_i y_i(t)/y$ is the share of total income produced by type i workers.

Finally, the welfare effect dW—sum of the individual welfare effects dW_i—is equal to the Pareto weights weighted reduction in consumption, since the indirect effect on welfare through changes in hours of work is zero by the envelope theorem:

$$dW = \sum_i dW_i = -\left(\sum_i \mu_i y_i(t)\right)dt$$

Denote the Pareto-weights weighted income shares by $\bar{y} \equiv \sum_i \mu_i y_i(t)/y$. $\bar{y}$ measures the concentration of income relative to redistributive preferences. Whenever the social welfare function puts the same weight on each type as his proportion in the population ($\lambda_i = \mu_i$, $\forall i$), $\bar{y} = 1$. If Pareto weights are concentrated mostly on those with low incomes, then $\bar{y} << 1$. Hence,

$$dW = -(y\bar{y})dt$$

The optimal tax is the one at which the sum of these three effects $dM + dB + dW$ is zero, which yields the familiar (implicit) tax formula:[8]

$$\frac{t^{SB}}{1-t^{SB}} = \frac{1-\bar{y}}{\varepsilon_y} \qquad (3)$$

This formula highlights the two usual forces determining the optimal tax, namely, the equity concern, proxied by the income distribution and Pareto weights in $\bar{y}$, and the efficiency concern, captured by the taxable income elasticity ε_y. Note that the revenue-maximizing tax rate is $t^R/(1-t^R) = 1/\varepsilon_y$, while the Rawlsian tax rate (when $\mu_1 = 1$) is $t^{Rawls}/(1-t^{Rawls}) = \frac{1-\alpha_1}{\varepsilon_y}$. If $\mu_i = \lambda_i$, for all i, the utilitarian criterion, combined with quasilinear utility yields $t^{SB} = 0$.

3. LINEAR TAXES WITH ADVERSE SELECTION

Suppose now that firms do not know workers' types. An adverse selection problem arises if, at the first best allocation, lower type workers would like to pretend they are higher types, *i.e.* if and only if:

$$\theta_{i+1}h_{i+1}^*(0) - \phi_i\left(h_{i+1}^*(0)\right) > \theta_i h_i^*(0) - \phi_i\left(h_i^*(0)\right) \quad \forall i \leq N-1 \qquad (4)$$

where $h_i^*(0)$ is as defined in (1) at $t = 0$. I assume that (4) holds throughout this section.

---

8. With $N = 2$, this can also be rewritten as:

$$\frac{t^{SB}}{1-t^{SB}} = \frac{(\mu-\lambda)\left(\frac{\alpha_2(t)}{1-\lambda} - \frac{a_1(t)}{\lambda}\right)}{\varepsilon_y}$$

Note that $\alpha_2(t)/(1-\lambda)$ and $a_1(t)/\lambda$ are the shares of total income per *worker* of each type respectively. The greater this difference, and the greater inequality. Whenever the low type workers are valued more at the margin than is justified by their population share ($\mu > \lambda$), the optimal tax is positive.
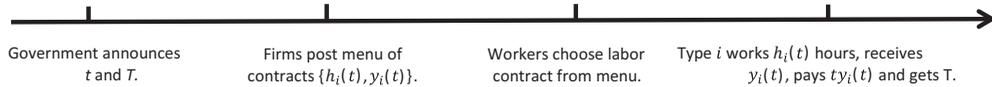
FIGURE 1

Timeline

A two-stage game takes place, with the government first setting taxes t, and the corresponding transfer T, and firms then choosing what labour contracts $\{h_i(t), y_i(t)\}_{i=1}^{N}$ to offer (see Figure 1). Working backwards from the second stage, I study the reaction functions of firms to any given tax, and then solve the government's optimal tax program, taking the responses of the labour market as additional constraints. To build the intuition, I start with $N=2$.

3.1. Second stage: the private sector's reaction to taxes

Two types model $N=2$:

In a MWS setting with $N=2$, as shown by Miyazaki (1977) for the case without taxes, firms offer a menu of contracts $\{h_i, y_i\}_{i=1}^{2}$ solving program $P^{MWS}(t)$, conditional on a given tax level t and a transfer T. The transfer T will not affect the firm's problem thanks to quasilinear utility and is omitted. Let $\lambda_1 = \lambda$ and $\lambda_2 = 1-\lambda$.

$$\left(P^{MWS}(t)\right): \max_{\{y_1, y_2, h_1, h_2\}} (1-t)y_2 - \phi_2(h_2)$$

$$(IC_{12}): (1-t)y_1 - \phi_1(h_1) \geq (1-t)y_2 - \phi_1(h_2)$$

$$(IC_{21}): (1-t)y_2 - \phi_2(h_2) \geq (1-t)y_1 - \phi_2(h_1)$$

$$(profit): \lambda y_1 + (1-\lambda)y_2 = \lambda\theta_1 h_1 + (1-\lambda)\theta_2 h_2$$

$$(RS_1): (1-t)y_1 - \phi_1(h_1) \geq (1-t)y_1^{RS} - \phi_1\left(h_1^{RS}\right) \equiv u_1^{RS}$$

The first two constraints are the incentive compatibility constraints for the low and high type respectively, ensuring that each type self-selects into the appropriate work contract. The third one is the zero profit condition on the full portfolio of contracts. The final constraint ensures that the low productivity worker always receives at least his utility from the separating allocation,[9] defined by $h_1^{RS}(t) = h_1^*(t)$ and $y_1^{RS}(t) = \theta_1 h_1^{RS}(t)$. Note that for some t, (IC_{12}) could become slack with hours at their efficient levels, even if it would be binding at $t=0$ (when (4) holds). Then we would have:

$$(1-t)\theta_2 h_2^*(t) - \phi_1\left(h_2^*(t)\right) < (1-t)\theta_1 h_1^*(t) - \phi_1\left(h_1^*(t)\right) \tag{5}$$

The following proposition characterizes the private market equilibrium for any tax t as a function of a threshold $\tilde{\lambda}(t)$ for the fraction of low types (defined in the Appendix).

Proposition 1. *For a given t, constraint (profit) is binding and (IC_{21}) is slack. The low type always works an efficient amount of hours $h_1^*(t)$, and there are three possible equilibrium*

---

9. This separating allocation is also known as the "Rothschild-Stiglitz" allocation. Why this constraint appears in the program is explained in detail in the original Miyazaki (1977) paper for the case without income taxes. In short, if this constraint was not satisfied, there would be a profitable deviation for some firm, consisting in offering a slightly worse contract than the fully separating one, attracting all low types, and making a positive profit.

configurations:

(i) If (5) holds, $h_2(t) = h_2^*(t)$, and the allocation is equal to the Second Best one.

(ii) If (5) does not hold and $\lambda > \tilde{\lambda}(t)$ (called case AS1), constraint (RS_1) is binding, each worker earns his marginal product, and there is full separation. $h_2(t)$ is above the efficient level, and is the solution to $\theta_1 h_1^*(t)(1-t) = \theta_2 h_2(t)(1-t) - \left(\phi_1(h_2(t)) - \phi_1\left(h_1^*(t) \right) \right)$.

(iii) If (5) does not hold and $\lambda \leq \tilde{\lambda}(t)$ (called case AS2), constraint (RS_1) is not binding and there is cross-subsidization from high to low productivity workers. $h_2(t)$ is above the efficient level, and is the solution to $\phi_2'(h_2(t)) = (1-t)(1-\lambda)\theta_2 + \lambda\phi_1'(h_2(t))$.

In addition,

(iv) $dh_i(t)/dt < 0$, for $i = 1, 2$

(v) $dh_2(t)/d\lambda > 0$ for $\lambda \leq \tilde{\lambda}(t)$, $dh_2(t)/d\lambda = 0$ for $\lambda > \tilde{\lambda}(t)$.

The first case occurs if the low type no longer wants to pretend to be a high type at t and is not of great interest: it is unlikely to occur with $N > 2$, and can be ruled out by assumption 2 below. More generally, low productivity workers work an efficient number of hours, but high productivity workers work excessively. There is a critical level of the fraction of low types, $\tilde{\lambda}(t)$, which determines whether firms find it profitable to cross-subsidize workers or not. The intuition is that, for a low λ, it is beneficial to reduce the distortion in the labour supply of high types in exchange for a higher cross-subsidy to low types. When the fraction of low types increases, however, this subsidy to each of them becomes too costly, and there is separation.[10] Hours of work of the high type are increasing in λ for $\lambda \leq \tilde{\lambda}(t)$ because of the standard trade-off in screening models between the distortion imposed on the high type and the informational rent forfeited to the low type (see Laffont and Martimort, 2001). The higher the fraction of bad types, the costlier it becomes for the firm to give up an informational rent to each of them. As a result, the hours of the high type must be distorted more.

An additional assumption, namely that the disutility of labour is isoelastic, simplifies the exposition but is not needed for the derivation of the optimal tax in Subsection 3.2.

Assumption 2. $\phi_i(h) = a_i h^\eta$, *for $i = 1, 2$.*

To satisfy assumption 1 with this specification would require $a_1 > a_2$.

Proposition 2.   *If assumption 2 holds:*

i) IC_{12} binds at all t,

ii) $\tilde{\lambda}(t)$ is independent of t: $\tilde{\lambda}(t) = \tilde{\lambda}, \forall t$.

Result (i) states that, if at $t = 0$ there is an adverse selection problem (*i.e.* (4) holds) and assumption 2 holds, then there is an adverse selection problem at all tax levels. Then, the marginal utility of the low type from his own efficient allocation and from deviating to the high type's efficient allocation grow at the same rate, and the relative rewards from cheating versus revealing truthfully are unaffected by the tax. The second result guarantees that the type of equilibrium does not depend on the tax rate. $\tilde{\lambda}(t)$ is equal to the ratio of the marginal welfare loss of the high type $\theta_2(1-t) - \phi_2'(h_2)$ and the marginal informational rent gain of the low type, $\theta_1(1-t) - \phi_1'(h_1)$. As long as the cost of distortion in h_2 remains low relative to the informational rent ($\lambda > \tilde{\lambda}(t)$), the contract is separating. If it grows too high ($\lambda < \tilde{\lambda}(t)$), it becomes better to grant the low type

---

10. But each worker earns his product at the equilibrium levels of hours only. It is not the case that a worker would earn his marginal product had he chosen another level of hours, unlike in standard competitive labour markets.

a cross-subsidy rather than to keep distorting hours of work. With isoelastic disutility functions these two effects grow at the same rate with the tax, so that their ratio is independent of t.

N types model, for $N \geq 2$:

For any t set by the government, define a sequence of programs $\left(P_i^{MWS}(t)\right)_{i=1}^N$ and utilities $\bar{u}_i$ such that:

$$\left(P_1^{MWS}(t)\right): \bar{u}_1 = \max_h \theta_1 h(1-t) - \phi_1(h)$$

For $2 \leq i \leq N$:

$$\left(P_i^{MWS}(t)\right): \bar{u}_i = \max_{\{y_j, h_j\}_{j=1}^i} y_i(1-t) - \phi_i(h_i)$$

subject to:

$$y_j(1-t) - \phi_j(h_j) \geq \bar{u}_j, \; j < i$$

$$y_j(1-t) - \phi_j(h_j) \geq y_{j+1}(1-t) - \phi_j(h_{j+1}), \; j < i$$

$$\sum_{j=1}^i (\theta_j h_j - y_j)\lambda_i = 0$$

The equilibrium with N types is the set of income and hour pairs $\{y_i, h_i\}_{i=1}^N$ which solve, for a given t, program $P^{MWS,N}(t) \equiv P_N^{MWS}(t)$.

Proposition 3. *In the MWS equilibrium with N types, $N \geq 2$:*

(i) There is a number of "break agents" $k_1, k_2, ...k_n$ with $n \leq N$ such that:

- Firms make losses on all subsets of types of the form $\{1,..,i\}$, $\{k_1+1,...,i\}$,..., $\{k_{n-1}+1,...,i\}$ for $i \neq k_1, k_2, ...k_n$.

- Firms break even on the subsets of types of the form $\{1,..,k_1\}$, $\{k_1+1,...,k_2\}$,..., $\{k_{n-1}+1,...,k_n\}$, called "cross-subsidization groups."

- If $\left(IC_{j,j+1}\right)$ is not binding for some j, types j and $j+1$ are in two different cross-subsidization groups, called "disjoint."

(ii) The lowest productivity agents of each disjoint cross-subsidization group (including type 1) work efficient hours.

(iii) All other types work excessively much, i.e. $h_i(t) > h_i^*(t)$, $\forall t < 1$.

The cross-subsidization groups are subsets of agents such that the firms breaks even on the group as a whole, but within which some types cross-subsidize others. With N types, if $IC_{j,j+1}$ is not binding, then j and $j+1$ are in different cross-subsidization groups (see the Appendix), and the population is split into (at least) two non-interacting sets, above and including $j+1$ and strictly below $j+1$.

3.2. *First stage: the optimal linear tax problem with two types*

In the first stage, the government chooses the optimal linear tax to maximize the weighted sum of individual utilities in (2), taking as given the reaction functions of the private market, $\{y_1(t), y_2(t), h_1(t), h_2(t)\}$. With $N = 2$ and $\mu_1 = \mu$, the program is:

$$\left(P^{AS}(\mu)\right): \max \mu(y_1(t)(1-t) - \phi_1(h_1(t)) + T) + (1-\mu)(y_2(t)(1-t) - \phi_2(h_2(t)) + T)$$

$$s.t.: T = t(\lambda \theta_1 h_1(t) + (1-\lambda)\theta_2 h_2(t))$$

The behavioural and mechanical revenue effects are still the same as in the Second Best (Subsection 2.2), but the welfare effects on the two types, dW_1 and dW_2, are now different. They can be decomposed into a *direct* effect from reduced consumption, and, if and only if the envelope condition does not hold, as is the case for h_2 here, additional *indirect* effects from changing labour supplies.

First, when taxes increase, the excessively high hours of work of the high type are reduced, which has a positive marginal effect on his own welfare, called the "rat race" effect and denoted by ξ_2:

$$\xi_2 \equiv (1-t)\theta_2 - \phi_2'(h_2) \leq 0$$

Second, the "informational rent" effect captures how the rent forfeited to induce the low type to reveal his true type changes with taxes, and is denoted by κ_2:

$$\kappa_2 \equiv \left[(1-t)\theta_2 - \phi_1'(h_2)\right] \leq 0$$

As is usual in screening models, there is a trade-off for the firm between reducing the informational rent of the low type and the distortion in the hours of the high type. As taxes increase, the high type is made to work less, which reduces the distortion in his labour supply, increases the rent transfer to the low type, and hence indirectly redistributes income. When firms cross-subsidize workers, they redistribute from high to low types. Hence, the welfare effects are:

$$dW_1 = -\mu y_1 dt + \mu(1-\lambda)I^c(t)\kappa_2 \frac{dh_2(t)}{dt}dt \tag{6}$$

$$dW_2 = -(1-\mu)y_2 dt + (1-\mu)\left\{\xi_2 - \lambda I^c(t)\kappa_2\right\}\frac{dh_2(t)}{dt}dt \tag{7}$$

where the indicator variable $I^c(t) = 1$ if there is cross-subsidization, and 0 otherwise. The informational rent effect only enters when there is cross-subsidization. Setting the sum $dM + dB + dW$ to zero, we obtain the optimal tax.[11]

Proposition 4. *The optimal tax rate with adverse selection is:*

$$\frac{t^{AS}}{1-t^{AS}} = \frac{(1-\bar{y})}{\varepsilon_y} + \frac{\frac{y_2/\theta_2}{y}(1-\mu)\frac{\varepsilon_{y2}}{1-t}}{\varepsilon_y}(-\xi_2) + \frac{\frac{y_2/\theta_2}{y}I^c(\mu-\lambda)\frac{\varepsilon_{y2}}{1-t}}{\varepsilon_y}(-\kappa_2) \tag{8}$$

In general, $I^c(t)$, the elasticities, and incomes depend on t. The tax formula is thus as usual endogenous. Recall that assumption (2), however, makes the type of equilibrium, and hence I^c, independent of taxes t. It also guarantees that the Second Best never occurs (hence, $\xi_2 < 0$). Table 1 specializes formula (8) for the three possible cases.

The usual "sufficient statistics" $\bar{y}$ and ε_y are no longer sufficient, as they do not capture the rat race and informational rent effects. The latter require knowledge of the underlying disutility of effort functions.

Comparison of tax rates with and without adverse selection: There are two complementary ways of comparing the optimal tax with and without adverse selection. The first one is to take the primitives of the model, *i.e.* the production and utility functions, as given; conceptually, this

---

11. Note that $\bar{y}$, ε_y, and ε_{y2} are not the same *functions* of the tax as in the Second Best case.

TABLE 1

Case	Tax Rate Formula
Second Best (no adverse selection)	$\frac{t^{SB}}{1-t^{SB}} = \frac{1-\bar{y}}{\varepsilon_y}$
Adverse Selection with full separation (case AS1)	$\frac{t^{AS1}}{1-t^{AS1}} = \frac{(1-\bar{y})}{\varepsilon_y} - \frac{\frac{y_2/\theta_2}{y}(1-\mu)\frac{\varepsilon_{y2}}{1-t^{AS1}}}{\varepsilon_y}\xi_2$
Adverse Selection with cross-subsidization (case AS2)	$\frac{t^{AS2}}{1-t^{AS2}} = \frac{(1-\bar{y})}{\varepsilon_y} - \frac{\frac{y_2/\theta_2}{y}(1-\mu)\frac{\varepsilon_{y2}}{1-t^{AS2}}}{\varepsilon_y}\xi_2 + \frac{\frac{y_2/\theta_2}{y}(\lambda-\mu)\frac{\varepsilon_{y2}}{1-t^{AS2}}}{\varepsilon_y}\kappa_2$

is akin to comparing tax rates in two economies which are exactly the same, except that one of them suffers from adverse selection while the other does not. The second possibility is to take as given the empirically measurable parameters, namely, the elasticities of taxable income and the distributional factors, and to compare the taxes which would be optimal if it was a market with adverse selection versus one without which generated them. This approach, adopted here, is more policy relevant: it reflects the situation of a government equipped with widely available measures of elasticities and statistics about the income distribution, but unaware of the true market structure.

Proposition 5. *At given ε_y and $\bar{y}$:*

(i) In the separating equilibrium (case AS1): $t^{AS1} > t^{SB}$ and $t^{AS1} \to t^{SB}$ as $\mu \to 1$.

(ii) With cross-subsidization (case AS2): If $\mu \geq \lambda$, $t^{AS2} > t^{SB}$ and $t^{AS2} > t^{AS1}$.

At given $\bar{y}$ and ε_y, the corrective Pigouvian term in ξ_2 leads to a higher tax rate destined to reduce the distortion in hours of work of the high type, the more so when the government cares about the welfare loss of the high type (μ small). In addition, if there is cross-subsidization, a higher tax redistributes toward the lower type in two ways: directly through a higher transfer T, but also indirectly through the informational rent term κ_2. If the government puts a high weight on low type agents ($\mu \geq \lambda$), this pushes the tax up.

Comparison of welfare with and without adverse selection: To compare welfare, on the other hand, the individuals' utility functions are held constant, since welfare is measured relative to them.

Proposition 6. *For the same economy: (i) When the government has highly redistributive preferences ($\mu = 1$), welfare is higher when there is adverse selection in the labour market than when firms can perfectly observe workers' types.*

(ii) When the government only cares about high type workers ($\mu = 0$), welfare is higher when there is no adverse selection in the labour market.

(iii) The low type is always weakly better off when there is adverse selection.

The counter-intuitive result in (i) implies that the inability of firms to observe workers' productivities and their reliance on non-linear compensation contracts for screening are not necessarily detrimental when the government wants to redistribute and can only use distortionary taxes. Like in traditional second-best theory, fixing a distortion in one place (here, adverse selection in the labour market) need not be good when there is another irremovable distortion (here, the absence of non-distortionary taxation for redistribution). If the government had lumpsum taxation available, or if it did not want to redistribute, adverse selection would only cause a deadweight loss. It is the *interaction* of the imperfect instruments available to fulfil strongly

redistributive goals with adverse selection which improves welfare. This result becomes most relevant if the informational structure in this economy is endogenous, a point discussed in Section 5.

What are the sources of this welfare gain? First, the use of hours as a screening tool and the resulting rat race limit the ability of the high type to reduce his labour supply as a response to taxes. Hence, revenue is higher at any t, which is beneficial for the low type. Second, at a given μ, with adverse selection and cross-subsidization—a form of redistribution done by firms—the optimal tax required to achieve the same level of redistribution could be lower, which is beneficial for both types.[12]

3.3. *Optimal linear tax with N types*

With $N \geq 2$ types, the government maximizes social welfare as in (2), taking as given the private sector's reaction functions $\{h_i(t), y_i(t)\}_{i=1}^N$ derived in Proposition 3 and its proof. For any set of Pareto weights μ, the program is:

$$\left(P^{AS,N}(\mu)\right) : \max_t \left\{ \sum_{i=1}^N \mu_i (y_i(t)(1-t) - \phi_i(h_i(t)) + T) \right\}$$

$$s.t. : T = t \sum_{i=1}^N \lambda_i y_i(t)$$

Let $I_{ij}(t)$ be the indicator function equal to 1 if j is in i's cross-subsidization group at tax t. Define $\bar{\lambda}^i$ (respectively, $\underline{\lambda}^i$) as the proportion of types strictly better (respectively, strictly worse) than i in i's cross-subsidization group:

$$\bar{\lambda}^i \equiv \frac{\sum_{j>i} I_{ij}\lambda_j}{\sum_{j=1}^N I_{ij}\lambda_j}, \qquad \underline{\lambda}^i \equiv \frac{\sum_{j<i} I_{ij}\lambda_j}{\sum_{j=1}^N I_{ij}\lambda_j}$$

Let $\bar{\mu}^i = \sum_{j>i} I_{ij}\mu_j$ (respectively, $\underline{\mu}^i = \sum_{j<i} I_{ij}\mu_j$) be the cumulative Pareto weights on types strictly better (respectively, strictly worse) than i in i's cross-subsidization group. As before, denote the rat race term of type i by $\xi_i(t)$, and the informational rent forfeited by type i by $\kappa_i(t)$:

$$\xi_i(t) \equiv \theta_i(1-t) - \phi_i'(h_i)$$

$$\kappa_i(t) \equiv \theta_i(1-t) - \phi_{i-1}'(h_i)$$

Proposition 7. *The optimal tax for any $N \geq 2$ is:*[13]

$$\frac{t^{AS}}{1-t^{AS}} = \frac{1-\bar{y}+\Delta^{AS}}{\varepsilon_y} \tag{9}$$

with

$$\Delta^{AS} = \frac{1}{y} \sum_{j=1}^N \left[\left(\bar{\mu}^j + \mu_j - \bar{\lambda}^j\right)\left(-\xi_j\right) + \left(\underline{\mu}^j - \underline{\lambda}^j\right)\left(-\kappa_j\right) \right] \varepsilon_{y_j} \frac{y_j/\theta_j}{1-t^{AS}} \tag{10}$$

---

12. Result (i) will hold as long as $\mu \geq \tilde{\mu}$ for some threshold $\tilde{\mu}$, as explained in the general $N \geq 2$ case below.
13. To reduce notational clutter, most dependences on the tax rate t are left implicit.

ε_{y_j}, ε_y, y, and $\bar{y}$ are as defined in Subsection 2.2. The formula in (9) highlights the same basic effects that were at play in the two types case, but allows for all possible equilibria configurations that can endogenously occur. Each type j (except type 1 and those at the bottom of each disjoint cross-subsidization group) now has other types below him with binding incentive constraints, leading to an upward distortion in his hours of work, and a rat race term ξ_j, which tends to push the tax rate up whenever $\bar{\lambda}^j < \mu_j + \bar{\mu}^j$. Put differently, if the government cares sufficiently about the welfare of types higher than j, it will raise the tax to correct for their excessive work.

Similarly, each type (except N and the highest types of each disjoint cross-subsidization group) now receives an informational rent, κ_j. This will tend to increase the tax if the cumulative Pareto weights exceed the fractions in the population, *i.e.* $\underline{\mu}^j > \underline{\lambda}^j$, which is the analog of the condition $\mu > \lambda$ for two types. The intuition lies again in the trade-off between the informational rent earned by j (and, hence, all lower types) and the distortion imposed on all higher types. When the government disproportionately cares about j and lower types, it wants to raise the tax, reducing the hours distortions above j, and simultaneously increasing the informational rent to j and below.

In the limit, an elitist government with $\mu_N = 1$ mostly cares about the rat race terms of high types, while trying to minimize informational rents transferred to low types. At the other extreme, a Rawlsian government with $\mu_1 = 1$ would mostly focus on increasing the transfer and informational rents to low types. If all agents are in the same cross-subsidization group, and the population weights are equal to the Pareto weights, then all redistributive concerns drop out, and only the corrective terms for the rat race remain, yielding a positive Pigouvian tax:

$$\frac{t^{Pigou}}{1-t^{Pigou}} = \frac{-\sum_{j=1}^{N} \lambda_j \xi_j \varepsilon_{y_j} \frac{y_j}{\theta_j y} \frac{1}{1-t^{Pigou}}}{\varepsilon_y}$$

If some worker groups in the economy are not affected by adverse selection, they will only appear in the $1-\bar{y}$ and ε_y terms, but not in Δ^{AS}. Hence, the discrepancy Δ^{AS} between the optimal Second Best and Adverse Selection taxes is directly linked to the fraction of the population affected by adverse selection.

Proposition 8. *At given $\bar{y}$ and ε_y:*

(i) With fully separating contracts, $t^{AS} > t^{SB}$ and $t^{AS} \to t^{SB}$ as the Pareto weights converge to Rawlsian weights ($\mu_1 \to 1$).

(ii) With full cross-subsidization, when all types are in the same cross-subsidization group, if $\underline{\mu}^j - \underline{\lambda}^j \geq 0$ ($\forall j > 1$), $t^{AS} > t^{SB}$.

Condition $\underline{\mu}^j - \underline{\lambda}^j \geq 0$, $\forall j > 1$, is a generalization of the condition $\mu \geq \lambda$ in Proposition 5, with the same intuitions.[14] Proposition 9 is the direct analog of Proposition 6.

Proposition 9. *With a Rawlsian social welfare function, in the same economy, welfare is higher when there is adverse selection than when there is not.*

This result can be extended from Rawlsian weights to weights mostly concentrated on lower productivity agents (in Section 4 a rigorous condition is given). The intuition is that there are two

---

14. As when $\mu < \lambda$ in the two types case, there are intermediate cases involving different configurations of cross-subsidization groups, Pareto weights, and population weights in which t^{AS} and t^{SB} cannot be unambiguously ranked.

sources of welfare gain from adverse selection. First, because of the rat race, revenues raised at any tax level are higher. This effect unambiguously makes all types who are at the bottom of a disjoint cross-subsidization groups better off. All other types directly suffer from their upward distortion in work, but also indirectly benefit from the raised revenue. The net effect is ambiguous, but lower types are more likely to gain on net, especially if higher types are much more productive. Secondly, the optimal tax could be lower with adverse selection for a given set of Pareto weights, benefitting most or all agents, especially whenever the government has highly redistributive preferences (*i.e.* Pareto weights μ_i are concentrated on low θ_i agents), and the optimal tax in the Second Best would have been very high and costly to high types. Overall, there is a range of Pareto weights, mostly concentrated on low types, for which welfare is higher with adverse selection.

4. NON-LINEAR TAXATION

When non-linear income taxation is available, the goal is to compare the full Pareto Frontiers under three informational regimes, illustrated in Figure 1. The first regime is the standard Mirrlees one, in which firms pay workers their marginal products, and the government, who does not know workers' types, sets non-linear taxation subject to truth-telling constraints. The second regime, the "Second Best with Adverse Selection", refers to a situation in which firms do not know workers' types either, but the government sees private market contracts. Alternatively, one can imagine a government-run firm which takes over private firms, and directly sets the hours and pay contracts so as to screen workers.[15] The most novel case, called Regime 3 or "Adverse Selection and unobservable private contracts", is the one in which neither firms nor the government see workers' types, but the government is in addition either unable (or unwilling) to take over private firms or to manipulate labour contracts *directly*. It only observes total realized pay, not the underlying labour contract. Unlike in the Mirrlees case, it must anticipate that workers are not free to choose their hours of work at a given wage, but face a non-linear screening wage schedule. Unlike in the Second Best, it must ensure not only that workers self-select appropriately, but also that *firms* do not deviate by offering different types of contracts in response to taxes. The main conclusion from the linear tax case is still true with non-linear taxation: whenever the government wants to redistribute from high to low types, adverse selection improves welfare.[16] I formulate and solve the general problem with $N \geq 2$ types, but start with the more intuitive and graphically appealing solution for $N = 2$.

4.1. *Characterizing the Mirrlees frontier and the second-best frontier with adverse selection*

4.1.1. Mirrlees frontier. In the traditional Mirrlees framework (regime 1), the government sees total pay y and sets a menu of contracts specifying consumption and hour pairs

---

Regime 1: Mirrlees

Regime 2: Second Best with Adverse Selection

Regime 3: Adverse Selectionand unobservable private contracts

Government

Government

Government
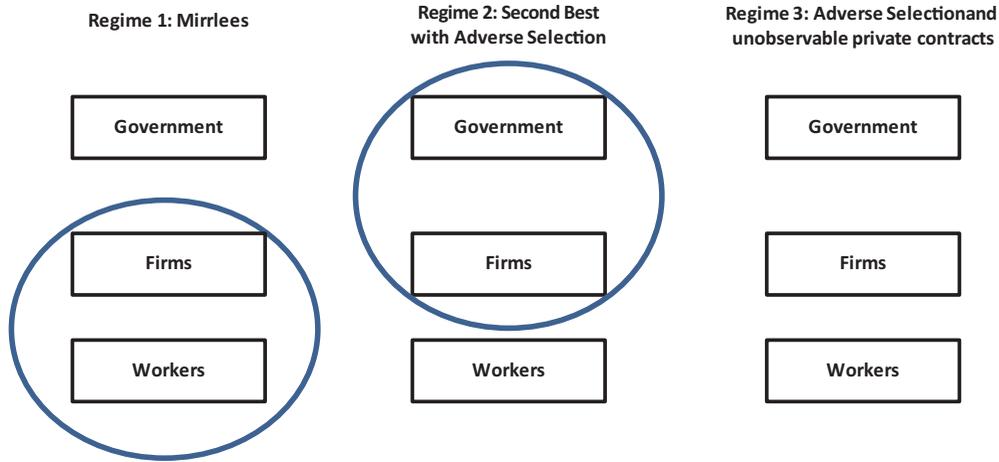
Firms

Firms

Firms

Workers

Workers

Workers

FIGURE 2

Informational Regimes

Agents within the same circle have the same information.

$(\{c_i, h_i\})_{i=1}^N$ to solve the program $(P^{Mirr,N})$:

$$\left(P^{Mirr,N}(\mu)\right): \max_{\{c_i, h_i\}_{i=1}^N} \sum_{i=1}^N \mu_i(c_i - \phi_i(h_i)) \tag{11}$$

$$\left(IC_{i,i+1}\right): c_i - \phi_i(h_i) \geq c_{i+1} - \phi_i\left(\frac{h_{i+1}\theta_{i+1}}{\theta_i}\right) \quad \forall i < N$$

$$\left(IC_{i+1,i}\right): c_{i+1} - \phi_{i+1}(h_{i+1}) \geq c_i - \phi_{i+1}\left(\frac{h_i\theta_i}{\theta_{i+1}}\right) \quad \forall i < N$$

$$(RC): \sum_{i=1}^N \lambda_i c_i \leq \sum_{i=1}^N \lambda_i \theta_i h_i$$

The constraints $\left(IC_{i,i+1}\right)$ (respectively, $\left(IC_{i+1,i}\right)$) are called "upward incentive compatibility constraints" (respectively, "downward incentive compatibility constraints"), as they ensure type i does not pretend he is higher (respectively, lower) productivity. The final constraint (RC) ensures aggregate resources balance. Alternatively, the government's problem can be specified as maximizing the utility of the highest type $(c_N - \phi_N(h_N))$, subject to incentive compatibility constraints, the resource constraint (RC), and minimal utility constraints on all other types. Under this formulation, for $N = 2$, the low type needs to obtain at least some threshold utility u, *i.e.* $c_1 - \phi_1(h_1) \geq u$. By varying u, we can trace out the whole frontier. This latter formulation will be more convenient for the graphical exploration and occasionally used.

The following Proposition characterizes the familiar Mirrlees frontier for two types.[17]

---

17. See also Bierbrauer and Boyer (2010). The result here is reformulated in terms of the relative proportions of types in the population and extended to N types below.

Proposition 10. *The Mirrlees frontier can be characterized by three regions.*

Region 1: When $\mu = \lambda$, none of the incentive constraints are binding, hours of work are efficient, and the Pareto frontier is linear in this region.

Region 2: When $\mu > \lambda$, (IC_{21}) is binding, the low type works inefficiently little, the high type works an efficient number of hours, and the Pareto frontier is strictly concave.

Region 3: When $\mu < \lambda$, (IC_{12}) is binding, the low type works an efficient number of hours, the high type works inefficiently much, and the Pareto frontier is strictly concave.

Whenever the low type is granted a disproportionate Pareto weight ($\mu > \lambda$), the incentive constraint of the high type is binding—the most typical case in the optimal taxation literature. The threshold for μ translates into thresholds for u. In particular, there exist four cut-off levels $u_{\min}^{Mirr} < \underline{u} < \overline{u} < u_{\max}^{Mirr}$, defined in the Appendix, such that the regions are delimited by, respectively, $\underline{u} \leq u \leq \overline{u}$ (Region 1), $\overline{u} \leq u \leq u_{\max}^{Mirr}$ (Region 2), and $u_{\min}^{Mirr} \leq u \leq \underline{u}$ (Region 3). To interpret them, note that in Region 1, work hours are fixed at their efficient levels, and utility is transferred one-for-one (because of quasilinearity) from one type to the other, by varying only consumption. As the consumption of the low type keeps increasing, however, constraint (IC_{21}) will become binding. This point defines $\overline{u}$ as the utility of the low type when work hours are efficient, and constraint (IC_{21}) has just become binding. It is the highest utility level that can be granted to the low type without the high type wanting to mimic him, *i.e.* before hours h_1 have to be distorted. The threshold $\underline{u}$ is defined symmetrically as the utility level of type 1 when, at efficient hours, (IC_{12}) has just become binding.

4.1.2. Second best frontier with adverse selection. In the Second Best case with Adverse Selection, neither firms nor the government know workers' types, but the government can directly set private labour contracts. The program is now:

$$\left(P^{SB,N}(\mu) \right): \max_{\{c_i, h_i\}_{i=1}^N} \sum_{i=1}^{N} \mu_i (c_i - \phi_i(h_i)) \tag{12}$$

$$\left(IC_{i,i+1} \right): c_i - \phi_i(h_i) \geq c_{i+1} - \phi_i(h_{i+1}) \quad \forall i < N$$

$$\left(IC_{i+1,i} \right): c_{i+1} - \phi_{i+1}(h_{i+1}) \geq c_i - \phi_{i+1}(h_i) \quad \forall i < N$$

$$(RC): \sum_i \lambda_i c_i \leq \sum_i \lambda_i \theta_i h_i$$

The constraints look very similar to the ones in the Mirrlees model, with one crucial difference in the downward constraints $\left(IC_{i+1,i} \right)$, which drives all of the subsequent results. In the Mirrlees case, when a high productivity agent deviates to a lower level of income in response to taxes, he can take advantage of his higher productivity to generate the same level of income as the low type, but with less hours of work. In other words, he receives the same *wage per hour* for any level of hours worked. With adverse selection, this is no longer true because the wage, which serves as part of a screening mechanism, is a non-linear function of hours worked. When a high type wants to mislead the government into thinking that he is a lower type, by producing the lower type's income level, he unavoidably also misleads the firm. The firm then pays him the lower type's wage, so that he still needs to work as many hours as the lower type to earn the same income.

This makes the downward deviation less attractive. The following proposition characterizes the Second Best frontier with two types:

Proposition 11.   *The Second Best Frontier with Adverse Selection is characterized by:*
   Region 1: For $\mu = \lambda$, both incentive constraints are slack, both workers work efficient hours, and the Pareto frontier is linear.
   Region 2: For $\mu > \lambda$, (IC_{21}) is binding, the high type works efficient hours, the low type works too little, and the Pareto frontier is strictly concave.
   Region 3: For $\mu < \lambda$, (IC_{12}) is binding, the low type works efficient hours, the high type works too much, and the Pareto frontier is strictly concave.

Again, there exist four thresholds for u, $u_{\min}^{SB} < \underline{u}' < \overline{u}' < u_{\max}^{SB}$ (defined in the Appendix) which delimit the three regions, with an interpretation analogous to the Mirrlees case.

4.1.3. Comparing welfare with and without adverse selection.
Proposition (12) compares the frontiers in the Mirrlees and the Second Best with Adverse Selection cases.

Proposition 12.   *For $\mu > \lambda$, welfare is higher in the Second Best with Adverse Selection regime than in the Mirrlees regime. For $\mu < \lambda$, welfare is lower.*

Hence, whenever the government disproportionately cares about low types relative to their share in the population, welfare is higher under adverse selection. It is also instructive to rephrase this result more visually using the Pareto frontiers. There are two cases depending on whether condition *NL*1 holds or not.
   Condition **NL1**: $\left(\phi_1\left(h_1^*\right) - \phi_1\left(h_2^*\right)\right) \leq \phi_2\left(\frac{h_1^*\theta_1}{\theta_2}\right) - \phi_2\left(h_2^*\right)$ where h_i^* is the first-best effort level for agent of type i, defined by $\phi_i'\left(h_i^*\right) = \theta_i$.
   If condition **NL1** holds, then (i) $\underline{u} \leq \underline{u}' \leq \overline{u} \leq \overline{u}'$, (ii) for $\underline{u} \leq \underline{u}'$, the Mirrlees Pareto frontier is above the Adverse Selection Pareto frontier, (iii) for $\underline{u}' \leq u \leq \overline{u}$, the Adverse Selection and Mirrlees frontiers coincide and are linear, and (iv) for $u \geq \overline{u}$, the Adverse Selection Pareto frontier is above the Mirrlees Pareto frontier. If, condition NL1 does not hold, then (i) $\underline{u} \leq \overline{u} \leq \underline{u}' \leq \overline{u}'$, (ii) for $u \leq \overline{u}$, the Mirrlees Pareto frontier is above the Adverse Selection Pareto frontier, (iii) for $\overline{u} \leq u \leq \underline{u}'$, it is possible to have either frontier above the other one, and (iv) for $u \geq \underline{u}'$, the Adverse Selection Pareto frontier is above the Mirrlees Pareto frontier. Figure 3 illustrates the relative position of the frontiers when **NL1** holds.
   The welfare result can be extended to $N > 2$ types, in a sharper way than with linear taxes. To simplify the proof, we assume that at the optimum, different types are not pooled completely—in the sense of being assigned exactly the same contract.

Proposition 13.   *For $N \geq 2$, if the government cannot observe workers' types, can use non-linear income taxation, and does not pool different types at the optimum, then:*
   (i) if $\sum_{i=1}^{j} \mu_i > \sum_{i=1}^{j} \lambda_i \ \forall j \leq (N-1)$, welfare is higher when there is adverse selection,
   (ii) if $\sum_{i=1}^{j} \mu_i < \sum_{i=1}^{j} \lambda_i \ \forall j \leq (N-1)$, welfare is lower when there is adverse selection.

These conditions on the Pareto weights make explicit how strong or weak the government's redistributive preferences have to be for welfare to be higher or lower with adverse selection. Highly redistributive preferences are those that place higher cumulative welfare weights up to
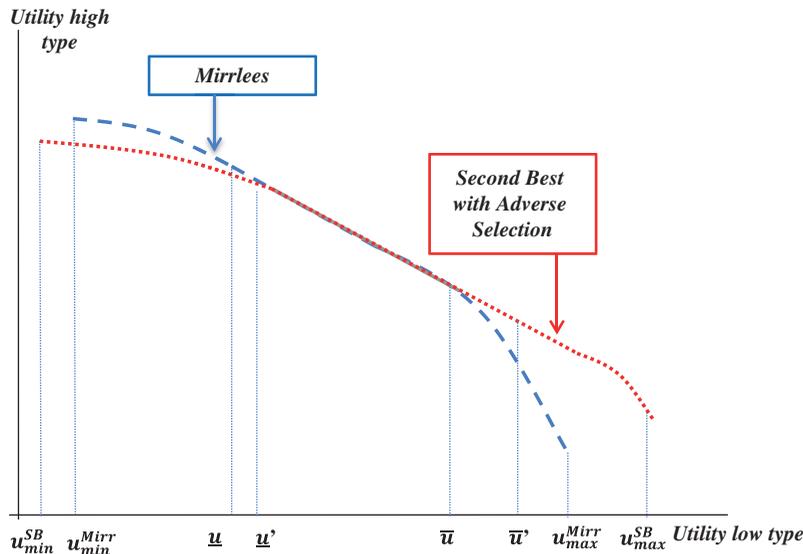
FIGURE 3

Mirrlees and Second-Best with Adverse Selection Frontiers

a given type than the corresponding cumulative proportions in the population.[18] Whenever the government wants to redistribute heavily towards low types, having adverse selection in the labour market helps him do so with a lower deadweight loss. The intuition for this was already captured in the relaxed incentive compatibility constraints. It is now less attractive for any worker to try to lie to the government by pretending to have lower productivity, because, by doing so, he also misleads the firm, and is paid a lower wage per hour. The rat race reduces a worker's capacity to respond negatively to taxes.

4.1.4. Link to the literature on hidden trades and screening.

Several differences with some important papers in the abundant literature on screening and hidden trades explain why adverse selection can be welfare-improving.

Prescott and Townsend (1984) essentially consider a version of regime 2 in their analysis of a Rothschild–Stiglitz insurance market. In the current study, there exists "double" adverse selection, namely between the government and workers, and between firms and workers, which are conflated in Prescott and Townsend. The welfare result here crucially depends on the existence of firms, with potentially more information than the government, as a middle layer between the latter and workers, something which is missing in Prescott and Townsend.[19]

There is also a literature that assesses the welfare effects of improving information when there is adverse selection and that highlights the detrimental redistributive effects and positive

---

18. Note that there is an intermediate range of Pareto weights such that welfare cannot be unambiguously ranked—analogous to the case $\mu = \lambda$ with $N = 2$, when the ranking of the frontiers depended on whether condition NL1 held or not.

19. Prescott and Townsend also note that the Second Best allocation is problematic to implement competitively because of the absence of individualized prices—but these can be imitated by non-linear income taxes. When private firms can act in potential discordance with the government, the decentralization problem arises even with non-linear taxation and thus also requires non-linear prices for *firms* (through the nonlinear payroll taxes considered in the next subsection).

efficiency effects of allowing categorical discrimination in insurance markets (see Hoy, 1984, 1989; Crocker and Snow, 1986, among others). While Schmalensee (1984) cautions against the idea that more information is always welfare improving, there is general agreement that *perfect* information is better than imperfect information, unless information acquisition is costly. The big difference is that, in these papers, any information that firms discover is immediately known to the government as well. Of course, moving from a world with an uninformed social planner to one with a perfectly informed one is welfare improving (this represents a shift from the Second Best with Adverse Selection to the First Best case). But when the government wants to redistribute, while firms have other objectives, improvements in the differential information set of *firms* can be welfare-reducing and increase the efficiency cost of taxation, as represented in the move from the Second Best to the Mirrlees frontier.

Related is the wide literature on hidden trades, in which trades adjust endogenously to government policies as do the private labour contracts here. In Golosov and Tsyvinski (2007) the government tries to insure agents who can engage in hidden trades in a private insurance market. Their private market equilibrium is inefficient because of the externality imposed by a firm's contracts on other firms' contracts through the work incentives of workers. The government can correct for the externality and improve welfare using taxes and subsidies.

Within many such models, the government can create Pareto improvements relative to the competitive equilibrium using tax tools—which might sound identical to the result in this article. Rothschild and Stiglitz (1976) themselves showed that public subsidies for insurance contracts can be Pareto improving, by essentially replacing the cross-subsidy which guarantees efficiency in the MWS setting. Greenwald and Stiglitz (1986) show that asymmetric information generates externalities which typically cause competitive equilibria to be constrained inefficient,[20] and that linear taxation can be Pareto improving. Guesnerie (1998) and Geanakoplos and Polemarchakis (2008) focus on how differential commodity taxation can improve upon a private market equilibrium with hidden trades. But in the current paper, the private market *is already constrained efficient,* and the government, armed with weakly less information than firms, is not generating a Pareto improvement. It merely moves the economy along the Pareto frontier. I take the redistributive preferences of the government as given and study the effects of different market structures to find that a market with adverse selection may be better for welfare when redistributive preferences are high. The distinctive result emerges by looking at adverse selection from the different angle of Mirrleesian optimal taxation.

4.2. *Welfare in the adverse selection and unobservable private contracts regime*

I now turn to regime 3, in which the government no longer sees private labour contracts or cannot directly control them. It can only see the income paid by firms to workers, but neither the underlying menus offered by firms, nor actual work hours. Hence, for each desired allocation $\{c_i,h_i\}$, it needs to set income levels and taxes $\{y_i,T_i\}$, with $T_i=y_i-c_i$, such that (i) firms do not find it profitable to deviate and offer another contract outside of the menu $\{h_i,y_i\}_{i=1}^N$ ("firms' incentive compatibility constraints"), (ii) workers indirectly choose the pair $\{c_i,h_i\}$ destined for them by directly choosing $\{h_i,y_i\}$, and paying income taxes $T_i=T(y_i)$ ("workers' incentive compatibility constraints"), (iii) firms break even on the portfolio of contracts offered, so that

$$\sum_{i=1}^N \lambda_i y_i = \sum_{i=1}^N \lambda_i \theta_i h_i \tag{13}$$

20. The MWS equilibrium satisfies all their stringent conditions for *not* being constrained inefficient.

and (iv) the government's budget constraint holds:

$$\sum_{i=1}^{N} \lambda_i y_i = \sum_{i=1}^{N} \lambda_i c_i \tag{14}$$

If constraint (i) could be omitted, then the constraints for problem $\left(P^{SB,N}\right)$ in (12) would be sufficient and setting $y_i = \theta_i h_i$ would be feasible for all i, so that (13) would be equivalent to (14). But firms too can deviate and offer different contracts than those the government intended. To limit such deviations, the government needs to set prohibitively high taxes (say, 100%) on incomes not in $\{y_i\}_{i=1}^{N}$. Even then firms can still undertake many possible deviations. A profitable deviation must (i) involve only one or several of the allowed incomes in $\{y_i\}_{i=1}^{N}$, (ii) make non-negative profits, even after other contracts rendered unprofitable by it are dropped. Formally, let $\mathcal{P}(\{1,...,N\})$ be the power set of $\{1,...,N\}$. Let $\bar{\theta}_{A_k}$ denote the average productivity within any subset $A_k \in \mathcal{P}(\{1,...,N\})$. A deviation is a collection of K triples $\left\{A_k, y^k, h_{k,A_k}\right\}_{k=1}^{K}$, specifying which groups of agents A_k (potentially singletons) are targeted by a contract offering income level y^k in exchange for an amount of work h_{k,A_k}.[21] Profits from the deviating sets of contracts must be non-negative if accepted by their targeted groups, i.e. $\sum_k h_{k,A_k} \bar{\theta}_{A_k} \geq 0$. The required work hours h_{k,A_k} can be smaller than (respectively, larger than or equal to) $y^k/\bar{\theta}_{A_k}$, in which case we say that group A_k is being cross-subsidized by other groups (respectively, is cross-subsidizing others or breaking even).[22] In general, all possible configurations of pooling groups and pooling income levels need to be considered. This underscores that firms still have a lot of leeway to trick the government by offering new contracts.[23] Proposition 14 shows that despite this hurdle the ranking of the frontiers is the same as when private contracts are observable (regime 2).

Proposition 14. *For $N \geq 2$, if the government cannot observe private labour contracts, the result from Proposition 13 still holds.*

The essence of the proof is that, no matter what deviations firms consider, they can never offer workers as profitable "downward" deviation opportunities as in the Mirrlees case without making losses. Another way to gain intuition is to once more think in terms of the wage per hour. In the Mirrlees case, the wage per hour of type i is equal to his marginal product θ_i for any amount of hours worked. In the Second Best with Adverse Selection, type i would only be paid a wage of $\theta_j < \theta_i$ per hour were he to deviate to a lower income level $y_j < y_i$. With unobservable contracts, the situation is in between those two. Firms can potentially provide a higher wage per hour than θ_j at y_j (the Second Best with Adverse Selection case), but will never be able to pay a wage of θ_i at any income level $y_j \neq y_i$ (the Mirrlees case) without violating workers' incentive constraints. The natural next question is, how closely the government can come to his desired Second Best allocation when private contracts are unobservable, or, the question of implementation.

---

21. To be profitable, naturally, y^k must be part of the allowed income levels $\{y_i\}_{i=1}^{N}$.

22. For example, consider the deviation which consists in pooling workers θ_{i+1} and θ_i at y_i. Using the newly introduced notation, $y^1 = y_i$, $A_1 = \{i, i+1\}$. The required work hours for such a deviation would have to be at least $h_{1,\{i,i+1\}} \geq y_i/\bar{\theta}_{\{i,i+1\}}$ with $\bar{\theta}_{\{i,i+1\}} = \frac{\lambda_i}{\lambda_i + \lambda_{i+1}} \theta_i + \frac{\lambda_{i+1}}{\lambda_i + \lambda_{i+1}} \theta_{i+1}$. This deviation will attract both workers if $y_i = \theta_i h_i$, because then $y_i/\bar{\theta}_{\{i,i+1\}} < h_i$, and, by the binding $(IC_{i+1,i})$, type $i+1$, who was just indifferent between $\{h_{i+1}, y_{i+1}\}$ and $\{h_i, y_i\}$, will strictly prefer the deviating contract $\{h_{1,\{i,i+1\}}, y_i\}$.

23. However, whenever the allocation that the government desires to implement is such that the upward constraints are binding, then then there is no conflict between firms and government. See the proof of Proposition 14.

4.3. *Implementation with adverse selection and unobservable private contracts*

A little thought experiment can highlight the peculiarities of this adverse selection situation, in which the government needs to ensure that firms, as well as workers, comply with its recommendations. Consider all potential choices available to the government. First, it could force firms to break even on each contract separately, and do all of the redistribution itself through taxes, so that $y_i = \theta_i h_i$ and $T_i = \theta_i h_i - c_i$ for all i. At the other extreme, it could let firms do all the redistribution, by assigning gross incomes to be the desired consumption levels, and setting taxes to zero, *i.e.* $y_i = c_i$ and $T_i = 0$. In between those two extremes, given the target allocations $\{c_i, h_i\}_{i=1}^N$, the government could set any incomes $\{y_i\}_{i=1}^N$ satisfying simultaneously (13) and (14).

The question then becomes how much of the redistribution the government can leave to firms. The choices of income and tax levels $\{y_i, T_i\}_{i=1}^N$ determine the profitable deviation opportunities available to firms. In particular, at a fully separating contract ($y_i = \theta_i h_i$), as illustrated above, firms are tempted to pool workers of type θ_i and θ_{i+1} at y_i. On the contrary, if income taxes are zero and firms take care of all the redistribution ($c_i = y_i, \forall i$), then, if agent i is strongly cross-subsidizing others ($\theta_i h_i >> c_i$), the incentives for firms are toward cream-skimming worker i into actuarially fairer contracts. For $N = 2$, the government can perfectly implement any Second-Best allocation despite unobservable labour contracts.

Proposition 15. *With unobservable labour contracts and $N = 2$, the government can implement any allocation from the Second Best with Adverse Selection (when labour contracts are observable) using only nonlinear income taxes.*

The second-best allocation can be implemented by assigning any income levels (y_1, y_2) such that constraints (A9), (A10), and (A11) in the Appendix hold. They could bear only weak relation to consumption levels—leading to a potentially unusual and non-monotone tax system—because, no matter what their assigned *income* levels are, workers only care about their final *consumption* levels. Firms on the other hand only care about the assigned income levels, and are happy to offer any pair satisfying the aforementioned constraints. The tax system is indeterminate even at equilibrium income levels, in the sense that many (y_1, y_2) pairs can sustain the second-best consumption levels. This is not true in the Mirrlees model, where, for any level of recommended hours, earned income is hours times marginal product. Here, it is always necessary for the government to let firms do some of the desired redistribution through cross-subsidization between workers (the conditions on y_2 imply that the high type is paid less than his product).

In general, however, it will not be possible to implement any arbitrary Second-Best allocation. As the number of types increases, the requirements on each income level become more stringent, since any configuration of deviating contracts needs to be ruled out. There is no guarantee that the ranges for the income levels needed to prevent all deviations will contain non-negative values only, and yet allow firms to break even. The Appendix illustrates these difficulties for $N = 3$.

The problem can be resolved by non-linear payroll taxes, levied on firms, and which vary with the income paid to workers. Since the Second Best allocation is resource-compatible, there always exist transfers between firms and governments which allow firms to break even. The optimal payroll taxes compensate for the net profits or losses that firms *would have made* if they offered the income levels recommended by the government. Direct profit taxation is ruled out because of the unobservable output. First, the government determines the admissible income levels, which do not allow firms any profitable deviations. It then announces a menu of payroll taxes (or transfers), $T_i^F = T^F(y_i) = \theta_i h_i - y_i$ as a function of the income paid by firms to workers,

so as to either tax away a net gain or to compensate for a net loss.[24] The two-tier tax system is crucial. A pure income tax system may not allow firms to break even while satisfying their no-deviation constraints. Conversely, a pure payroll tax need not satisfy all the deviation constraints by firms.

Proposition 16. *With $N \geq 2$ types, any allocation $\{c_i, h_i\}_{i=1}^N$ solving problem $(P^{SB,N})$ in the Second Best with Adverse Selection can be implemented, even if private contracts are unobservable, by a sequence of incomes $\{y_i\}_{i=1}^N$, income taxes $\{T_i\}_{i=1}^N$, and payroll taxes $\{T_i^F\}_{i=1}^N$ such that:*

$y_i > \max_{m>i} \bar{\theta}_{\{i,..,m\}} \phi_m^{-1}(c_i - c_m + \phi_m(h_m)) \quad \forall i < N \text{ and } y_N < \phi_{N-1}^{-1}(c_N)\theta_N.$

$T(y) = y - c_i \text{ if } y = y_i \text{ for some } i, \text{ and } T(y) = 2y \text{ otherwise.}$

$T^F(y) = \theta_i h_i - y \text{ if } y = y_i \text{ for some } i, \text{ and } T^F(y) = 2y \text{ otherwise.}$

The addition of payroll taxes to the government's toolbox can allow to implement any allocation from the Second Best with Adverse Selection, for $N \geq 2$, even if private contracts are unobservable. This setting, in which the government cannot observe private contracts or, equivalently, cannot make taxes dependent on the labour contract itself, seems more realistic. Proposition 16 highlights that policies which would incentivize firms to reveal their contracts offered would be helpful only insofar as the government did not have access to the non-linear payroll taxes needed.

5. EMPIRICAL AND POLICY IMPLICATIONS

The findings in this article have two implications for policy design. The first is that estimates of labour supply and taxable income elasticities that are obtained using standard empirical methods may not capture the underlying Marshallian elasticities that are the key inputs for optimal tax calculations. The second involves the normative analysis of tax rates and their design, as well as labour market policies that may affect the degree of adverse selection.

5.1. *Interpretation, measurement, and use of elasticities*

Interpretation of measured elasticities: Because the wage depends on the tax structure in equilibrium, one cannot directly map measured elasticities—the change in hours or income associated with a given change in net wages—to structural elasticities, the fundamental parameters of preferences, without a knowledge of the underlying market structure. For example, the labour supply elasticity of the high type is no longer just a function of his disutility of effort, but also of the low type's preferences, his proportion in the population, and the type of equilibrium. This is because the high type's labour supply is determined by firms in general equilibrium, subject to the low type's reaction.[25] Related, the elasticity of taxable income is not directly mapped into the elasticity of labour supply, since taxable income is also the result of the wage per hour, which is

24. Note that, since the transfers are conditional on money which actually changes hands between firms and workers (assumed to be observable, say, on the paystubs of employees), firms cannot game the system and collude with workers, by pretending to pay some income level, when they in fact pay another one, in order to get a payroll transfer that they could use to pay workers more.

25. This is reminiscent of other papers in which work hours are part of a job "package", and where knowledge of the market structure is required in order to map estimated elasticities to primitives (Altonji and Paxson, 1988; Dickens and Lundberg, 1993; Chetty *et al.*, 2011).

endogenously determined, and depends on taxes (see also Feldstein, 1999; Slemrod and Yitzhaki, 2002; Chetty, 2009).

Estimating taxable income elasticities: The second lesson is that the measurement and estimation of the relevant elasticities is more difficult than typically assumed. Even policy reforms used as "natural experiments" might not be able to correctly capture the elasticities, which are determined in general equilibrium when different groups are interconnected. In the U.S., the largest changes in tax rates have been for the top of the income distribution, typically used as the treatment group, with lower incomes acting as control groups in a difference-in-difference analysis (for comprehensive and critical assessment of this literature see Slemrod, 1998; Giertz, 2004; Saez *et al.*, 2012). Unfortunately, in the presence of adverse selection, a reform affecting high incomes (the high types of the model) will also affect the labour contracts offered to lower incomes (low types), turning the latter into an invalid control group.[26] Paradoxically, the problem is greatest when the groups are more comparable, *i.e.* closer in the income distribution, as they are likely to interact in the same labour market and have interdependent labour contracts.[27]

Use of elasticities for tax design: Taxable income elasticities may not be sufficient statistics for the welfare cost of taxation, as they do not capture the externalities arising from the distortions in labour supply (the rat race and the informational rent effects). This is similar to the limitations of the taxable income elasticity as the sole measure of the efficiency cost of taxation, when there are additional channels through which households react to taxes, such as avoidance and income shifting, which generate fiscal externalities (see Chetty, 2009; Saez *et al.*, 2012).

These three implications could also be valid for other environments in which adverse selection is thought to be a problem, such as health care insurance or markets for used durables.

5.2. *Construction of optimal tax schedules*

This model has three important implications for the praxis of tax policy design, as well as an application to social insurance.

Setting the optimal linear tax rate: At any given set of measured elasticities and income distribution, adverse selection will tend to push tax rates higher, as long as the government wants to redistribute towards lower types (see Propositions 5 and 8). Knowledge of the underlying market structure is hence important for the government; a government armed with estimates of taxable income elasticities and inequality will set the tax rate too low if it wrongly assumes there is no adverse selection.

The strength of this effect depends on the proportion of markets in the economy affected by adverse selection and subject to the same income tax schedule. Markets more prone to adverse selection can be defined among others by age groups (younger workers without a track record), by type of job (more complex, multifaceted jobs), by profession (less automated jobs where worker quality matters more). Ideally, the government could set market specific taxes according to the formulas in this article. Age-dependent taxation could be viewed in this light: if younger

---

26. If there is cross-subsidization, the detrimental effect of tax hikes would be overestimated: As taxes increase and the labour supply of high incomes decreases, the pre-tax incomes of the control group increase (through their informational rents).

27. This is reminiscent of models in which the effects of public policies act through coordinated changes in institutions (here, labour contracts), rather than only individual behaviours (see for example Lindbeck's (1995) "Social Multiplier" idea, or Alesina *et al.* (2005)).

people are more prone to adverse selection, their income tax schedule would optimally be shifted upwards at all income levels relative to older people.[28] But if taxes only condition on income, the optimal tax would be based on an average weighted taxable income elasticity, and the externality term Δ^{AS} would only take into account those affected by adverse selection (see Subsection 3.3).

Adverse Selection is endogenous to government policy: While imperfect information about heterogeneous workers' types might be a common feature in most markets and economies, its consequences, *i.e.* adverse selection *per se* and the use of screening, depend on the structure of the economy, which is endogenous to government action, mostly to regulatory policies.

Statistical discrimination: The government can influence firms' opportunities to engage in statistical discrimination—that is, selecting workers based on characteristics correlated with productivity—through regulations on labour contracts and anti-discrimination laws. For instance, through the lens of firms in my model, women with children are lower productivity workers. If direct discrimination against them is prevented—as is the case in many countries—firms will have to indirectly screen through the labour contract.[29] They might then offer a menu of contracts: a low-paying, part-time contract with shorter hours and more maternity leave, likely to be taken up by working mothers, and a high-paying, full-time contract with overtime bonuses, late-afternoon and week-end meetings, and little parental leave, likely to be taken up by workers without small children.[30]

Firing costs: The more difficult it is to fire a worker once his type is discovered, the costlier adverse selection will be for firms. Kugler and Saint Paul (2004) review empirical studies which find that increasing the stringency of employment protection legislation shifts the composition of employment away from young people and female workers, perceived as being of lower productivity. If however increasing firing costs are coupled with stricter regulation on ex ante statistical discrimination, screening through menus of work contracts becomes more attractive to firms.

Pay structure: Adverse selection cannot occur if there is perfect pay for performance, such as piece rates or purely bonus-based pay, because then the firm would directly reward the worker as a function of his output. On the contrary, contracts specifying the wage as a function of inputs (*e.g.* required number of hours per day, or set of obligatory tasks) are prone to adverse selection since the firm bears the full risk of having hired a low type. Most pay structures are in between these two extremes, including some pay-for-performance as well as some fixed or input-oriented components. By reducing the prevalence of pay-for-performance, the government can shift more of the risk to firms, and increase the consequences of adverse selection for them, thus augmenting their need to engage in screening through labour contracts.[31]

These policies—widely used in the real world—might not have been introduced explicitly to deal with adverse selection, but, once in place, need to be taken into account by the tax system. A government interested in redistribution might choose not to reduce adverse selection in the labour market, even though it had the aforementioned tools to do so.

---

28. Of course, this ignores considerations of age-specific labour supply elasticities or credit constraints for younger people.

29. This assumes that the market does not unravel once direct discrimination is forbidden, *i.e.* firms have access to a screening tool such as hours.

30. This result hinges on the inability of the government to leverage that information to set taxes. If the government could itself use the information extracted by firms in the tax system (for example, base taxes on gender or IQ tests performed by firms), this would pose a very different problem, in the spirit of the "tagging" literature (Akerlof, 1976).

31. Incentive pay of course fulfils a useful role when there is moral hazard.

If adverse selection were a relevant issue, higher redistribution should, all else equal, go hand in hand with more anti-discrimination policies against lower productivity groups (potentially, working mothers or inexperienced youth), more stringent employment protection, and less pay-for-performance. Indeed, it might seem, at least anecdotally, that Continental Europe, with its more rigid labour market, more generous youth and maternal employment policies, and regulated labour contracts that alleviate workers from risk can afford a higher level of redistribution at a lower efficiency cost than the U.S.

The government may hamper screening: If the government wants to redistribute, firms' abilities to indirectly screen through work hours and contracts should not be excessively restricted through, for instance, constraints on hours of work such as the 35-hour work week in France.

Application to social insurance: The model in this article can be directly applied to social insurance—such as health insurance—when it coexists with private insurance providers. If the government chooses a subsidy on health insurance expenditures to maximize a weighted sum of utilities of people with high and low health risks, redistribution toward higher health risks will be facilitated by adverse selection. Intuitively, the government and private insurers have conflicting objectives, which relieve insurees' incentive compatibility constraints. Hence, policies to reduce adverse selection through mandates or regulations may be misguided under some conditions.

6. CONCLUSION

Empirical evidence suggests that there is asymmetric information between firms and workers regarding the latter's ability, and that, accordingly, firms may be screening workers through non-linear compensation contracts. Because work effort is used as a screening device for unobserved talent, labour supply decisions and responses to income taxes are different from those in the traditional optimal taxation literature. Firms have a more active role in setting hours of work and pay than is typically assumed, while workers are more constrained in their labour supply choices.

This article considered the problem of optimal linear and non-linear income taxation when there is adverse selection in the labour market because of workers' private information about their ability. Higher productivity workers are trapped in a rat race, in which they are forced to work excessively, so that firms can screen them from low productivity ones. The nonlinear wage schedule imposed by firms affects the response to taxes, with several implications for optimal tax policy. Most importantly, if the government has sufficiently strong redistributive goals, welfare is higher when there is adverse selection, both with linear and nonlinear taxes. The informational structure of the economy is potentially endogenous to government policies such as bans on discrimination or firing and pay regulations, and a government with strong redistributive goals might find some degree of adverse selection useful. Secondly, the optimal linear tax formula contains additional terms: corrective terms for the rat race distortions, as well as redistributive terms due to the informational rents, whenever firms cross-subsidize workers. At given taxable income elasticities and a given income distribution, taxes are higher with adverse selection whenever the government has highly redistributive preferences. Thirdly, the usual interpretation, estimation, and use of taxable income elasticities may be problematic when labour market contracts are interconnected, and hours of work are determined not just by workers, but also by firms.

At the most general level, the idea of this paper is that there are endogenous private market contracts which react to, and interact with, the government's tax contract. This "contract inside a contract" setup modifies responses to taxes. In future research, it would be interesting to consider the other ways in which wages and labour supply are part of private market contracts, such as

incentive or screening schemes, and their implications for optimal tax policy. The consequences of these labour market imperfections for our interpretation of the estimated taxable income elasticities would be important. It would also be useful to extend the analysis to other labour market imperfections which could affect responses to taxes, among others moral hazard or rent-seeking.

APPENDIX

APPENDIX 1: PROOFS OF SECTION 3

Proof of Proposition (1).

Substituting for y_1 from the budget constraint, $y_1 = \frac{1}{\lambda}[(\lambda\theta_1 h_1 + (1-\lambda)\theta_2 h_2) - (1-\lambda)y_2]$, the maximization problem of the firm is (multipliers are in brackets after the corresponding constraint):

$$\left(P^{MWS}(t)\right): \max_{y_2, h_2, h_1} (1-t)y_2 - \phi_2(h_2)$$

$$(IC_{12}): \frac{1}{\lambda}[(\lambda\theta_1 h_1 + (1-\lambda)\theta_2 h_2) - (1-\lambda)y_2](1-t) - \phi_1(h_1) \geq y_2(1-t) - \phi_1(h_2) \quad [\lambda_{12}]$$

$$(IC_{21}): y_2(1-t) - \phi_2(h_2) \geq \frac{1}{\lambda}[(\lambda\theta_1 h_1 + (1-\lambda)\theta_2 h_2) - (1-\lambda)y_2](1-t) - \phi_2(h_1) \quad [\lambda_{21}]$$

$$(RS_1): \frac{1}{\lambda}[(\lambda\theta_1 h_1 + (1-\lambda)\theta_2 h_2) - (1-\lambda)y_2](1-t) - \phi_1(h_1) \geq u_1^{RS} \quad [\varphi]$$

The general FOCs are:

$$[y_2]: \lambda = \lambda_{12} - \lambda_{21} + (1-\lambda)\varphi$$

$$[h_1]: \left[\theta_1(1-t) - \phi_1'(h_1)\right]\lambda_{12} - \lambda_{21}\left((1-t)\theta_1 - \phi_2'(h_1)\right) + \varphi\left(\theta_1(1-t) - \phi_1'(h_1)\right) = 0$$

$$[h_2]: -\phi_2'(h_2) + \lambda_{12}\left[\frac{(1-\lambda)}{\lambda}\theta_2(1-t) + \phi_1'(h_2)\right] - \lambda_{21}\left[\frac{(1-\lambda)}{\lambda}(1-t)\theta_2 - \phi_2'(h_2)\right] + \frac{1-\lambda}{\lambda}\theta_2(1-t)\varphi = 0$$

Note that whenever $\varphi = 0$, we require that $\lambda_{12} > 0$, or else we would have $\lambda_{21} = -\lambda < 0$, which is not possible. Secondly, the incentive compatibility constraint of the high type (IC_{21}) should never be binding since the firm is trying to maximize that type's utility. Finally, it could happen that the incentive constraint of the low type IC_{12} is not binding, although this will never occur if the cost functions satisfy assumption (2). Whenever IC_{12} is slack, with $\lambda_{12} = 0$, then necessarily $\varphi = \frac{\lambda + \lambda_{21}}{1-\lambda} > 0$, so that the contract is fully separating.

Hence, the three possible cases are: (i) $\varphi > 0$, $\lambda_{12} = \lambda_{21} = 0$ (ii) $\varphi > 0$, $\lambda_{12} > 0$, $\lambda_{21} = 0$ iii) $\varphi = 0$, $\lambda_{12} > 0$, $\lambda_{21} = 0$

(i) Case SB, Second Best: $\varphi > 0$, $\lambda_{12} = \lambda_{21} = 0$. This is immediate, since no incentive compatibility constraint is binding, hence the allocation is as in the second best.

(ii) Case AS1, Separation: $\varphi > 0$, $\lambda_{12} > 0$, $\lambda_{21} = 0$: The necessary equilibrium conditions are:

$$\theta_1 h_1(t)(1-t) = \theta_2 h_2(t)(1-t) - (\phi_1(h_2(t)) - \phi_1(h_1(t))) \tag{A1}$$

$$\phi_2'(h_2) = (\lambda - (1-\lambda)\varphi + \varphi)\frac{(1-\lambda)}{\lambda}\theta_2(1-t) + (\lambda - (1-\lambda)\varphi)\phi_1'(h_2)$$

$$y_1(t) = \theta_1 h_1(t), \, y_2(t) = \theta_2 h_2(t)$$

and the second order condition is: $\left(\lambda - \frac{\phi_2''(h_2)}{\phi_1''(h_2)}\right) < \varphi(1-\lambda)$. To rewrite the characterization as in the main text, let $\delta = \varphi + 1 > 1$, so that: $\phi_2'(h_2) = \theta_2(1-t)\delta(1-\lambda) + (1 - \delta(1-\lambda))\phi_1'(h_2)$. Note that there could be several solutions to equation (A1). If this were the case, the one we pick is the one which yields the highest utility to the type 2. For this case, we need $\varphi > 0$. Since the high type is supplying more labour than $h_2^*(t)$, we need:

$$\lambda > \widetilde{\lambda}(t) = \frac{(1-t)\theta_2 - \phi_2'(h_2)}{(1-t)\theta_2 - \phi_1'(h_2)} \tag{A2}$$

Note that this is indeed a well-defined threshold since h_2 (and hence the right-hand side) does not depend on λ in case AS1. To check that (IC_{21}) is indeed slack, note that from the binding (IC_{12}): $\theta_2 h_2(t)(1-t) - \theta_1 h_1(t)(1-t) = (\phi_1(h_2(t)) - \phi_1(h_1(t)))$. Combined with the Spence Mirrlees single-crossing condition $\phi_1(h_2(t)) - \phi_1(h_1(t)) \geq \phi_2(h_2) - \phi_2(h_1)$, this guarantees that ($IC_{21}$) is slack.

(iii) Case AS2, Cross-subsidization: $\varphi = 0$, $\lambda_{12} > 0$, $\lambda_{21} = 0$. The FOCs become:

$$\lambda = \lambda_{12}, \quad \theta_1(1-t) = \phi_1'(h_1), \quad \phi_2'(h_2) = (1-\lambda)\theta_2(1-t) + \lambda\phi_1'(h_2) \qquad (A3)$$

and the income levels are determined by the binding (IC_{12}):

$$y_1(t) = \lambda\theta_1 h_1(t) + (1-\lambda)\theta_2 h_2(t) - \frac{1-\lambda}{1-t}(\phi_1(h_2(t)) - \phi_1(h_1(t)))$$

$$y_2(t) = \lambda\theta_1 h_1(t) + (1-\lambda)\theta_2 h_2(t) + \frac{\lambda}{(1-t)}(\phi_1(h_2(t)) - \phi_1(h_1(t)))$$

The Second-order condition is: $\lambda\phi_1''(h_2) < \phi_2''(h_2)$. This case can apply only if the last equation in (A3) has a solution at the given t, which requires that $\lambda \leq \widetilde{\lambda}(t)$. To check that (IC_{21}) is indeed slack, note that from the expression for y_2 above: $y_2(1-t) - \phi_2(h_2) = (1-t)\lambda\theta_1 h_1(t) + (1-t)(1-\lambda)\theta_2 h_2(t) + \lambda(\phi_1(h_2(t)) - \phi_1(h_1(t))) - \phi_2(h_2)$ which is this greater than:

$$y_1(1-t) - \phi_1(h_1) = (1-t)\lambda\theta_1 h_1(t) + (1-t)(1-\lambda)\theta_2 h_2(t) - (1-\lambda)(\phi_1(h_2(t)) - \phi_1(h_1(t))) - \phi_2(h_1)$$

because, by the Spence–Mirrlees single crossing condition, and by monotonicity in the hours of work $h_2(t) \geq h_1(t)$, $(\phi_1(h_2(t)) - \phi_1(h_1(t))) \geq \phi_2(h_2) - \phi_2(h_1)$. At *any* solution to (A3) the high type is supplying too much labour relative to the second best case, since: $(1-t)(1-\lambda)\theta_2 = \phi_2'(h_2) - \lambda\phi_1'(h_2) \leq (1-\lambda)\phi_2'(h_2)$, so that $(1-t)\theta_2 \leq \phi_2'(h_2)$, and since the cost function $\phi_2()$ is convex, this implies that $h_2(t) \geq h_2^*(t)$.

(iv) The result is straightforward for the low type. For the high type, consider the two cases separately. In case $AS2$, $(\lambda \leq \widetilde{\lambda}(t))$, $\frac{dh_2}{dt} = -\frac{(1-\lambda)\theta_2}{\phi_2''(h_2) - \lambda\phi_1''(h_2)}$, which is negative by the SOC. In case $AS1$, $(\lambda \geq \widetilde{\lambda}(t))$,

$$\frac{dh_2}{dt} = \frac{\theta_2[\phi_2'(h_2) - \phi_1'(h_2)]}{[(1-t)\theta_2 - \phi_2'(h_2)]\phi_1''(h_2) - [(1-t)\theta_2 - \phi_1'(h_2)]\phi_2''(h_2)}$$

which is again negative by the SOC and the Spence–Mirrlees condition.

(v) In case $AS1$, $(\lambda \geq \widetilde{\lambda}(t))$, h_2 is obtained directly from the binding (IC_{12}), and, hence, does not depend on λ. In case $AS2$, $(\lambda \leq \widetilde{\lambda}(t))$, $dh_2/d\lambda = (\phi_1'(h_2) - \theta_2(1-t))/\phi_2''(h_2) - \lambda\phi_1''(h_2)$. This is positive because the high type's excessive labour supply and the Spence–Mirrlees condition together imply that $\phi_1'(h_2) > \phi_2'(h_2) > \theta_2(1-t)$. The denominator is positive because of the SOC.

Proof of Proposition (3):

In problem $P^{MWS,N}(t)$, let φ_j be the multiplier on the constraint guaranteeing utility $\bar{u}_j$ for type j, $\beta_{j,j+1}$ the multiplier on the incentive constraint ensuring that j does not pretend to be type $j+1$, and δ the multiplier on the resource constraint. The FOCs are:

$$[h_i]: \quad -\phi_i'(h_i)\varphi_i - \beta_{i,i+1}\phi_i'(h_i) + \beta_{i-1,i}\phi_{i-1}'(h_i) + \lambda_i\theta_i\delta = 0$$

$$[y_i]: \quad \varphi_i + \beta_{i,i+1} - \beta_{i-1,i} - \frac{1}{1-t}\delta\lambda_i = 0$$

By convention, normalize $\varphi_N = 1$, $\beta_{0,1} = 0$, and $\beta_{N,N+1} = 0$. Define the modified multipliers $\beta_{i,i+1}/\delta \equiv \beta_i$ and $\eta_i \equiv (\varphi_i + \beta_{i,i+1})/\delta$, so that the FOCs become:

$$\eta_i = \frac{1}{1-t}\lambda_i + \beta_{i-1}, \quad \frac{\eta_i}{\beta_{i-1}} = \frac{[\phi_{i-1}'(h_i) - (1-t)\theta_i]}{[\phi_i'(h_i) - (1-t)\theta_i]}$$

(i) It is immediately clear that for the lowest type $h_1(t) = h_1^*(t)$, $\forall t$, since $\beta_{0,1} = 0$. For all other groups, hours of work are inefficiently high since $[\phi_i'(h_i) - (1-t)\theta_i] > 0$, unless $\varphi_i = \beta_i = 0$, in which case the market splits into two non-interacting groups strictly below and weakly above agent of type $i+1$.

(ii) Whenever $\varphi_j = 0$, firms lose money on all groups $1,...,j$. On the other hand, whenever $\varphi_j > 0$, firms break even on agents $1,...,j$ as a group. To see why, note that firms cannot make strictly positive profits on any subset of agents. Else, it would be possible for some firm to enter, offer slightly lower hours of work at the same pay, and still make a positive profit. Whenever $\varphi_j > 0$, we have $y_j(1-t) - \phi_j(h_j) = \bar{u}_j$. By definition of $\bar{u}_j$, firms can then break even on agents $1,...,j$, and they will not provide those agents with additional utility (all surplus resources could instead be used to increase type N's utility). Whenever $\varphi_j = 0$, $y_j(1-t) - \phi_j(h_j) > \bar{u}_j$, and by the definition of $\bar{u}_j$, this means that the firm is losing money on the subset $1,...,j$. Hence, the cross-subsidization groups referred to in the main text are defined by the break points $k_1,...,k_n$ at which $\varphi_k > 0$.

Generic solution for the equilibrium income levels of each type (no assumptions on which constraints are binding):

Note that whenever $\beta_i = 0$, given that $\eta_i > 0$, we must have $\varphi_i > 0$ (if $IC_{i,i+1}$ is not binding, agent i must be a break agent and is not part of agent $i+1$'s cross-subsidization group). Thus, within a cross-subsidization group, all ICs bind. Let N_i be the highest index of the types who are together with i in a cross-subsidization group (if m and k are in the same cross-subsidization group, then $N_m = N_k$). Symmetrically, let n_i be the smallest index in the cross-subsidization group. Let I_{ij} be the indicator function equal to 1 if j is in i's cross-subsidization group. $\bar{\theta}\bar{h}_i$ is the average production in i's cross-subsidization group, i.e. $\bar{\theta}\bar{h}_i = \sum_j \lambda_j I_{ij} \theta_j h_j / \sum_j \lambda_j I_{ij}$. Let $\bar{\lambda}^i_j = (\lambda_{j+1} + ... + \lambda_{N_i}) / \sum_j I_{ij}\lambda_j$ (respectively, $\underline{\lambda}^i_j = (\lambda_{j-1} + ... + \lambda_{n_j}) / \sum_j I_{ij}\lambda_j$) denote the population weights on those strictly above j (respectively, strictly below j) in i's cross subsidization group. Using the binding ICs and setting the weighted profit in each cross-subsidization group to 0 allows to write each type's income as:

$$y_i = \bar{\theta}\bar{h}_i + \sum_{j=i+1}^{N_i} \frac{\bar{\lambda}^i_{j-1}}{(1-t)} [\phi_{j-1}(h_{j-1}) - \phi_{j-1}(h_j)] - \sum_{m=n_i}^{i-1} \frac{\underline{\lambda}^i_{m+1}}{(1-t)} [\phi_m(h_m) - \phi_m(h_{m+1})] \tag{A4}$$

Proof of Proposition (4) and (7):

The welfare effect on i of changing taxes is:

$$dW_i = -\mu_i y_i dt + \mu_i dt \left\{ \sum_j \frac{dy_i}{dh_j} \frac{dh_j}{dt} (1-t) - \phi'_i(h_i) \frac{dh_i}{dt} \right\}$$

$$\frac{dy_i}{dh_j} = \frac{\lambda_j}{\sum_m I_{im}\lambda_m} \theta_j + \frac{\bar{\lambda}^i_j}{1-t} \phi'_j(h_j) - \frac{\bar{\lambda}^i_{j-1}}{1-t} \phi'_{j-1}(h_j) \quad \text{for } j > i, I_{ij} = 1$$

$$\frac{dy_i}{dh_j} = \frac{\lambda_j}{\sum_m I_{im}\lambda_m} \theta_j + \frac{\underline{\lambda}^i_j}{1-t} \phi'_{j-1}(h_j) - \frac{\underline{\lambda}^i_{j+1}}{1-t} \phi'_j(h_j) \quad \text{for } j < i, I_{ij} = 1$$

$$\frac{dy_i}{dh_j} = \frac{\lambda_j}{\sum_m I_{im}\lambda_m} \theta_j + \frac{\bar{\lambda}^i_i}{(1-t)} \phi'_i(h_i) + \frac{\underline{\lambda}^i_i}{(1-t)} \phi'_{i-1}(h_i) \quad \text{for } j = i, I_{ij} = 1$$

Hence:

$$dW_i = -\mu_i y_i dt + \mu_i dt \left\{ \begin{array}{l} (1-t)\sum_{j=1} I_{ij} \frac{\lambda_j}{\sum_k I_{ik}\lambda_k} \theta_j \frac{dh_j}{dt} + \sum_{j>i} I_{ij} \left(\bar{\lambda}^i_j \phi'_j(h_j) - \bar{\lambda}^i_{j-1} \phi'_{j-1}(h_j) \right) \frac{dh_j}{dt} \\ + \sum_{j<i} I_{ij} \left(\underline{\lambda}^i_j \phi'_{j-1}(h_j) - \underline{\lambda}^i_{j+1} \phi'_j(h_j) \right) \frac{dh_j}{dt} + \left(\bar{\lambda}^i_i \phi'_i(h_i) - \underline{\lambda}^i_i \phi'_{i-1}(h_i) \right) \frac{dh_i}{dt} - \phi'_i(h_i) \frac{dh_i}{dt} \end{array} \right\}$$

Using the definitions for ξ_i and κ_i from the main text, some cumbersome algebra yields:

$$dW_i = -\mu_i y_i dt - \mu_i dt \left\{ \sum_{j>i} \left(\bar{\lambda}^i_j \xi_j - \bar{\lambda}^i_{j-1} \kappa_j \right) \frac{dh_j}{dt} - \sum_{j \le i} \left(\underline{\lambda}^i_{j+1} \xi_j - \underline{\lambda}^i_j \kappa_j \right) \varepsilon_{y_j} \frac{y_j}{1-t} \frac{1}{\theta_j} \right\}$$

(with $\varepsilon_{y_j} \equiv d\log y_j / d\log(1-t)$). Define $\bar{\mu}^i \equiv \sum_{j>i} I_{ij}\mu_j$ (respectively, $\underline{\mu}^i \equiv \sum_{j<i} I_{ij}\mu_j$). Then:

$$\frac{dW}{dt} = \sum_i \frac{dW_i}{dt} = -\sum_j \left[\left(\bar{\lambda}^j_j - \bar{\mu}^j_j - \mu_j \right) \xi_j + \left(\underline{\lambda}^j_j - \underline{\mu}^j_j \right) \kappa_j \right] \varepsilon_{y_j} \frac{y_j}{1-t} \frac{1}{\theta_j}$$

With an abuse in notation, let $\bar{\lambda}^j \equiv \bar{\lambda}^j_j$, $\underline{\lambda}^j \equiv \underline{\lambda}^j_j$, $\bar{\mu}^j \equiv \bar{\mu}^j_j$, and $\underline{\mu}^j \equiv \underline{\mu}^j_j$. The behavioural and mechanical revenue effects are as in Subsection 2.2, $dB = -y \frac{t}{1-t} \varepsilon_y dt$ and $dM = y dt$. Setting $dW + dB + dM = 0$ yields the formula in the proposition.

Proof of Proposition (5):

Both results follow from the fact that $\xi_2 < 0$, $\kappa_2 < 0$, and the terms $\bar{y}$ and ε_y are held constant in the comparison.

Proof of Proposition (6):

(i) When $\mu = 1$: In case $AS1$ $\left(\lambda \ge \widetilde{\lambda}(t) \right)$, welfare is $(1-t)\theta_1 h_1(t) - \phi_1(h_1(t)) + t(\lambda\theta_1 h_1(t) + (1-\lambda)\theta_2 h_2(t))$, which is higher at any tax level since the $h_1(t)$ function is the same while the $h_2(t)$ function is higher at any tax level. In case $AS2$, $\left(\lambda \ge \widetilde{\lambda}(t) \right)$, the difference in welfare with the second best is:

$$W^{SB} - W^{AS2} = t(1-\lambda)\theta_2 \left(h_2^*(t) - h_2(t) \right) + (1-\lambda) \left[(1-t)\theta_1 h_1^*(t) - \phi_1 \left(h_1^*(t) \right) - ((1-t)\theta_2 h_2(t) - \phi_1(h_2(t))) \right] \tag{A5}$$

But by cross-subsidization and the binding (IC_{12}), we have that:

$$\left[(1-t)\theta_1 h_1^*(t) - \phi_1\left(h_1^*(t)\right) - ((1-t)\theta_2 h_2(t) - \phi_1(h_2(t))) \right]$$

$$\leq \left[(1-t)y_1(t) - \phi_1\left(h_1^*(t)\right) - ((1-t)\theta_2 h_2(t) - \phi_1(h_2(t))) \right]$$

$$\leq \left[(1-t)y_1(t) - \phi_1\left(h_1^*(t)\right) - ((1-t)y_2(t) - \phi_1(h_2(t))) \right] = 0$$

so that the last term in (A5) is negative. The first term is negative, since the high type is working more under adverse selection than in the second best. Hence $W^{SB} \leq W^{AS2}$.

(ii) If $\mu = 0$: In both the second best case and the adverse selection cases, the government maximizes the utility of the high type exclusively. Hence, even with adverse selection, it acts as a single agent with the firms. The addition of the incentive compatibility constraint makes the best achievable allocation with adverse selection for the high type worse than in the second best because: (1) hours are distorted relative to the second-best level $h_2^*(t)$ (the level that maximizes the high type's utility), and (2) because pay is weakly lower than the true product, for any level of hours, i.e. $y_2(t) \leq \theta_2 h_2(t)$. The second-best allocation is no longer feasible for the high type with the added incentive compatibility constraint.

(iii) When there is adverse selection, the low type is always working the same amount, yet consuming weakly more due to the higher transfer T and the cross-subsidization transfer. Hence, he must be better off.

Proof of Proposition (8):

At fixed ε_y and $\bar{y}$, t^{SB} and t^{AS} differ only by the term Δ^{AS}, so the result will follow if we show $\Delta^{AS} > 0$.

(i) With fully separating contracts, Δ^{AS} becomes simply $-\frac{1}{y}\sum_{j=1}^{N}\mu_j\xi_j\varepsilon_{y_j}\frac{y_j/\theta_j}{1-t} > 0$.

(ii) With a single cross-subsidization group, $\sum_m I_{jm}\omega_m = \sum_m I_{jm}\lambda_m = 1$. Δ^{AS} can be rewritten as: $\Delta^{AS} = \frac{1}{y}\sum_j\left[\left(\underline{\mu}^j - \underline{\lambda}^j\right)(\xi_j - \kappa_j) - \lambda_j\xi_j\right]\varepsilon_{y_j}\frac{y_j/\theta_j}{1-t}$. In this case $\forall j$, $\xi_j < 0$, $\kappa_j < 0$, $(\xi_j - \kappa_j) = \phi'_{j-1}(h_j) - \phi'_j(h_j) > 0$ (by assumption 1), and hence $\Delta^{AS} > 0$ follows from the condition in the Proposition.

Proof of Proposition (9):

Identical to the proof of Proposition (6), since the lowest type works the same hours, but benefits from more revenues from the increased work of all other types.

APPENDIX 2: PROOFS OF SECTION 4

Proof of Proposition (12):

The proofs of Propositions 10 and 11 (in the Supplementary Appendix) showed that whenever $\mu > \lambda$, IC_{21} is binding both in the Mirrlees and SB with Adverse Selection case. But IC_{21} in the Mirrlees case is more stringent than in the SB with Adverse Selection: namely, for each $\{c_i, h_i\}_{i=1}^2$, $c_1 - \phi_2\left(\frac{h_1\theta_1}{\theta_2}\right) > c_1 - \phi_2(h_1)$. The set of incentive compatible allocations is hence smaller and welfare is lower. The exact opposite applies when $\mu < \lambda$, as then IC_{12}—which is more stringent in the SB with Adverse Selection—is binding.

Proof of Proposition (13):

The problem is reformulated conditional on the set of utilities to be provided to types lower than N, $\mathbf{u} = \{u_i\}_{i=1}^{N-1}$. Multipliers are in brackets.

$$\left(P^{SB,N}(\mathbf{u})\right): \max_{\{c_i, h_i\}} c_N - \phi_N(h_N)$$

$$(IC_{i,i+1}): c_i - \phi_i(h_i) \geq c_{i+1} - \phi_i(h_{i+1}) \quad [\beta_{i,i+1}] \quad i = 1,..,N-1$$

$$(IC_{i+1,i}): c_{i+1} - \phi_{i+1}(h_{i+1}) \geq c_i - \phi_{i+1}(h_i) \quad [\beta_{i+1,i}] \quad i = 1,..,N-1$$

$$(RC): \sum_i \lambda_i c_i \leq \sum_i \lambda_i \theta_i h_i \quad [\delta]$$

$$c_i - \phi_i(h_i) \geq u_i \quad [\gamma_i] \quad i = 1,..,N-1$$

The weights from the main text can be mapped into the multipliers of the utility constraints using: $\mu_i = \gamma_i / \sum_{j=1}^{N}\gamma_j$, and the normalization $\gamma_N = 1$. Note that if the Pareto frontier is a linear hyperplane along some dimensions in some regions,

then the same set of Pareto weights could correspond to several different threshold utilities **u**. The FOCs are:

$$[c_i] : \beta_{i,i+1} + \beta_{i,i-1} - \beta_{i-1,i} - \beta_{i+1,i} - \lambda_i \delta + \gamma_i = 0$$

$$[h_i] : -\phi_i'(h_i)\beta_{i,i+1} - \phi_i'(h_i)\beta_{i,i-1} + \beta_{i-1,i}\phi_{i-1}'(h_i) + \phi_{i+1}'(h_i)\beta_{i+1,i} + \lambda_i\theta_i\delta - \phi_i'(h_i)\gamma_i = 0$$

$$[c_N] : 1 + \beta_{N,N-1} - \beta_{N-1,N} - \lambda_N \delta = 0$$

$$[h_N] : -\phi_N'(h_N) - \beta_{N,N-1}\phi_N'(h_N) + \phi_{N-1}'(h_N)\beta_{N-1,N} + \lambda_N\theta_N\delta = 0$$

$$[c_1] : \beta_{1,2} - \beta_{2,1} - \delta\lambda_1 + \gamma_1 = 0 \qquad [h_1]: -\phi_1'(h_1)\beta_{1,2} + \phi_2'(h_1)\beta_{2,1} + \delta\lambda_1\theta_1 - \phi_1'(h_1)\gamma_1 = 0$$

Lemma 1. *In the Second-Best with Adverse Selection, if $\sum_{i=1}^{j}\mu_i > \sum_{i=1}^{j}\lambda_i$, $\forall j \leq (N-1)$, all downward incentive compatibility constraints $IC_{i+1,i}$ are binding, and all upward incentive compatibility constraints $IC_{i,i+1}$ are slack for all $i \leq N-1$.*

Proof Given the mapping from Pareto weights to multipliers, the condition $\sum_{i=1}^{j}\mu_i > \sum_{i=1}^{j}\lambda_i$, $\forall i \leq N-1$ corresponds to:

$$\sum_{k=i}^{N}\lambda_k > \sum_{k=i}^{N}\frac{\gamma_k}{\sum_{j=1}^{N}\gamma_j}, \forall i \geq 2 \tag{A6}$$

Suppose that, for all $i \geq 2$, condition (A6) holds. First, let us show that there cannot be any upward binding constraint. Start from $i = N-1$, and suppose by contradiction that constraint $IC_{N-1,N}$ is binding, so that $\beta_{N-1,N} > 0$, and $\beta_{N,N-1} = 0$ (since we assumed that pooling is not optimal). The FOC for c_{N-1} would imply that $\beta_{N-1,N} - \beta_{N-2,N-1} = \lambda_{N-1}\delta - \gamma_{N-1}$ (with $\beta_{N-2,N-1}$ either strictly positive or zero), while the FOC for c_N implies that: $1 - \beta_{N-1,N} = \lambda_N\delta$. Adding these two expressions yields:

$$-\beta_{N-2,N-1} = \lambda_N\delta + \lambda_{N-1}\delta - \gamma_{N-1} - 1 \tag{A7}$$

But by the assumption on the parameters in (A6), $\lambda_N\delta + \lambda_{N-1}\delta - \gamma_{N-1} - 1 > 0$, which implies $\beta_{N-2,N-1} < 0$, a contradiction. Hence, $\beta_{N-1,N} = 0$.

Proceeding recursively, consider agent $N-2$ and suppose that constraint $IC_{N-2,N-1}$ binds, so that $\beta_{N-2,N-1} > 0$ and $\beta_{N-1,N-2} = 0$. The FOC for c_{N-2} then implies that $\beta_{N-2,N-1} - \beta_{N-3,N-2} - \lambda_{N-2}\delta + \gamma_{N-2} = 0$ with $\beta_{N-3,N-2}$ either strictly positive or zero. The FOC for c_{N-1} implies $-\beta_{N-2,N-1} - \beta_{N,N-1} - \lambda_{N-1}\delta + \gamma_{N-1} = 0$. The FOC for c_N implies: $\beta_{N,N-1} - \lambda_N\delta + \gamma_N = 0$. Adding these three expressions, we get: $-\beta_{N-3,N-2} = (\lambda_N\delta + \lambda_{N-2}\delta + \lambda_{N-1}\delta - \gamma_N - \gamma_{N-1} - \gamma_{N-2}) > 0$ (by condition (A6)). Hence, $\beta_{N-3,N-2} < 0$, a contradiction. We can continue in this fashion up to type 1 to show that no constraint of the form $IC_{i,i+1}$ binds, hence $\beta_{i,i+1} = 0$.

To show that the downward constraints are not slack but binding, let us now show that it is not optimal to have both $IC_{i,i+1}$ and $IC_{i+1,i}$ slack for some i. Start from agent $i = N-1$ and suppose that $IC_{N-1,N}$ and $IC_{N,N-1}$ are both slack, so that $\beta_{N-1,N} = \beta_{N,N-1} = 0$. Then, the FOC for c_N implies that $1 = \lambda_N\delta$, which violates the strict inequality in (A6). Continuing recursively, suppose that $IC_{N-2,N-1}$ and $IC_{N-1,N-2}$ are both slack. Then, we can decrease c_N and c_{N-1} by the same small amount $dc > 0$ (leaving constraint $IC_{N,N-1}$ unaffected) and increase all c_i for $i \leq N-2$ by the same amount dc' such that the resource constraint is unaffected (this leaves all incentive constraints for types below $N-2$ unaffected as well):

$$dc'(\lambda_1 + ... + \lambda_{N-2}) = dc(\lambda_{N-1} + \lambda_N)$$

The change in welfare from this resource neutral transfer is:

$$dc\left(-(\gamma_N + \gamma_{N-1}) + \frac{(\lambda_{N-1} + \lambda_N)}{(\lambda_1 + ... + \lambda_{N-2})}(\gamma_1 + ... + \gamma_{N-2})\right)$$

Which is positive, from the assumption on parameters in (A6). ‖

Lemma 2. *In the Second-Best with Adverse Selection, if $\sum_{i=1}^{j}\mu_i < \sum_{i=1}^{j}\lambda_i$ $\forall j \leq (N-1)$, all upward incentive compatibility constraints $IC_{i,i+1}$ are binding, and all downward incentive compatibility constraints $IC_{i+1,i}$ are slack for all $i \leq N-1$.*

Proof The proof is symmetric to the one above, starting from the opposite strict inequality than in (A6) and proceeding recursively from type $i = 1$, using the condition on multipliers:

$$\sum_{k=i}^{N}\lambda_k < \sum_{k=i}^{N}\frac{\gamma_k}{\sum_{j=1}^{N}\gamma_j}, i \geq 2 \tag{A8}$$

‖

Lemma 3. *In the Mirrlees regime, if* $\sum_{i=1}^{j}\mu_i > \sum_{i=1}^{j}\lambda_i \ \forall j \le (N-1)$, *all constraints* $IC_{i+1,i}$ *are binding and all constraints* $IC_{i,i+1}$ *are slack for all* $i \le N-1$.

  If $\sum_{i=1}^{j}\mu_i < \sum_{i=1}^{j}\lambda_i \ \forall j \le (N-1)$, *all constraints* $IC_{i,i+1}$ *are binding and all constraints* $IC_{i+1,i}$ *are slack for all* $i \le N-1$.

Proof The program of the Planner with N types can also be reformulated as maximizing the utility of type N, conditional on the utilities of other types being above some thresholds, and subject to the same $(IC_{i,i+1})$, $(IC_{i+1,i})$, and (RC) as in $\left(P^{Mirr,N}(\mu)\right)$ in the text. The proof is then exactly as for Lemmas 1 and 2, since the only thing that differs between the Mirrlees and the Second-Best with Adverse Selection cases is how hours of work enter the incentive compatibility constraints, but the aforementioned proofs only used the FOCs with respect to consumption levels $\{c_i\}_{i=1}^{N}$. ‖

  Thus, when condition (A6) holds, the downward incentive compatibility constraints ($IC_{i+1,i} \ \forall i \le N-1$) are binding both in the Mirrlees and Adverse Selection case. But these constraints are more stringent in $\left(P^{Mirr,N}(\mu)\right)$ than in $\left(P^{SB,N}(\mu)\right)$. Namely, for each $\{c_i,h_i\}$, $\forall i$, $\phi_i\left(\frac{h_{i-1}\theta_{i-1}}{\theta_i}\right) < \phi_i(h_{i-1})$. The incentive compatible set of allocations is hence smaller and welfare is lower. Inversely, when condition (A8) holds, the upward constraints $\left(IC_{i,i+1}, \forall i \le N-1\right)$ are binding, and are more stringent in $\left(P^{SB,N}(\mu)\right)$ than in $\left(P^{Mirr,N}(\mu)\right)$.

Proof of Proposition (14):

  (i) Any second best allocation $\{h_i,c_i\}_{i=1}^{N}$ for which the upward incentive compatibility constraints $\left(IC_{i,i+1}, \forall i \le N-1\right)$ are binding can be implemented by assigning $y_i = \theta_i h_i$ and $T_i = c_i - y_i$ (and prohibitively high tax levels on all other incomes $y \notin \{y_i\}_{i=1}^{N}$). The maximization program of firms then becomes the same as the government's and, since the Second-Best allocation was optimal, there is no possible deviation which could make some type better off without violating the ICs. Hence, welfare in this region is equal to welfare in the Second-Best case, and we showed in Proposition 13 that in this region, the Second-Best frontier is below the Mirrlees frontier.

  (ii) The proof proceeds by finding a lower bound for the Pareto frontier when $\sum_{j=1}^{i}\mu_j > \sum_{j=1}^{i}\lambda_j \ \forall i \le (N-1)$. In this case, Proposition 13 showed that in the Second Best all downward incentive constraints are binding. We already know that, if the only ICs are those from $\left(P^{SB,N}\right)$, the Pareto frontier is above the Mirrlees one in that region. The incentive compatibility constraints in $\left(P^{SB,N}\right)$ are still necessary with unobserved contracts. Are the additional constraints needed to prevent firms from deviating, if any, weaker than those in $\left(P^{Mirr,N}\right)$?

  Suppose that the government artificially strengthens constraint (13) to $y_i = \theta_i h_i \ \forall i$, limiting its choice variables to only h_i and T_i. Starting from $i=N$, we will now rule out all deviations which involve type i being offered a new contract (together with a pool of types), in which he is either cross-subsidizing other deviating agents or earning exactly his product. By doing this for all i, no i can be attracted to a deviating contract in which he is cross-subsidized, since the types made to cross-subsidize him will never join any such deviation.

  Start with agent N and suppose firms try to attract him to a pool with some subset A_k of workers at income level $y^k = y_j = \theta_j h_j (< y_N)$ for some of the available income levels y_j. This requires hours of work of at least $h_{k,A_k} \ge y_j / \bar{\theta}_{A_k} = \left(\theta_j / \bar{\theta}_{A_k}\right) h_j$. In the Mirrlees case, on the other hand, N would have had to work only $y_j / \theta_N = \left(\theta_j / \theta_N\right) h_j < y_j / \bar{\theta}_{A_k} \le h_{k,A_k}$ $\forall A_k$, for the same pay. Thus, ruling out even the most attractive (non-loss making) pool $\left\{A_k, y_j, h_{k,A_k}\right\}$ for type N is strictly easier than to rule out his most attractive deviation in the Mirrlees case: the incentive compatibility constraint for type N has to be strengthened relative to $IC_{N,N-1}$ in program $P^{SB,N}$, but it will never have to be strengthened as much as to become stricter than $IC_{N,N-1}$ in $P^{Mirr,N}$.

  Continue with agent $N-1$. Given that we have ruled out even the most attractive deviation for agent N, any deviation offered to agent $N-1$ must have him as the highest type in any pool he is part of. Again, no matter at which income level y_m the pool occurs, a deviation which is not cross-subsidized by another contract (which we are ruling out for each type) cannot be more profitable than in the Mirrlees case, since agent $N-1$ will necessarily be pooled with lower types and his pay per hour diluted to some $\bar{\theta}_{A_k}$ for some A_k. Continuing recursively this way, we see that for every desired allocation $\{c_i, h_i\}_i$, the downward binding IC for each i will be easier to satisfy than in the Mirrlees case. Removing the artificially imposed constraint $y_i = \theta_i h_i$ will then allow the government to reach even higher social welfare. Hence, *a fortiori*, welfare will be higher than in the Mirrlees case.

Proof of Proposition (15):

  Suppose the government wants to implement the second best consumption and hour levels, $(\{h_i,c_i\})_{i=1}^{2}$, which are characterized by (using the same notation as in the previous section):

$$c_2 = \lambda\theta_1 h_1 + (1-\lambda)\theta_2 h_2 + \lambda(\phi_2(h_2) - \phi_2(h_1))$$

$$c_1 = \lambda\theta_1 h_1 + (1-\lambda)\theta_2 h_2 - (1-\lambda)(\phi_2(h_2) - \phi_2(h_1))$$

$$h_2 = h_2^*, \ (\gamma - \lambda\gamma - \lambda)(\phi_2'(h_1) - \theta_1) + \gamma\theta_1 = \gamma\phi_1'(h_1)$$

Take any arbitrarily assigned income levels (y_1, y_2), such that firms' break-even constraints holds. There are no profitable deviations attracting only the low type, who is already weakly subsidized in the second-best allocation; any such contract making him better off would make a loss. It is also not possible to attract both types to a pooling contract at income level y_2. If it were, then this contract would have made both types better off, be budget feasible for the government, and, hence, would violate the Pareto optimality of the second best allocation. The only deviations that a firm could make are hence:

(1) Offer to pay y_2 for hours of work $h_2 = y_2/\theta_2$. This will be accepted by the high type if y_2, was originally not actuarially fair, *i.e.* $y_2 < \theta_2 h_2$. Under the MWS assumption, other firms will then drop the loss-making cross-subsidization contract. If the low type joins this new contract, it becomes unprofitable. Hence an equilibrium requires that $c_2 - \phi_1\left(\frac{y_2}{\theta_2}\right) \geq 0$ so the low type prefers joining the deviating contract rather than staying out of the market (at utility 0).

(2) Pool both types at y_1 with hours h'_1 such that: $\bar{\theta}_{\{1,2\}} h'_1 = y_1$. The low type prefers this if $h'_1 \leq h_1$ (since he would be working less for the same pay). The constraint needed to render this deviation unprofitable is hence: $h_1 \leq \frac{y_1}{\bar{\theta}_{\{1,2\}}}$. In order to implement the second-best allocation, we thus only need to find two assigned income levels y_1 and y_2 such that (using that from the break even requirement, $y_1 = \frac{c}{\lambda} - \frac{(1-\lambda)}{\lambda} y_2$):

$$\lambda y_1 + (1-\lambda) y_2 = \lambda \theta_1 h_1 + (1-\lambda)\theta_2 h_2 = \lambda c_1 + (1-\lambda)c_2 := c \tag{A9}$$

$$0 \leq c_2 - \phi_1\left(\frac{y_2}{\theta_2}\right) \tag{A10}$$

$$h_1 \leq \frac{\frac{c}{\lambda} - \frac{(1-\lambda)}{\lambda} y_2}{\bar{\theta}_{\{1,2\}}} \tag{A11}$$

where $c := \lambda c_1 + (1-\lambda)c_2$.[32] Thus, it is sufficient to find a y_2 such that: $0 \leq y_2 \leq \min\left\{\phi_1^{-1}(c_2)\theta_2, \frac{c}{(1-\lambda)} - \frac{h_1 \bar{\theta}_{\{1,2\}}}{(1-\lambda)}\lambda\right\}$. Such a level will exist if and only if $\frac{c}{(1-\lambda)} - \frac{h_1\bar{\theta}_{\{1,2\}}}{(1-\lambda)}\lambda \geq 0$, or alternatively, if $c \geq h_1\bar{\theta}_{\{1,2\}}\lambda$. Using the resource constraint, this requires $\lambda\theta_1 h_1 + (1-\lambda)\theta_2 h_2 = \lambda c_1 + (1-\lambda)c_2 \geq h_1\lambda(\lambda\theta_1 + (1-\lambda)\theta_2) \Leftrightarrow h_1\theta_1\lambda + \theta_2(h_2 - h_1\lambda) \geq 0$, which is always true since $h_2 \geq h_1$ (and hence $h_2 \geq \lambda h_1$) in the Second Best.

Proof of Proposition (16):

Illustration with $N = 3$

The income level for type 3, y_3, must be such that firms are not tempted to cream-skim type 3. This requires that, if it occurs, and the unprofitable contracts of type 1 and 2 are dropped in response, type 2 (at least) joins the new contract, *i.e.* $\theta_3 \phi_2^{-1}(c_3) > y_3$. Income level y_2 must be such that no pooling of 2 and 3 can occur, either because 1 would then join the pool and a pool with 1 would not be profitable for 3 to join (*i.e.* $\bar{\theta}_{\{2,3\}}\phi_1^{-1}(c_2) > y_2 > \bar{\theta}_{\{1,2,3\}}h_2$), or because the pooling would not attract type 3 in the first place, *i.e.* $y_2 > h_2\bar{\theta}_{\{2,3\}}$. Income level y_1 must be such that pooling 1 and 2 at 1 is not profitable, and pooling all 3 is impossible because 3 would not join such a pool. Hence, we need $y_1 > \max\left\{h_1\bar{\theta}_{\{1,2\}}, \bar{\theta}_{\{1,2,3\}}\phi_3^{-1}(\phi_3(h_2) - (c_2 - c_1))\right\}$. Finally, all income levels must be non-negative, and firms must break even on average: $\sum_{i=1}^{3}\lambda_i y_i = \sum_{i=1}^{3}\lambda_i \theta_i h_i$. In general, we cannot ensure that there are non-negative income levels $\{y_i\}_{i=1}^{N}$ which will satisfy all these constraints.

General implementation with $N \geq 2$:

We now focus on the implementation of Second Best allocations $\{c_i, h_i\}_{i=1}^{N}$, in which the downward ICs are binding. First, suppose that type $m > i$ is attracted by a deviating contract, paying y_i for $y_i/\bar{\theta}_{\{i,m\}}$ hours of work. Using the binding $(IC_{m,m-1})$, we have $c_i - \phi_m(y_i/\bar{\theta}_{\{i,m\}}) > c_m - \phi_m(h_m) = c_{m-1} - \phi_m(h_{m-1})$. But, by the Spence–Mirrlees single crossing condition in 1, if type θ_m prefers the allocation with less work and less consumption $\{c_i, y_i/\bar{\theta}_{\{i,m\}}\}$ to the one with more work and consumption $\{c_{m-1}, h_{m-1}\}$, then so must type $m-1$. Hence, if m is attracted by this deviation, so is $m-1$. Repeating this argument iteratively, all types $i, .., m-1$ will be attracted if m is. Thus, we only need to consider connected intervals from i to m, for some $m > i$. For each y_i, pick the type for whom a deviation to y_i would be most attractive, and set y_i such that the deviation is not preferred to the type's own allocation:

$$y_i > \max_{m \geq i} \bar{\theta}_{\{i,...,m\}}\phi_m^{-1}(c_i - c_m + \phi_m(h_m)) \quad \forall i < N$$

The constraint on y_i implies that there is no profitable pool that could attract all workers of types i through m to income level y_i, for any m, even if they were just made to work sufficient hours for the firm to break even. By extension, this also implies that no contract could be offered which allowed to cross-subsidize other contracts. For type N, y_N must be set

---

32. To ensure that y_1 is non-negative, we need $y_2 \leq \frac{c}{1-\lambda}$ but this is guaranteed by the third constraint, $y_2 \leq \frac{c}{(1-\lambda)} - \frac{h_1\bar{\theta}_{\{1,2\}}}{(1-\lambda)}\lambda \leq \frac{c}{(1-\lambda)}$.

sufficiently low, so that it would be attractive for type $N-1$ to join if type N was offered an actuarially fair contract with y_N/θ_N hours of work for a pay y_N, *i.e.* we need $y_N < \phi_{N-1}^{-1}(c_N)\theta_N$ (this assumes that N is not alone in a cross-subsidization group. If N is disjoint from other types, there is no profitable deviation for firms to start with). Given these income levels, income taxes are set according to $c_i = y_i - T_i > 0$, $\forall i$, and to 100% for income levels not in the recommended set. The income levels thus specified are potentially very large, and would cause losses for the firms overall. The government can rebate the losses or tax away the profits from each individual contract, by setting a payroll tax schedule $\{T_i^F\}_{i=1}^N$ such that $T_i^F = T^F(y_i) = \theta_i h_i - y_i$, and $T^F(y) = 2y$ for $y \notin \{y_i\}_{i=1}^N$.

Supplementary Data

Supplementary data are available at *Review of Economic Studies* online.

REFERENCES

ACEMOGLU, D. and PISCHKE, J. (1998), "Why do Firms Train? Theory and Evidence", *Quarterly Journal of Economics*, **113**, 79–119.
AKERLOF, G. (1976), "The Economics of Caste and of the Rat Race and Other Woeful Tales", *Quarterly Journal of Economics*, **90**, 599–561.
ALBANESI, S. and SLEET, C. (2006), "Dynamic Optimal Taxation with Private Information", *Review of Economic Studies*, **73**, 1–30.
ALESINA, A., GLAESER, E. and SACERDOTE, B. (2005), "Work and Leisure in the United States and Europe: Why so Different"? (*NBER Macroeconomics Annual 20*).
ALTONJI, J. and PAXSON, C. (1988), "Labour Supply Preferences, Hours Constraints, and Hours-Wage Trade-offs", *Journal of Labour Economics*, **6**, 254–276.
AUTOR, D. (2001), "Why do Temporary Help Firms Provide Free General Skills Training?", *Quarterly Journal of Economics*, **116**, 1409–1448.
BAKER, G., GIBBONS, R. and MURPHY, K. J. (1994), "Subjective Performance Measures in Optimal Incentive Contracts", *Quarterly Journal of Economics*, **109**, 1125–56.
BIERBRAUER, F. and BOYER, P. C. (2010), "The Pareto-Frontier in a Simple Mirrleesian Model of Income Taxation", *Preprints of the Max Planck Institute for Research on Collective Goods*.
CHETTY, R. (2009), "Is the Taxable Income Elasticity Sufficient to Calculate Deadweight Loss? The Implications of Evasion and Avoidance", *American Economic Journal: Economic Policy*, **1**, 31–52.
CHETTY, R., FRIEDMAN, J., OLSEN, T. and PISTAFERRI, L. (2011), "Adjustment Costs, Firm Responses, and Micro vs. Macro Labour Supply Elasticities: Evidence from Danish Tax Records", *Quarterly Journal of Economics*, **126**, 749–804.
CHETTY, R. and SAEZ, E. (2010), "Optimal Taxation and Social Insurance with Endogenous Private Insurance", *American Economic Journal: Economic Policy*, **2**, 85–114.
CROCKER, K. and SNOW, A. (1985), "The Efficiency of Competitive Equilibria in Insurance Markets with Asymmetric Information", *Journal of Public Economics*, **26**, 207–219.
CROCKER, K. and SNOW, A. (1986), "The Efficiency Effects of Categorical Discrimination in the Insurance Industry", *Journal of Political Economy*, **94**, 321–344.
DIAMOND, P. (1998), "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates", *American Economic Review*, **88**, 83–95.
DICKENS, W. and LUNDBERG, S. (1993), "Hours Restrictions and Labour Supply", *International Economic Review*, **34**, 169–92.
FELDSTEIN, M. (1999), "Tax Avoidance and the Deadweight Loss of the Income Tax", *Review of Economics and Statistics*, **81**, 674–680.
GEANAKOPLOS, J. and POLEMARCHAKIS, H. M. (2008), "Pareto Improving Taxes", *Journal of Mathematical Economics*, **44**, 682–696.
GIBBONS, R. and KATZ, L. (1991), "Layoffs and Lemons", *Journal of Labour Economics*, **9**, 351–380.
GIBBONS, R and MURPHY, K. (1992), "Optimal Incentive Contracts in the Presence of Career Concerns: Theory and Evidence", *Journal of Political Economy*, **100**, 468–505.
GIERTZ, S. (2004), "Recent Literature on Taxable Income Elasticities", *US Congressional Budget Office Technical Paper*, 2004–16.
GOLOSOV, M. and TSYVINSKI, A. (2007), "Optimal Taxation with Endogenous Insurance Markets", *Quarterly Journal of Economics*, **122**, 487–534.

GOLOSOV, M., TSYVINSKI A. and WERNING, I. (2006), *New Dynamic Public Finance: A User's Guide* (NBER Macroeconomics Annual).

GREENWALD, B. and STIGLITZ, J. (1986), "Externalities in Economies with Imperfect Information and Incomplete Markets", *The Quarterly Journal of Economics*, **101**, 229–264.

GUESNERIE, R. (1998), *A Contribution to the Pure Theory of Taxation* (Cambridge, UK: Cambridge University Press).

HOLMSTROM, B. (1999), "Managerial Incentive Problems: A Dynamic Perspective", *Review of Economic Studies*, **66**, 169–182.

HOY, M. (1984), "The Impact of Imperfectly Categorizing Risks on Income Inequality and Social Welfare", *The Canadian Journal of Economics*, **17**, 557–568.

HOY, M. (1989), "The Value of Screening Mechanisms under Alternative Insurance Possibilities", *Journal of Public Economics*, **39**, 177–206.

KRUEGER, D. and PERRI, F. (2010), "Public versus Private Risk Sharing", *Journal of Economic Theory*, **146**, 920–956.

KUGLER, A. and SAINT-PAUL, G. (2004), "How do Firing Costs affect Worker Flows in a World with Adverse Selection", *Journal of Labour Economics*, **22**, 553–584.

LAFFONT, J.-J. and MARTIMORT, D. (2001), *The Theory of Incentives: The Principal-Agent Model* (Princeton, NJ: Princeton University Press).

LANDERS, R., REBITZER, J. and TAYLOR, L. (1996), "The Rat Race Redux: Adverse Selection in the Determination of Work Hours in Law Firms", *American Economic Review*, **86**, 329–348.

LINDBECK, A. (1995), "Welfare State Disincentives with Endogenous Habits and Norms", *The Scandinavian Journal of Economics*, **97**, 477–494.

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

MIYAZAKI, H. (1977), "The Rat Race and Internal Labour Markets", *The Bell Journal of Economics*, **8**, 394–418.

PRESCOTT, E. C. and TOWNSEND, R. (1984), "Pareto Optima and Competitive Equilibria with Adverse Selection and Moral Hazard", *Econometrica*, **52**, 21–45.

ROTHSCHILD, M. and STIGLITZ, J. (1976), "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information", *The Quarterly Journal of Economics*, **90**, 629–649.

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

SAEZ, E., SLEMROD, J. and GIERTZ, S. (2012), "The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review", *Journal of Economic Literature*, **50**, 3–50.

SCHMALENSEE, R. (1984), "Imperfect Information and the Equitability of Competitive Prices", *The Quarterly Journal of Economics*, **99**, 441–460.

SPENCE, M. (1978), "Product Differentiation and Performance in Insurance Markets", *Journal of Public Economics*, **10**, 427–447.

SCHEUER, F. (2013a), "Entrepreneurial Taxation with Endogenous Entry", *American Economic Journal: Economic Policy*, forthcoming.

SCHEUER, F. (2013b), "Adverse Selection in Credit Markets and Regressive Profit Taxation", *Journal of Economic Theory*, **148**, 1333–1360.

SLEMROD, J. B. (1998), "Methodological Issues in Measuring and Interpreting Taxable Income Elasticities", *National Tax Journal*, **51**, 773–88.

SLEMROD, J. B. and YITZHAKI, S. (2002), "Tax avoidance, evasion, and administration", in Auerbach, A. and Feldstein, M. (eds) *Handbook of Public Economics* (Amsterdam: North-Holland) 1423–1470.

WEINZIERL, M. (2011), "The Surprising Power of Age-Dependent Taxes", *Review of Economic Studies*, **78**, 1490–1518.

WILSON, C. (1977), "A Model of Insurance markets with Incomplete Information", *Journal of Economic Theory*, **12**, 167–207.