DISCUSSION PAPER SERIES

IZA

IZA DP No. 9154

Optimal Wage Redistribution in the Presence of Adverse Selection in the Labor Market

Spencer Bastani
Tomer Blumkin
Luca Micheletto

June 2015

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Optimal Wage Redistribution in the Presence of Adverse Selection in the Labor Market

Spencer Bastani
Uppsala University, Uppsala Center for Fiscal Studies, Uppsala Center for Labor Studies and CESifo

Tomer Blumkin
Ben Gurion University, CESifo and IZA

Luca Micheletto
University of Milan, Uppsala Center for Fiscal Studies and CESifo

Discussion Paper No. 9154
June 2015

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 9154
June 2015

ABSTRACT

Optimal Wage Redistribution in the Presence of Adverse Selection in the Labor Market*

In this paper we allude to a novel role played by the non-linear income tax system in the presence of adverse selection in the labor market due to asymmetric information between workers and firms. We show that an appropriate choice of the tax schedule enables the government to affect the wage distribution by controlling the transmission of information in the labor market. This represents an additional channel through which the government can foster the pursuit of its redistributive goals.

JEL Classification: D82, H21, J31

Keywords: adverse selection, labor market, optimal taxation, pooling, redistribution

Corresponding author:

Tomer Blumkin
Department of Economics
Ben-Gurion University
P.O. Box 653
Beer-Sheva 8410501
Israel
E-mail: tomerblu@exchange.bgu.ac.il

* We are grateful to two anonymous referees for very helpful comments.

1 Introduction

The modern approach to taxation emphasizes information as the fundamental constraint on public policy. The key assumption in the standard Mirrlees (1971) framework is that the government is unable to observe individual productivities (earning capacities) and hence has to redistribute based on observed levels of income. This might invite high-skilled workers to engage in "mimicking", that is, to reproduce the earned income of a low-skilled worker, in order to benefit from a more lenient tax treatment and thereby derive a higher utility. This means that the income tax must be designed in a way which renders such mimicking unattractive; namely, the income tax must be incentive-compatible.

A standard assumption in the optimal tax literature is that there is symmetric information between workers and firms. In a recent paper, Stantcheva (2014) relaxes this assumption by assuming that firms cannot observe the productivities of workers. Assuming in addition that higher-skilled workers have a weaker taste for leisure, firms have the possibility to screen between high- and low-skilled workers by offering an increased compensation conditional on a higher labor effort. This gives rise to adverse selection where high-skilled agents work more than the efficient amount. Stantcheva (2014) shows that when the government is sufficiently egalitarian, social welfare would be higher in the presence of adverse selection than under the Mirrleesian benchmark with symmetric information. The reason for this is that under adverse selection, as labor contracts cannot be conditioned on (unobserved) labor productivity, high-skilled mimickers are not fully remunerated for their higher earning capacity. That is, they have to work longer hours than under a symmetric information regime in order to reproduce the income of the low-skilled workers. This makes less tempting for the high-skilled workers to mimic their low-skilled counterparts and thereby enhances redistribution.

In principle, the government can promote redistributive goals through two different channels: (i) by changing the income distribution, and/or (ii) by affecting the underlying wage distribution. In the standard Mirrlees (1971) setting, the production technology is assumed to be linear, which implies that the wage distribution is exogenous, thereby leaving no scope for the government to further equity goals through the wage channel. By relaxing the assumption of linearity, the subsequent literature has introduced a role for the income tax to affect the wage distribution. Stiglitz (1982) demonstrates that, when skill types are complements in the production technology, it is socially optimal to marginally subsidize the labor supply of high-skilled workers in order to reduce wage dispersion. This in turn renders the optimal taxes less progressive than under the standard Mirrlees setup with a linear production technology. More recently, Rothschild and Scheuer (2013) have extended the discrete Stiglitz (1982) framework to a continuum of types that differ along a multidimensional skill vector and have allowed for endogenous occupational choices. They show that the redistributive wage channel emphasized by Stiglitz carries over to the more general setting. However, the additional features associated with the occupational choice margin mitigate the general equilibrium effects and make the optimal taxes more progressive (but still less progressive than under the standard Mirrleesian setting).[1]

In this paper we connect the analysis of Stantcheva (2014) with the aforementioned strand of the literature, which emphasizes the wage channel for redistribution. Stantcheva considers a standard linear production technology and restricts attention to separating allocations, in which each type of worker is offered a distinct consumption-labor bundle. In a separating equilibrium, when the relevant equilibrium concept is of the Rothschild-Stiglitz (*RS*) type, each worker is remunerated according to his/her marginal productivity. This implies that no redistribution is carried out through the wage channel.[2] Employing a similar framework, we show that the government can, by choosing an appropriate tax system, make it more difficult for firms to engage in screening, and thereby implement wage pooling.[3] When designing the optimal redistributive policy the government has to balance the efficiency gains from screening, associated with implementing a separating allocation, and the equity gains from wage

[1]The above mentioned literature has limited attention to the role of income taxation in affecting the wage distribution by relying on the complementarity between production factors. Cremer et al. (2011) consider a setting with a linear production technology (that is, no complementarities) but where the government can supplement the nonlinear income tax with education policy that affects the wage distribution. They show that, for a given mean of the individual earning abilities, social welfare is a convex function of the variance of the individual earning abilities. In particular, the most unequal distribution of wages is desirable from the standpoint of social welfare maximisation when the permissible degree of wage differentiation is large. When the permissible degree of wage differentiation is small, they demonstrate that an equal-wage outcome (which obviates the redistributive role of income taxation) may be socially desirable. Notice that in their model, independently on whether redistribution is achieved through the tax or the educational policy, each worker is always paid a wage rate equal to his/her marginal productivity. In the model developed in our paper, productivities are exogenously given and the government has no instrument to affect them. Moreover, in our setting, the difference between tax-and-transfer redistribution and wage redistribution is that in the second case a wedge is created between the gross wage rate paid by firms to a given worker and the worker's marginal productivity.

[2]An alternative equilibrium concept developed in the literature is the so-called Miyazaki-Wilson-Spence (*MWS*) equilibrium [following Miyazaki (1977), Wilson (1977) and Spence (1978)]. The crucial difference between the *RS* and *MWS* equilibrium concepts is in the degree of cross-subsidization across types that derives in a separating equilibrium given the permissible forms of contracts that can be signed between the firms and the workers. Under the *RS* equilibrium concept each contract offered in equilibrium has to break even separately (yielding zero profits); whereas, under the alternative *MWS* equilibrium concept, firms make zero profits on their overall portfolio of contracts.

[3]In her contribution, Stantcheva (2014) mostly relies on the *MWS* equilibrium concept and confines to the online Appendix the analysis of the *RS* equilibrium. Our focus on the *RS* equilibrium is justified by our purpose of comparing income (ex-post) redistribution and wage (ex-ante) redistribution as alternative redistributive channels that can be exploited by the government through the design of an appropriate tax policy. As we will see, this comparison will be undertaken by contrasting a separating tax regime (where, due to our focus on a *RS* equilibrium concept, wage redistribution cannot exist) with a pooling tax regime.

pooling.

We first consider a two-type model and show that when optimizing the tax schedule the government can implement a pooling allocation with full wage equalization. This pooling allocation turns out to be socially superior when both the differences in productivities and the differences in labor-leisure preferences are not too large. In a setting with two types we are limited to comparing the welfare effects of the separating equilibrium and the equilibrium where both types are pooled together. To gain insights into the possibility of having an equilibrium involving partial pooling of types we then consider two extensions of our benchmark model. First, we analyze a three-type model and provide numerical examples where the optimal policy of the government is to implement a hybrid equilibrium (where two out of three types are pooled together). Second, we show analytically that in a model with a continuum of types, full pooling is never optimal and that some redistribution through the wage channel is always desirable.

The general message of our analysis is that one can highlight a novel role played by the non-linear income tax system in the presence of adverse selection in the labor market due to asymmetric information between workers and firms. Under symmetric information, firms observe workers' productivities and therefore remunerate each worker according to his/her marginal productivity in a competitive labor market. Under asymmetric information, however, the translation of differences in productivities into differences in wage rates hinges on the mechanism by which workers and firms exchange information. In line with Stantcheva (2014) we focus on the particular mechanism in which firms screen workers through non-linear labor contracts and work effort is used as a screening device for unobserved talent.[4] In this case we show that the optimal nonlinear income tax is not only used to redistribute income, but also enables the government to perform redistribution through the wage channel as the nonlinear tax allows the government to affect the transmission of information (and thereby the remuneration of workers) in the labor market. Under certain circumstances, this additional role for the income tax enhances the redistributive capacity of the government.

The screening device that we focus on is empirically justified by the 'rat-races' occurring in many work places where employees are required to work long hours in order to be eligible for a high compensation. For example, Landers et al. (1996) report on the long working hours required by lawyers for them to become eligible as partners in law-firms. Another example is the segmentation of the labor market into career jobs

[4]Notice that screening through nonlinear labor contracts is an example of an indirect screening mechanism which is relevant when other (direct) screening measures based on observable characteristics (such as age or gender) are rendered infeasible by anti-discrimination legislation. Another example of an indirect screening mechanism is job training programs. The screening through nonlinear labor contracts is most relevant at the stage of hiring or the early stages of employment, as later on, firms get familiarized with their workers' abilities and may choose to promote/remunerate them accordingly.

and part-time jobs to allow the separation between workers who are more family oriented versus those who are more career oriented. Such screening through nonlinear labor contracts may arise if discrimination based on family status/number of children is ruled out by antidiscrimination legislation.

Clearly, a key assumption in our framework is that there exists asymmetric information in the labor market. This is the case whenever incumbent firms have more information about worker quality than other potential employers do. This is supported by Gibbons and Katz (1991) who found that workers who were laid off experienced a larger wage loss than workers who were displaced by plant closings. Acemoglu and Pischke (1998) find evidence of adverse selection in the labor market for German apprentices. More recently, Kahn (2013) estimates a model of employer learning using data on nationally representative sample of workers and finds strong support for asymmetric information in the labor market.

The paper is organized as follows. In section 2 we first consider the two-type model. In section 3 we generalize our model to three types and in section 4 we discuss how our model generalizes to a continuum of types. Finally, section 5 offers concluding remarks.

2 The Two-Type Case

We use the simplest possible model with just the key ingredients necessary to demonstrate our point. Consider an economy with low- and high-skilled workers (indexed by l and h, respectively) that produce a single consumption good (the price of which is normalized to unity) using a production technology exhibiting constant returns to scale and perfect substitutability between the two skill levels. We normalize the workers' population to a unit measure and let the measures of low- and high-skilled workers be given, respectively, by γ^l and γ^h.

Let the productivity (which is equal to the hourly wage rate under a perfectly competitive labor market) of a low- and a high-skilled worker be denoted by w^l and w^h respectively, where $w^h > w^l \geq 0$.

The two types of workers differ in their labor-leisure preferences. The utility of the high-skilled workers is given by $u^h \equiv c^h - g(n^h)$ where c represents consumption, n represents working hours, and where, $g(0) = 0$, $g^{'} > 0$, $g^{''} > 0$ and $\lim_{n \to 0} g^{'}(n) = 0$. The utility of the low-skilled workers is given by $u^l \equiv c^l - kg(n^l)$, where $k > 1$. That is, low-skilled workers incur a higher disutility (both total and marginal) from work relative to their high-skilled counterparts for the same working hours supplied.[5]

[5]The quasi-linear specification, which is common in the literature [see Diamond (1998) and Salanié (2011) amongst others], is invoked for tractability. Our qualitative results remain robust to incorporation of income effects on labor supply.

2.1 Labor Market Equilibrium under Asymmetric Information

We deviate from the standard Mirrlees (1971) framework and assume that firms cannot observe the types of their workers when signing a labor contract.[6] An alternative interpretation of the setting would be that firms do observe the types but are not allowed to offer separate contracts due to anti-discrimination legislation. As is well known since the seminal contribution of Rothschild and Stiglitz (1976), adverse selection may arise in such contexts. Before turning to present the optimal tax problem we briefly characterize the *laissez-faire* equilibrium (adopting the Rothschild and Stiglitz (*RS*) equilibrium concept) and demonstrate the resulting market failure.

2.2 The *RS* Equilibrium

A typical labor contract specifies the number of working hours, n, and the corresponding total compensation, c. Crucially, a labor contract cannot be made conditional on the type of worker, which is assumed to be private information of the worker and hence unobservable by the hiring firm. The *RS* equilibrium is defined by a set of labor contracts satisfying two properties: (i) firms make non-negative profits on each contract; and, (ii) there is no other potential contract that would yield non-negative profits if offered (in addition to the equilibrium set of contracts).

Having defined the equilibrium, we turn next to show that the *laissez-faire* allocation under symmetric information may become non incentive-compatible in the presence of asymmetric information.

In a competitive labor market with symmetric information each worker would be remunerated according to his/her earning capacity. Formally, a competitive equilibrium allocation is given by the two consumption-labor bundles $(c^{i*}, n^{i*}); i = l, h$, which satisfy:

$$c^{i*} = w^i n^{i*}; i = l, h, \tag{1}$$

$$w^i = k^i g'(n^{i*}); \quad i = l, h; \quad k^l = k \text{ and } k^h = 1. \tag{2}$$

The first condition is the individual budget constraint driven by the zero-profit free entry requirement, whereas the second condition states that workers optimally choose their labor supply by equating the marginal disutility from work with their hourly wage rate.

To be implementable in a setting with asymmetric information between firms and workers, the set of allocations defined in (1)-(2) must also satisfy two incentive compat-

[6] As we already mentioned, this was also a key assumption in Stantcheva's (2014) contribution. A more thorough discussion of the differences between her and our set-up will be provided at the end of Section 2.

ibility constraints. These constraints ensure that workers have no incentives to mimic each other. Formally,

$$c^{l*} - kg(n^{l*}) \geq c^{h*} - kg(n^{h*}), \tag{3}$$

$$c^{h*} - g(n^{h*}) \geq c^{l*} - g(n^{l*}). \tag{4}$$

Notice that a high-skilled worker has no incentive to mimic his/her low-skilled counterpart due to the higher hourly wage rate reflected in his/her symmetric information *laissez-faire* contract. Thus, the incentive compatibility constraint (4) for the high-skilled worker is slack.[7] However, the incentive compatibility constraint associated with the low-skilled worker may be violated. To see this, reformulate the incentive constraint associated with the low-skilled worker by substituting for w^i and c^{i*} from (1) and (2) into (3) to obtain:

$$kg'(n^{l*})n^{l*} - kg(n^{l*}) \geq g'(n^{h*})n^{h*} - kg(n^{h*}). \tag{5}$$

Consider now the limiting case where k converges to 1. Re-formulating (5) by taking the limit yields:

$$g'(n^{l*})n^{l*} - g(n^{l*}) \geq g'(n^{h*})n^{h*} - g(n^{h*}) \Longleftrightarrow H(n^{l*}) \geq H(n^{h*}), \tag{6}$$

where $H(n) \equiv g'(n)\, n - g(n)$.

Differentiation with respect to n yields $H'(n) = g''(n)n > 0$, where the inequality follows by the strict convexity of g. Thus, by virtue of (6), for the incentive constraint associated with the low-skilled worker to hold it is necessary that $n^{l*} \geq n^{h*}$. However, by virtue of (2), the strict convexity of g and the fact that $w^h > w^l$ it follows that $n^{h*} > n^{l*}$. Thus, by continuity considerations, for k sufficiently close to unity, the incentive constraint associated with the low-skilled workers is violated and, hence, the symmetric information *laissez-faire* allocation is not incentive compatible.[8]

When condition (5) is violated, the *laissez-faire RS* equilibrium allocation is given by

[7]Formally, $c^{h*} - g(n^{h*}) > w^h n^{l*} - g(n^{l*}) = \frac{w^h}{w^l} c^{l*} - g(n^{l*}) > c^{l*} - g(n^{l*})$, where the first inequality follows by virtue of the strict convexity of g, which implies that n^{h*} is the (unique) optimal labor supply choice of type-h workers under the symmetric information regime, the equality follows from the budget constraint in (1) and the latter inequality follows as $w^h > w^l$.

[8]Notice that when k is sufficiently large, namely when the disutility from work entailed by the low-skilled workers is sufficiently high, the symmetric information *laissez-faire* allocation would be incentive compatible (and hence the asymmetric information *laissez-faire* allocation would be first-best efficient).

the two consumption-labor bundles $(c^{i**}, n^{i**}); i = l, h$, which satisfy:

$$c^{i**} = w^i n^{i**}; i = l, h, \tag{7}$$
$$w^l = kg'(n^{l**}), \tag{8}$$
$$c^{l**} - kg(n^{l**}) = c^{h**} - kg(n^{h**}). \tag{9}$$

Comparing the equilibrium allocations under symmetric and asymmetric information [given, respectively, by conditions (1)-(2) and (7)-(9)] reveals that the labor supply condition for type-h workers under the symmetric information regime is being replaced by the binding incentive constraint of type-l workers under the asymmetric information regime, which implicitly defines the labor contract offered to type-h workers in the asymmetric information equilibrium. Under asymmetric information low-skilled workers are still offered their efficient (symmetric information) allocation, $(n^{l**} = n^{l*})$, whereas high-skilled workers' labor supply choice is distorted, as they work more hours than under their efficient allocation $(n^{h**} > n^{h*})$. This enables the firms to reduce the information-rent associated with type-l workers and render the allocation incentive compatible.

Two final remarks are in order. First, notice that in the *RS* setting the separating equilibrium characterized by conditions (7)-(9) exists when the fraction of low-skilled workers is sufficiently high [see Rothscild and Stiglitz (1976)]. A pooling equilibrium does not exist, as firms can engage in "cream-skimming", by offering a contract that would attract only type-h workers and yield positive profits. Notice further that our assumption that workers differ not only in their earning capacity [as in Mirrlees (1971)] but also in their labor-leisure preferences is essential for the existence of a separating equilibrium which relies on the ability of firms to screen between workers based on their differences in preferences (higher-skilled workers exhibit a weaker taste for leisure). In the absence of such screening capacity the only equilibrium that would sustain under asymmetric information would be one where all workers would be pooled together and each paid an hourly wage rate equal to the average productivity.

2.3 The Government's Problem

The government is seeking to design a non-linear tax-and-transfer system, which maximizes a welfare function given by a weighted average of the utilities of the two types of workers. Formally,

$$W = \sum_i \beta^i u^i; i = l, h, \quad \text{where} \quad \sum_i \beta^i = 1 \quad \text{and} \quad \gamma^l < \beta^l \leq 1. \tag{10}$$

The fact that the weight assigned to type-l workers strictly exceeds their share in the population reflects the strictly egalitarian preferences of the government with respect to redistribution.

We follow Stantcheva (2014) by considering the regime referred to as "adverse selection with unobservable private contracts" in which neither the firm nor the government observes workers' types and, in addition, the government has no control over labor contracts.[9]

In the previous subsection we have argued that a pooling equilibrium cannot exist in the *RS* setting under *laissez faire*, as it will invite "cream-skimming". However, in the presence of government intervention the government may block the possibility for such "cream-skimming" by an appropriate choice of the tax system. In the absence of screening, all workers will be pooled together in the (n,c)-space; they all receive the same wage rate, equal to the average productivity, and earn the same income. When designing the optimal re-distributive policy, therefore, the government has to account for the trade-off between the efficiency gains from screening associated with implementing a separating equilibrium, which induces high-skilled workers to work more hours than their low-skilled counterparts, and the equity gains from wage pooling associated with implementing a pooling equilibrium.

2.4 The Separating Equilibrium Regime

Invoking the self-selection approach common in the optimal tax literature, a non-linear income tax schedule is given by the tuple $\{y^i, T^i\}$; $i = l, h$, where y denotes gross income and T denotes the associated tax (possibly negative) which satisfies the balanced budget constraint: $\sum_i \gamma^i T^i = 0$, where we assume with no loss in generality that the government has no exogenous revenue needs. To abbreviate notation, letting $T^l \equiv -T$, it follows, using the balanced budget constraint, that $T^h = \frac{\gamma^l}{\gamma^h} \cdot T$.

We turn next to characterize the separating *RS* equilibrium given the tax schedule in place. Notice that the only margin of maneuver of firms is to set the working hours demanded, denoted by n^i (with $i = l, h$), for each level of gross income. As firms do not observe workers' types, the number of working hours demanded for each level of gross income will be independent of the worker's type. A separating equilibrium has to satisfy the following set of conditions.

First of all, the resulting allocation has to be incentive-compatible; namely, it has to satisfy the following two incentive constraints associated with type-l and type-h,

[9] Even though neither the firm nor the government knows "who is who", they both possess the statistical information on the distribution of types in the population and know the functional form of the utility functions.

respectively:

$$y^l + T - kg(n^l) \geq y^h - \frac{\gamma^l}{\gamma^h} \cdot T - kg(n^h), \qquad (\text{IC}^l)$$

$$y^h - \frac{\gamma^l}{\gamma^h} \cdot T - g(n^h) \geq y^l + T - g\left(\frac{y^l}{\sum_i \gamma^i w^i}\right). \qquad (\text{IC}^h)$$

The incentive constraint associated with type-l (IC^l) is rather standard, the only difference from the Mirrlees (1971) setting being that the mimicking type-l agent is working the same number of hours as his/her type-h counterpart.

The incentive constraint associated with type-h is instead non-standard [the argument is similar to Stantcheva (2014)]. Notice that in the second term on the right-hand side of the inequality in (IC^h), we replaced n^l with the term $\frac{y^l}{\sum_i \gamma^i w^i}$. Due to the requirement that firms earn non-negative profits in a separating equilibrium, n^l is necessarily bounded from below by the term $\frac{y^l}{w^l}$. Offering any lower level of n^l [assuming (IC^h) is not violated] would yield negative profits, provided that the contract is not chosen by the type-h workers too. However, by offering a sufficiently low level of n^l, the firm can attract also type-h workers, who are more productive than their type-l counterparts. Thus, although the firm suffers losses on type-l workers, it is compensated, by gaining on their type-h counterparts. The term $\frac{y^l}{\sum_i \gamma^i w^i}$ defines the level of n that would yield the firm zero profits in a pooling equilibrium associated with the income level y^l. That is, the term defines a lower bound on n that can be offered by the firm under such pooling equilibrium. Offering such a pooling equilibrium contract would be more attractive for both types of workers than the separating contract associated with y^l, as $\frac{y^l}{\sum_i \gamma^i w^i} < \frac{y^l}{w^l} \leq n^l$. Incentive compatibility then requires that a type-h worker weakly prefers his/her contract to mimicking type-l and getting the pooling contract that yields zero profits to the firm. Notice that any alternative pooling contract that would yield positive profits would require longer working hours and would hence be clearly dominated by the type-h separating equilibrium contract.[10]

In addition to being incentive compatible, the resulting allocation has to satisfy two zero-profit conditions associated with the contracts offered to type-l and type-h

[10]We would like to make the following technical remark. The binding incentive constraint associated with type-h workers implies that a type-h worker is indifferent between his/her separating contract and the pooling contract associated with y^l that yields zero profits. In principle, one might consider the possibility of firms offering this pooling contract as being a violation of property (ii) in the definition of the RS equilibrium on page 6. However, notice that the pooling contract is strictly preferred by *all* low-skilled workers to their separating contract. Thus, in order for the pooling contract to yield zero profits, *all* high-skilled workers would have to choose the pooling contract even though they are indifferent between the pooling and their separating contracts. We rule out this implausible possibility.

workers, respectively:

$$y^l = w^l n^l, \qquad (\mathrm{ZP}^l)$$
$$y^h = w^h n^h. \qquad (\mathrm{ZP}^h)$$

Condition (ZP^l) requires that a contract offered to a type-*l* worker would yield zero profits. Recall that according to the definition of the *RS* equilibrium any contract has to yield non-negative profits (but not necessarily zero profits). However, if the condition is violated and holds as a strict inequality, a firm can offer a contract that slightly reduces n^l. Clearly, this new contract would attract type-*l* workers and would yield positive profits, by continuity considerations; therefore, it would not be an equilibrium. In case type-*h* workers find this contract attractive as well, the firm's profits will further increase (as type-*h* workers are more productive than their type-*l* counterparts).

We turn next to condition (ZP^h). Notice that this condition may hold as a strict inequality. In such a case an additional (complementary-slackness) condition has to be satisfied, namely, requiring that the incentive constraint associated with type-*l* workers is binding. To see this, notice that when condition (IC^l) is slack and the zero profit condition for type-*h* workers does not hold, a firm can offer a new contract that slightly decreases n^h, attracting only type-*h* workers and yielding positive profits, by continuity considerations. When (IC^l) is binding, however, such a decrease in n^h will also attract type-*l* workers. The resulting pooling allocation has to be unprofitable to sustain the equilibrium. Thus, we need to add the condition $n^h \sum_i \gamma^i w^i \leq y^h$ which implies that for any $n < n^h$, the pooling contract would yield negative profits. However, any profits earned by a firm hiring type-*h* workers can be taxed away at a confiscatory 100 percent tax rate and paid back to the workers in an incentive compatible manner that renders both types of workers strictly better off. The modified profit cum income tax system is equivalent to an income tax system where $y^h = w^h n^h$, namely, the zero profit condition is satisfied, and the income tax paid by type-*h* workers is augmented by the amount paid by the firm as profit taxes. Consequently, we will henceforth assume that condition (ZP^h) holds with no loss in generality.

Reformulating the utilities of the two types of workers given the tax schedule in place, employing the two zero-profit conditions, and substituting into the welfare function in (10) yields:

$$W = \beta^l \left[y^l + T - kg\left(\frac{y^l}{w^l}\right) \right] + \beta^h \left[y^h - \frac{\gamma^l}{\gamma^h} \cdot T - g\left(\frac{y^h}{w^h}\right) \right]. \qquad (11)$$

Substituting from the zero profit conditions in (ZP^l) and (ZP^h) into the incentive compatibility conditions (IC^l) and (IC^h) yields:

$$y^l + T - kg\left(\frac{y^l}{w^l}\right) \geq y^h - \frac{\gamma^l}{\gamma^h} \cdot T - kg\left(\frac{y^h}{w^h}\right), \tag{12}$$

$$y^h - \frac{\gamma^l}{\gamma^h} \cdot T - g\left(\frac{y^h}{w^h}\right) \geq y^l + T - g\left(\frac{y^l}{\sum_i \gamma^i w^i}\right). \tag{13}$$

Under the separating equilibrium regime the government is choosing the tax parameters y^l, y^h and T so as to maximize the welfare in (11) subject to the two incentive compatibility constraints in (12) and (13).

2.5 The Pooling Equilibrium Regime

Under a pooling equilibrium regime, by choosing an appropriate tax schedule, the government can determine the common gross level of income $\bar{y}$.[11] Being unable to distinguish between the two types of workers, firms will pay all workers the same wage rate equaling the average productivity:

$$\bar{w} \equiv \sum_i \gamma^i w^i.$$

Moreover, all agents will work the same number of hours, $\bar{n}$, given, due to the zero profit condition, by:

$$\bar{n} = \bar{y}/\bar{w}. \tag{14}$$

By virtue of our assumption that the government has no exogenous revenue needs, with no loss in generality, there will be no tax levied at the income level chosen by both types of workers.

Substituting from (14) into the welfare function in (10) yields:

$$W = \beta^l \left[\bar{y} - kg\left(\bar{y}/\bar{w}\right)\right] + \beta^h \left[\bar{y} - g\left(\bar{y}/\bar{w}\right)\right]. \tag{15}$$

Under the pooling equilibrium regime the government is choosing the common gross level of income, $\bar{y}$, so as to maximize the welfare in (15).

We turn next to compare the two regimes.

[11] One simple tax schedule that implements the pooling allocation is the one in which the government levies a confiscatory 100 percent tax rate at any level of income other than $\bar{y}$. Notice that such a schedule prevents firms from engaging in "cream-skimming". The reason is that they become unable to attract only the high-skilled workers by offering them a higher compensation in exchange for longer working hours.

2.6 Comparison between the Separating and the Pooling Equilibria

In this section we present two propositions that characterize the second-best optimum associated with an egalitarian government, that assigns a relatively high welfare weight to the low-skilled workers, as a function of the differences in productivities and labor-leisure preferences between the two types of workers.

Proposition 1. *When the weight assigned to type-l workers in the social welfare function is sufficiently large and the differences in labor-leisure preferences between the two types of workers are sufficiently small there exists cutoff levels,* $0 < w^1 \leq w^2 < w^3 \leq w^4 < w^h$*, such that:*

(i) the separating allocation constitutes the second-best social optimum for $w^l \in [0, w^1] \cup [w^4, w^h]$

(ii) the pooling allocation constitutes the second-best social optimum for $w^l \in [w^2, w^3]$

(iii) $w^1 = w^2$ *and* $w^3 = w^4$*, if* $W^p(w^l, w^h, k) - W^s(w^l, w^h, k)$ *is strictly concave in* w^l*, where* W^p *and* W^s *denote, respectively, the Lagrangean expressions associated with the welfare-maximizing pooling and separating allocations.*

Proof See appendix A □

Thus, under the conditions specified in the Proposition, (i) the separating equilibrium is the second-best social optimum when the wage rates of the low and high skilled individual are either sufficiently close together or sufficiently far apart, and, (ii) there always exists an interior range where the pooling allocation constitutes the second-best social optimum. Part (iii) simply specifies a condition under which the sets in part (i) and (ii) are connected.[12]

The following proposition highlights the role of the variation in labor/leisure preferences in the determination of the optimal equilibrium configuration.

Proposition 2. *When the weight assigned to type-l workers in the social welfare function is sufficiently large and the differences in productivities between the two types of workers are sufficiently small there exists a cutoff level,* $k > 1$*, above which the separating allocation dominates the pooling allocation and vice versa. Moreover, the welfare gain associated with a switch from a separating to a pooling allocation is decreasing with respect to k.*

Proof See appendix B □

[12] The concavity property holds for a large class of the commonly used iso-elastic forms (of the disutility from work) including the specifications employed in our numerical analysis.

The two propositions demonstrate that the social desirability of a pooling regime, in which redistribution is confined to the wage channel, hinges both on the differences in productivities and the variation in labor-leisure preferences across the two types of workers.

Let us consider first the role of the heterogeneity in productivities. Its importance is due to the fact that, by implementing a pooling equilibrium and thereby forcing wage equalization, the government can eliminate all the information rent for the high-skilled type that derives from heterogeneity in productivities.[13] Thus, the larger is this information rent at the optimal separating equilibrium, the larger will be the equity gain from implementing a pooling equilibrium. Given that the aforementioned information rent is first increasing and then decreasing when we let the productivity of the low-skilled type increase from zero (as indicated by our numerical analysis below), it follows that the equity-motives for choosing a pooling over a separating allocation will tend to be larger at intermediate values of the productivity of the low-skilled type.[14]

Let us now consider the role of the heterogeneity in preferences in shaping the optimal choice of tax regime by the government. Its importance is due to the fact that it affects the cost at which the government can reap the equity gain that we have discussed above. To shed light on this issue, it is useful to consider the pattern of distortions generated by each tax regime. When the government implements a welfare maximizing separating allocation, the labor supply of the high-skilled type is either distorted upwards (when the IC^l-constraint is binding) or it is left undistorted (when the IC^l-constraint is not binding).[15] At the same time, the welfare maximizing separat-

[13] Notice that, because of the binding IC^h-constraint, the high-skilled type derives an information rent that depends both on the difference in productivities and on the difference in taste for leisure. Only the first source of information rent can be neutralized by the government through the implementation of a pooling equilibrium.

[14] To clarify this point, consider a standard two-type optimal taxation model with no asymmetric information between firms and workers and where workers can only differ in market ability (productivity). A government maximizing a *max-min* social welfare function would always equalize utilities across agents in a first-best setting. In a second-best setting with asymmetric information between the government and the private agents, the government would be thwarted by a downward binding incentive-compatibility constraint in its attempt to equalize utilities. This constraint generates an information rent for the high-skilled agents that manifests itself in a utility gap between the high- and the low-skilled agents. However, for a given productivity of the high-skilled agents, this utility gap is at its minimum when the productivity of the low-skilled type either approaches zero or it approaches the productivity of the high-skilled type. The reason is that the utility gap arises from the difference between the labor supply of a low-skilled agent and that of a high-skilled agent, both evaluated at the income point intended by the government for the low-skilled agents only. Denoting by w^1 and w^2 the wage rate of, respectively, low- and high-skilled agents, the difference between the labor supply of a low-skilled agent and that of a high-skilled mimicker is equal to y^1/w^1-y^1/w^2, which tends to zero when either y^1 approaches zero (which is the case when w^1 approaches zero) or when w^1 approaches w^2.

[15] The upward distortion on the labor supply of the high-skilled type is also a feature of the laissez-faire (separating) equilibrium. Because of the asymmetric information between firms and workers, this inefficiency is needed for firms to achieve separation and prevent mimicking by the low-skilled type. When the government redistributes in favor of the low-skilled type, it alleviates the incentive for low-skilled workers to disguise themselves as high-skilled ones. Therefore, when the welfare maximizing

ing allocation introduces a downward distortion on the labor supply of the low-skilled type (due to the binding IC^h-constraint). In comparison, when the government implements a welfare maximizing pooling allocation, the equalization of the labor supplies implies that the labor supply of the high-skilled type is downward distorted, whereas the labor supply of the low-skilled type is distorted upwards.

When low-skilled workers incur a significantly higher disutility from work than their high-skilled counterparts, the IC^l-constraint will not be binding at the welfare maximizing separating equilibrium, and therefore the labor supply of the high-skilled agents will be left undistorted.[16] At the same time, with k large and when the weight assigned to type-l workers in the social welfare function is large, the pooling regime will feature a low $\bar{y}$ (and a low $\bar{n}$) and therefore produce a large downward distortion on the labor supply of type-h workers. On the other hand, when there is not much heterogeneity in labor-leisure preferences, the pooling regime will produce a smaller downward distortion on the labor supply of type-h workers. Moreover, given the fact that both regimes feature a double distortion, and in particular the fact that the labor supply of the high-skilled type is distorted also in the separating regime, implies that the efficiency costs of switching from a separating- to a pooling tax regime tend to be lower.

It is important to notice the differences between our framework and the standard Mirrleesian setting. The possibility to pool wages (rather than merely incomes) due to the asymmetry in information between firms and workers enables the government, when differences in labor-leisure preferences are small, to nearly equalize the utility levels derived in equilibrium by the low- and high-skilled workers. This can never occur under a separating regime in the standard Mirrleesian setting, as it would yield a violation of the incentive compatibility constraint associated with the high-skilled workers. Under a separating regime, in light of our assumption that the welfare weight assigned to the low-skilled workers is sufficiently large, the incentive compatibility constraint associated with the high-skilled workers remains binding as in the standard Mirrleesian setting. Thus, the optimality of levying a positive marginal tax rate on the low-skilled workers carries over to our setting. The difference from the standard setting is reflected in the fact that the incentive constraint associated with the high-skilled worker is less binding, as the high-skilled mimicker is only remunerated according to the average productivity in the labor market rather than based on his/her high productivity level, which is unobserved by the firm. This clearly represents a mechanism that facilitates the pursuit of the government's redistributive objectives.[17] On the other

separating equilibrium maintains an upward distortion on the labor supply of the high-skilled workers, the magnitude of the distortion is smaller than that associated with the laissez-faire equilibrium.

[16]The variation in the taste for leisure will suffice to render mimicking by the low-skilled workers unattractive without the need to distort (upwards) the labor supply of their high-skilled counterparts.

[17]This result is not novel to our analysis. Instead, it is one of the main results of Stantcheva's (2014)

hand, the possibility that in our model the incentive constraint associated with the low-skilled agents is also binding implies another difference from the standard setting. In particular, rather than always obtaining a no-distortion at the top result, there are cases where the separating equilibrium implemented by a benevolent government entails an upward distortion on the labor supply of high-skilled agents.[18] A relevant, and previously unnoticed, implication of this second difference from the standard setting is that the asymmetric information between firms and workers does not necessarily make it easier to redistribute. This would clearly be the case when the IC^l-constraint does not bind at the separating tax regime. However, if this constraint happens to be binding (a possibility that is confirmed from the numerical simulations that we present below), the fact that the labor supply of the high-skilled workers is distorted in the separating tax regime with adverse selection in place, whereas it is left undistorted in a standard Mirrleesian setting, implies that social welfare may be lower in the presence of adverse selection in the labor market.

2.6.1 Numerical example

For illustrative purposes, we consider here a simple numerical example. We make the following parametric assumptions: $g(n) = \frac{n^4}{4}$, $k = 1.005$, $w^h = 100$, $\gamma^h = 0.6$. The figure below compares the max-min (Rawlsian) welfare levels associated with the pooling and the separating allocations for different values of w^l, the productivity of the low-skilled workers.

Two insights emerge from the figure. First, consistent with Proposition 1, there exists an interior range $w^2 \leq w^l \leq w^3$ in which the pooling equilibrium dominates, and outside this range the separating allocation prevails. In this example, w^2 is equal to 29.132 and w^3 is very close to but strictly less than 100 (the exact value is 99.998). Second, and perhaps most importantly, the superiority of the pooling allocation is not confined to knife-edge cases. Pooling turns out to be socially desirable over a large range of parameters and a shift from a separating to a pooling allocation may yield a substantial welfare gain. The maximal welfare gain from pooling is obtained when $w^l = 87.8$ and amounts (in equivalent-variation terms) to 15.6% of the total output produced in the separating equilibrium.[19] For the numerical example that we have considered, under the separating tax regime the IC^l-constraint turns out being binding for almost the entire range of possible values for the productivity of the low-skilled

paper.

[18]It is worth emphasizing that this property derives from an attempt to mitigate an upward binding incentive-compatibility constraint associated with low-skilled mimickers and not, as in the endogenous wages literature that relies on complementarities across different skill types in the production technology [see Stiglitz (1982)], deriving from an attempt to mitigate a downward binding incentive-compatibility constraint associated with high-skilled mimickers.

[19]We have also performed simulations considering less egalitarian preferences for redistribution and obtained qualitatively very similar results. The results are available upon request.

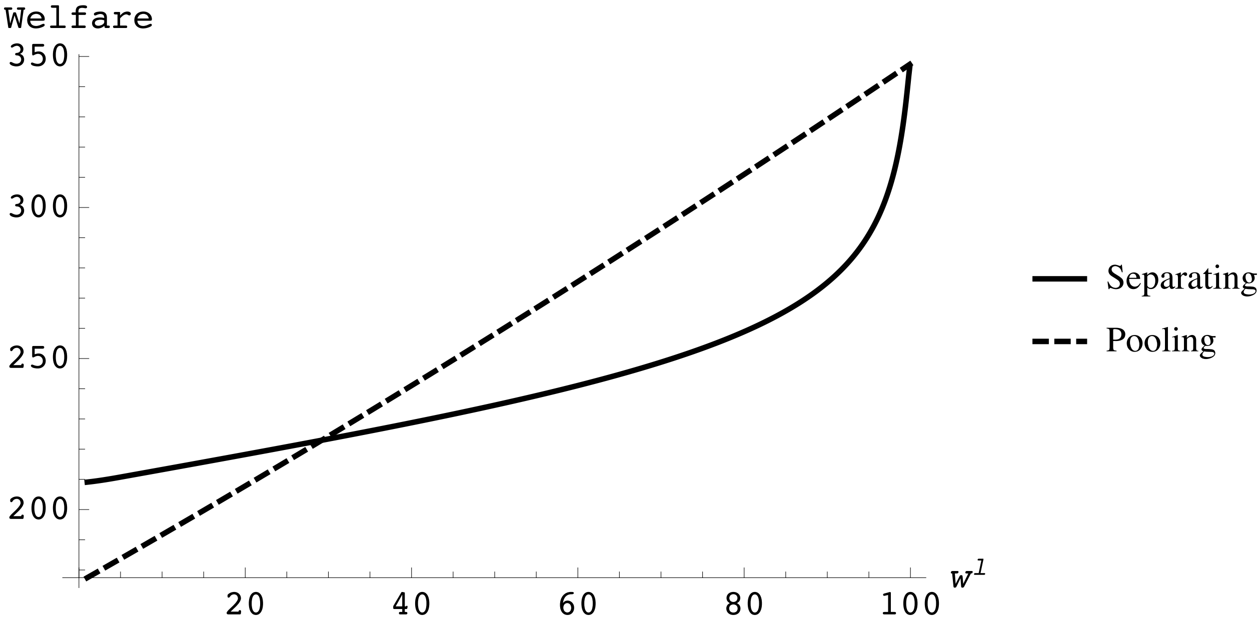

Figure 1: Welfare comparison between the Separating and Pooling Allocation.

workers. In particular, the IC^l-constraint is binding for all values of w^l between 4.705 and 99.996. This implies that, over this range of values for w^l, the separating tax regime entails an upward distortion on the labor supply of the high-skilled workers. Moreover, within this interval, the size of the upward distortion is first increasing, reaching a maximum when w^l is equal to 98.48 and then starts rapidly decreasing.[20] As we have already noticed in our discussion of Propositions 1 and 2, the equity-motives for choosing a pooling over a separating allocation will tend to be larger at intermediate values for the productivity of the low-type. We also pointed out how, when the separating tax regime entails a distortion on the labor supply of high-skilled workers, a shift from a separating- to a pooling tax regime entails a relatively small efficiency cost. Putting these two observations together, and given that in our numerical example the upward distortion induced by the separating tax regime is maximized at quite high values for w^l, one can explain why the welfare gains of switching from the separating- to the pooling tax regime are substantial at moderately large values for the productivity of the low-skilled type.[21]

[20] The value of the undistorted labor supply of high-skilled agents is 4.64159. At w^l equal to 98.48 the labor supply of the high-skilled workers is 5.03609 under the separating tax regime.

[21] In the discussion following Proposition 1 and 2 we have hinted at the fact that the asymmetric information between firms and workers does not necessarily make it easier for the government to redistribute, implying that social welfare might be lower in the presence of adverse selection in the labor market. This is confirmed by our numerical example; comparing the social welfare that we obtain in our setting under the optimal tax regime (either separating or pooling, depending on the value of w^l) with the social welfare that would have been obtained in a standard Mirrleesian setting, we have found that for $4.705 \leq w^l < 37.293$ social welfare is actually higher in a standard Mirrleesian setting.

2.7 Relation to the literature

A closely related study to our analysis is Stantcheva (2014). Whereas we focus on the *RS* equilibrium, the model invoked by Stantcheva (2014) assumes a Miyazaki-Wilson-Spence (*MWS*) solution concept which assumes that firms break even on their entire portfolio of labor contracts. This gives rise, by construction, to cross-subsidization across workers when the number of types is greater than two, but also occurs with two types when the fraction of low-skilled workers is relatively small. Clearly, in such a setting part of the redistribution is done via the wage channel; hence, when the government sets its optimal tax policy, it takes into account the effects of its policy on the degree of cross subsidization done by firms. This serves as a complementary channel for redistribution.

Stantcheva indeed demonstrates the interaction between cross-subsidization and optimal taxation in the case of a linear tax system by showing that the government should increase the linear tax rate above the level associated with the standard Mirrleesian setting where workers' types are observable by firms. The reason is that this induces firms to cross-subsidize low-skilled workers as a means to induce them to reveal their true type, which indirectly enhances redistribution.

In the non-linear case, however, Stantcheva does not refer to this issue and focuses on the role of asymmetric information between firms and workers in promoting the government's redistributive goals. We touch upon this issue too (highlighting, however, that social welfare might actually be lower in the presence of adverse selection in the labor market than under the standard Mirrleesian case) but set focus on the appropriate balance between the wage and income channels in promoting re-distributive goals. Invoking the *RS* solution concept in which cross-subsidization does not arise in equilibrium, and confining attention to the case of non-linear taxation, we examine the optimal mixture of the wage- and income channels as complementary tools to promote redistributive goals. Our key point is that the government can control this balance by choosing an appropriate tax schedule and that the wage channel may play a substantial role in enhancing redistribution.[22] [23]

3 The Three-Type Case

In this section we consider a three-type extension of our baseline model. We will then in the next section turn to discuss how the model generalizes to the case with a continuum of types.

[22]Stantcheva does provide an analysis of the *RS* equilibrium in her online appendix but does not consider the possibility of the government to affect the wage distribution.

[23]Note that Stantcheva disregards the possibility of pooling. Even though pooling cannot exist in the *laissez faire* regime, it may well exist when nonlinear taxes are available.

We assume that there are three types of workers, indexed by $i = 1, 2, 3$, who differ in their productivity denoted by w^i, where $w^3 > w^2 > w^1 \geq 0$. As before, productivities are assumed to be private information unobserved by either the firms or the government. The total workers' population is normalized to a unit measure and the fraction of type i workers is denoted by $0 < \gamma^i < 1$, where $\sum_{i=1}^{3} \gamma^i = 1$.

As compared to the two-type case, we introduce a slightly more general notation and let the preferences of a worker with ability w be represented by the utility function:

$$u\left(w, n, c\right) = c - g(w, n),$$

where c denotes consumption, n denotes working hours, $\frac{\partial g}{\partial n} > 0$, $\frac{\partial^2 g}{\partial n^2} > 0$ and $\frac{\partial^2 g}{\partial n \partial w} < 0$.

Several remarks are in order. First notice that g, which measures the disutility from work, is strictly increasing and strictly convex. Further, notice that the cross-derivative condition implies that the single-crossing property holds. In graphical terms, the property implies that the indifference curves (in the n-c space) become flatter as w increases. More formally, for any two bundles (n, c) and (n', c') where $n > n'$ and $c > c'$, the single crossing property implies that $u\left(w, n, c\right) \geq u\left(w, n', c'\right) \Longrightarrow u\left(\tilde{w}, n, c\right) > u\left(\tilde{w}, n', c'\right)$ for all $\tilde{w} > w$, and $u\left(w, n', c'\right) \geq u\left(w, n, c\right) \Longrightarrow u\left(\tilde{w}, n', c'\right) > u\left(\tilde{w}, n, c\right)$ for all $\tilde{w} < w$.

Three possible equilibrium configurations need to be considered: a fully separating equilibrium where each type is associated with a distinct bundle, a hybrid equilibrium featuring partial pooling, and a pooling equilibrium where all types are bunched together at the same bundle. We will formulate the government problem for each possible equilibrium configuration and then conduct a welfare comparison by resorting to numerical simulations. In all cases, to stay in line with the previous section, we assume a *max-min* government; namely, the government seeks to maximize the utility of type 1 workers, the least well-off individuals, subject to a balanced budget. We start with the simplest pooling configuration.

3.1 Pooling Equilibrium

In a pooling equilibrium all workers are offered the same labor contract $(\hat{n}, \hat{y})$ and taxes and transfers are set to zero (by virtue of our assumption that the government has no revenue needs). The government is maximizing the utility of type-1 workers, $u(w^1, \hat{n}, \hat{c})$, subject to the following revenue constraint:

$$\hat{c} = \hat{y}, \tag{16}$$

where $\hat{y} = \hat{n} \cdot \sum_{i=1}^{3} \gamma^i \cdot w^i$.

Condition (16) states that, indeed, there are no taxes or transfers set by the government at the income level associated with the pooling allocation, $\hat{y}$, to ensure that the budget is balanced. All workers are associated with the same labor contract, where the wage rate is equal to the average productivity. The stability of the pooling allocation is guaranteed by the fact that firms are unable to engage in 'cream-skimming' as all income levels other than that associated with the pooling allocation, $\hat{y}$, can be ruled out by an appropriate choice of the tax function.[24]

3.2 Hybrid Equilibria

With three types of agents, there are potentially three possible hybrid equilibria where two of the three workers' types are bunched together.[25] However, as we show in the Lemma below, a simple argument can be provided to rule out the possibility of a hybrid equilibrium where types 1 and 3 are pooled.

Lemma 1. *There is no hybrid equilibrium in which types 1 and 3 are bunched together.*

Proof See appendix E □

By virtue of the above Lemma we are left with two possible hybrid equilibria. Pooling of type 1 and 2, and, pooling of type 2 and 3. We consider these possibilities next.

3.2.1 Types 1 and 2 bunched together.

A first possibility is that types 1 and 2 are bunched together. In this case, let the net income (consumption) levels in the bundles associated with types 1 and 2 (who are bunched together), and type 3, be respectively denoted by c^{12} and c^3. Moreover, let the labor contracts associated, in equilibrium, with types 1 and 2, and type 3, be respectively denoted by (n^{12}, y^{12}) and (n^3, y^3).

The government is seeking to maximize the utility of type-1 workers, $u(w^1, n^{12}, c^{12})$, subject to the revenue constraint:

$$\sum_{i=1}^{2} \gamma^i \cdot (y^{12} - c^{12}) + \gamma^3 \cdot (y^3 - c^3) \geq 0,$$

[24]The argument is similar to the one described in footnote 11 for the two-type case.

[25]It is worth noticing that partial pooling equilibria (but not fully pooling equilibria) may also be optimal in the standard three-type version of the Mirrlees (1971) model, i.e. in a setting where workers only differ in skill and where there is no asymmetric information between firms and workers. This is for instance the case when there are relatively few workers of the intermediate type; as discussed by Stiglitz (1987, pp. 1010-1011), Pareto efficiency might then require a partial pooling equilibrium in which the lowest skilled individuals are offered the same bundle as the intermediate type. However, this sort of pooling is different from the one that characterizes our analysis as it does not entail redistribution through the wage channel. More precisely, whereas partial pooling in the standard version of the Mirrlees model (1971) implies that the hours of work differ for the different types that are bunched together (and only consumption is equalized), in our setting all pooled agents work the same amount of hours (and, of course, consumption is equalized too).

and the following set of incentive compatibility constraints:

$$u(w^1, c^{12}, n^{12}) \geq u(w^1, c^3, n^3), \tag{17}$$

$$u(w^2, c^{12}, n^{12}) \geq u(w^2, c^3, n^3), \tag{18}$$

$$u(w^3, c^3, n^3) \geq u(w^3, c^{12}, n^{12}), \tag{19}$$

$$u(w^3, c^3, n^3) \geq u(w^3, c^{12}, y^{12} / \sum_{i=1}^{3} \gamma^i \cdot w^i), \tag{20}$$

where $y^3 = n^3 w^3$ and $y^{12} = n^{12} \cdot \frac{\sum_{i=1}^{2} \gamma^i \cdot w^i}{\sum_{i=1}^{2} \gamma^i}$.

Several remarks are in order. Conditions (17)-(19) are the standard incentive compatibility constraints with one exception: the mimickers work the same number of hours as their mimicked types due to the asymmetric information between firms and workers. Condition (20) states an additional incentive compatibility constraint that derives from the fact that, by assumption, the government cannot directly control the working hours specified by each labor contract. Firms have some margin of maneuver to offer new contracts (in addition to those specified by the equilibrium allocation) by reducing the working hours (n) associated with each level of income (y). Condition (20) ensures that any such contract that yields non-negative profits will not be incentive compatible. To see this, suppose that the inequality sign in (20) is reversed and consider the contract $(\frac{y^{12}}{\sum_{i=1}^{3} \gamma^i \cdot w^i}, y^{12})$. The offered contract would be strictly preferred to any other contract by type 3 [by virtue of the presumed reversed inequality sign in (20)] as well as by types 1 and 2, recalling that $\frac{y^{12}}{\sum_{i=1}^{3} \gamma^i \cdot w^i} < n^{12} = y^{12} / \left(\frac{\sum_{i=1}^{2} \gamma^i \cdot w^i}{\sum_{i=1}^{2} \gamma^i} \right)$. The offered contract would therefore induce a pooling allocation that yields zero profits.[26] Incentive compatibility thus requires, as stated by condition (20), that type-3 would weakly prefer his/her equilibrium contract (n^3, y^3) to the contract $(\frac{y^{12}}{\sum_{i=1}^{3} \gamma^i \cdot w^i}, y^{12})$.[27]

Notice that the set of incentive compatibility constraints can be reduced, as some of the constraints will not be binding. First, condition (20) implies condition (19), as by construction, $\frac{y^{12}}{\sum_{i=1}^{3} \gamma^i \cdot w^i} < n^{12}$. Second, condition (18) implies condition (17) by virtue of the single-crossing property.[28]

---

[26]Offering contracts that induce pooling allocations associated with the higher income level, y^3, will never form a profitable deviation for the firm. Any such pooling contract $[(n, y^3)$with $n < n^3]$ would entail a mechanical decrease in profits, augmented by an additional behavioral decrease in profits due to the attraction of workers with lower ability levels (types 1 and 2).

[27]Notice that condition (20) further implies that type-3 would strictly prefer his/her equilibrium bundle to any pooling contract that yields *positive* profits $\left[\text{as the latter requires setting longer working hours, } n > \frac{y^{12}}{\sum_{i=1}^{3} \gamma^i \cdot w^i}\right]$.

[28]To see this, notice that in the hybrid equilibrium, by virtue of the single-crossing property, it necessarily follows that $c^3 > c^{12}$ and $n^3 > n^{12}$. Thus, if type-2 (weakly) prefers a bundle with a lower level of working hours and, correspondingly, a lower level of consumption, to a bundle with higher levels of working hours and consumption, then type-1 will (strictly) prefer this bundle as well.

3.2.2 Types 2 and 3 bunched together.

We turn next to the other possible hybrid equilibrium, the one in which types 2 and 3 are bunched together. For this purpose, let the net income (consumption) levels in the bundles associated with types 2 and 3 (who are bunched together), and type 1, be respectively denoted by c^{23} and c^1. Moreover, let the labor contracts associated, in equilibrium, with types 2 and 3, and type 1, be respectively denoted by (n^{23}, y^{23}) and (n^1, y^1).

The government aims at maximizing the utility of type-1 workers, $u(w^1, n^1, c^1)$, subject to the revenue constraint:

$$\gamma^1 \cdot \left(y^1 - c^1\right) + \sum_{i=2}^{3} \gamma^i \cdot (y^{23} - c^{23}) \geq 0,$$

and the following set of incentive compatibility constraints:

$$u(w^1, c^1, n^1) \geq u(w^1, c^{23}, n^{23}), \tag{21}$$

$$u(w^2, c^{23}, n^{23}) \geq u(w^2, c^1, n^1), \tag{22}$$

$$u(w^3, c^{23}, n^{23}) \geq u(w^3, c^1, n^1), \tag{23}$$

$$u(w^2, c^{23}, n^{23}) \geq u[w^2, c^1, y^1 / (\sum_{i=1}^{2} \gamma^i w^i / \sum_{i=1}^{2} \gamma^i)], \tag{24}$$

$$u(w^3, c^{23}, n^{23}) \geq u(w^3, c^1, y^1 / \sum_{i=1}^{3} \gamma^i w^i), \tag{25}$$

where $y^1 = n^1 w^1$ and $y^{23} = n^{23} \cdot \frac{\sum_{i=2}^{3} \gamma^i \cdot w^i}{\sum_{i=2}^{3} \gamma^i}$. Several remarks are in order. As in the previous hybrid configuration, conditions (21)-(23) are the standard incentive compatibility constraints with the exception that mimickers work the same number of hours as their mimicked types due to the asymmetric information between firms and workers. Conditions (24)-(25) state two additional incentive compatibility constraints that derive from the fact that, by assumption, the government cannot directly control the working hours specified by each labor contract.

Similar to condition (20) in the formulation of the previous hybrid configuration, conditions (24)-(25) ensure that firms cannot profitably deviate by reducing the working hours associated with the low level of income, y^1.[29] There are two such potentially profitable deviations to consider. The firm can either reduce the working hours moderately below n^1, so as to attract type-2 workers (but not their type-3 counterparts), or, alternatively, reduce the working hours sufficiently so as to attract *both* type-2 and type-3 workers. Conditions (24) and (25) ensure, respectively, that it is impossible for

[29] As in the previous hybrid configuration, decreasing the working hours associated with the high level of income, y^{23}, will unambiguously yield negative profits (see footnote 26 above).

such deviations to yield non-negative profits.

To see this, consider first condition (24). Suppose that the inequality sign in (24) is reversed and consider the contract $\left(y^1 / \frac{\sum_{i=1}^2 \gamma^i \cdot w^i}{\sum_{i=1}^2 \gamma^i},\, y^1\right)$. The offered contract would be strictly preferred to any of the two other contracts by both type-2 [by virtue of the presumed reversed inequality sign] and type-1 (as $y^1 / \frac{\sum_{i=1}^2 \gamma^i \cdot w^i}{\sum_{i=1}^2 \gamma^i} < n^1 = y^1 / w^1$) and would, by construction, yield zero profits (and strictly positive profits, if the new contract would be strictly preferred by type 3 as well). Incentive compatibility thus requires, as stated by condition (24), that type-2 would weakly prefer his/her equilibrium contract (n^{23}, y^{23}) to the contract $\left(y^1 / \frac{\sum_{i=1}^2 \gamma^i \cdot w^i}{\sum_{i=1}^2 \gamma^i},\, y^1\right)$.[30]

Consider next condition (25). Suppose that the inequality sign in (25) is reversed and consider the contract $\left(y^1 / \sum_{i=1}^3 \gamma^i \cdot w^i,\, y^1\right)$. The offered contract would be strictly preferred to any of the two other contracts by type-3 (by virtue of the presumed reversed inequality sign), type-2 (by virtue of the single-crossing property)[31] and type-1 (as $y^1 / \sum_{i=1}^3 \gamma^i \cdot w^i < n^1 = y^1 / w^1$) and would, by construction, yield zero profits. Incentive compatibility thus requires, as stated by condition (25), that type-3 would weakly prefer his/her equilibrium contract (n^{23}, y^{23}) to the contract $\left(y^1 / \sum_{i=1}^3 \gamma^i \cdot w^i,\, y^1\right)$.[32] As a final observation, notice that the set of incentive compatibility constraints can be reduced, as some of the constraints will not be binding. First, notice that condition (24) implies condition (22) as $y^1 / (\sum_{i=1}^2 \gamma^i \cdot w^i / \sum_{i=1}^2 \gamma^i) < n^1 = \frac{y^1}{w^1}$. Further, notice that condition (25) implies condition (23) as $\frac{y^1}{\sum_{i=1}^3 \gamma^i \cdot w^i} < n^1$.

3.3 Separating Equilibrium

We finally turn to characterize the fully separating equilibrium. In such equilibrium each type is offered a distinct bundle. Let c^i denote the net income (consumption) level associated with type-i workers and let (n^i, y^i) denote the labor contract associated with type-i workers ($i = 1, 2, 3$)

The government aims at maximizing the utility of type-1 workers, $u(w^1, n^1, c^1)$,

[30] Notice that condition (24) further implies that type-2 would strictly prefer his/her equilibrium bundle to any contract that induces bunching of types 1 and 2 and yields strictly positive profits $\Big[$as the latter requires setting longer working hours, $n > y^1 / \left(\frac{\sum_{i=1}^2 \gamma^i \cdot w^i}{\sum_{i=1}^2 \gamma^i}\right)\Big]$.

[31] To see this notice that in the hybrid equilibrium, by virtue of the single-crossing property, it necessarily follows that $c^{23} > c^1$ and $n^{23} > n^1$. Thus, if type-3 prefers the offered contract (which specifies a lower level of working hours and, correspondingly, a lower level of consumption) to his/her equilibrium contract (which specifies higher levels of working hours and consumption) then type-2 will prefer the offered contract to his/her equilibrium bundle as well.

[32] Condition (25) further implies that type-3 would strictly prefer his/her equilibrium bundle to any pooling contract that yields strictly positive profits [as the latter requires setting longer working hours, $n > y^1 / \sum_{i=1}^3 \gamma^i \cdot w^i$].

subject to the revenue constraint:

$$\sum_{i=1}^{3} \gamma^i \cdot (y^i - c^i) \geq 0,$$

and the following set of incentive compatibility constraints:

$$u(w^i, c^i, n^i) \geq u(w^i, c^j, n^j);\ i = 1, 2, 3; j \neq i, \tag{26}$$

$$u(w^2, c^2, n^2) \geq u[w^2, c^1, y^1 / (\sum_{i=1}^{2} \gamma^i \cdot w^i / \sum_{i=1}^{2} \gamma^i)], \tag{27}$$

$$u(w^3, c^3, n^3) \geq u(w^3, c^1, y^1 / \sum_{i=1}^{3} \gamma^i \cdot w^i), \tag{28}$$

$$u(w^3, c^3, n^3) \geq u[w^3, c^2, y^2 / (\sum_{i=2}^{3} \gamma^i \cdot w^i / \sum_{i=2}^{3} \gamma^i)] \tag{29}$$

$$\text{if} \quad u(w^1, c^1, n^1) \geq u[w^1, c^2, y^2 / (\sum_{i=2}^{3} \gamma^i \cdot w^i / \sum_{i=2}^{3} \gamma^i)];$$

$$u(w^3, c^3, n^3) \geq u(w^3, c^2, y^2 / \sum_{i=1}^{3} \gamma^i \cdot w^i), \tag{30}$$

where $y^i = n^i w^i; i = 1, 2, 3$.

Several remarks are in order. As in all the previous equilibrium configurations, condition (26) states the standard incentive compatibility constraints with the exception that mimickers work the same number of hours as their mimicked types due to the asymmetric information between firms and workers. Conditions (27)-(30) derive from the fact that, by assumption, the government cannot directly control the working hours specified by each labor contract. Conditions (27)-(28) ensure that firms cannot profitably deviate by reducing the working hours associated with the low level of income, y^1; whereas, conditions (29)-(30) ensure that firms cannot profitably deviate by reducing the working hours associated with the intermediate level of income, y^2.[33] For each income level (y^1 and y^2) there are several potentially profitable deviations to consider. Consider first the low level of income, y^1. The firm can either reduce the working hours moderately below n^1, so as to attract type-2 workers (but not their type-3 counterparts), or, alternatively, reduce the working hours sufficiently so as to attract both type-2 and type-3 workers. Conditions (27) and (28) ensure, respectively, that it is impossible for such deviations to yield non-negative profits [the arguments are similar to those concerning conditions (24) and (25) in the hybrid configuration and are hence omitted].

[33] As in the previous equilibrium configurations, decreasing the working hours associated with the high level of income, y^3, will unambiguously yield negative profits (see footnote 26 above).

Consider next the intermediate level of income y^2 and suppose that the firm offers to reduce the working hours to some $n < n^2 = y^2/w^2$. By construction, the offered contract would be strictly preferred by type-2 to any of the other three equilibrium bundles. In order to form a profitable deviation the offered contract has to be strictly preferred by type-3 (the high-productivity type) to any of the other three equilibrium bundles, so that by attracting type-3 workers the firm can offset the mechanical loss associated with the reduction of working hours (while maintaining the same compensation level).[34] There are two such potential profitable deviations to consider. Either the offered contract attracts type-3 only (and not their type-1 counterparts) or, alternatively, the offered contract attracts both types 1 and 3. Conditions (29)-(30) ensure, respectively, that such deviations would not yield non-negative profits.

As a final observation, notice that the set of incentive compatibility constraints can be reduced, as some of the constraint will not be binding. First, notice that the single-crossing property implies that it suffices to focus on the adjacent incentive compatibility constraints given in condition (26); that is, type j weakly prefers his/her bundle to those associate with types $j-1$ and $j+1$ (whenever applicable).

Further, notice that the set of constraints given in (26)-(30) may be further reduced as condition (27) implies that $u(w^2, \overset{2}{c}, n^2) \geq u(w^2, \overset{1}{c}, n^1)$.

3.4 Numerical Examples

We turn next to compare the different equilibrium configurations of the three-type model numerically. In this section we employ the utility function

$$u(w, n, c) = c - k\frac{n^\theta}{\theta},$$

and consider two sets of values for k for the different agents. In the first scenario we consider a "low" variance in k corresponding to the values $\{k_1, k_2, k_3\} = \{1.05, 1.025, 1\}$ and in the second scenario we consider a "high" variance corresponding to the values $\{k_1, k_2, k_3\} = \{1.25, 1.1, 1\}$. In the simulations we investigate how the welfare ranking of the different equilibria changes in response to changes in the degree of productivity dispersion. We fix the productivity of type 2 to $w^2 = 100$ and consider a mean-preserving spread where the productivity of type 1 varies from $w^1 = 50$ to $w^1 = 100$ and where we let the productivity of type 3 vary from $w^3 = 150$ to $w^3 = 100$. In all cases we employ $\gamma_1 = \gamma_2 = \gamma_3 = 1/3$. The results are shown in figure 2 and figure 3. Notice that a movement to the right along the horizontal axis is in both figures associated with a higher degree of productivity dispersion. In particular, the variable Δ on

[34]Notice that the offered contract may also attract type-1 (low-productivity) workers (whose productivity is lower than their type-2 counterparts). This behavioral effect will induce a further decrease in profits in addition to that associated with the mechanical reduction in working hours.

the horizontal axis represents the difference $100 - w^1$ (or, equivalently, $w^3 - 100$).

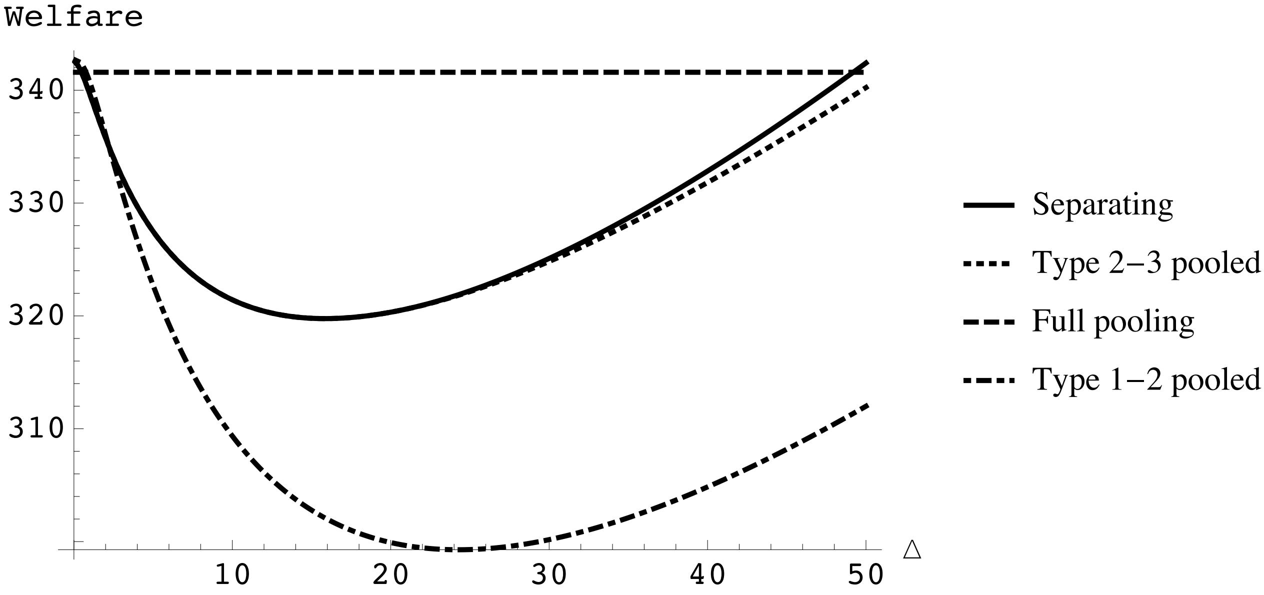

Figure 2: Three-type case with low degree of preference dispersion.

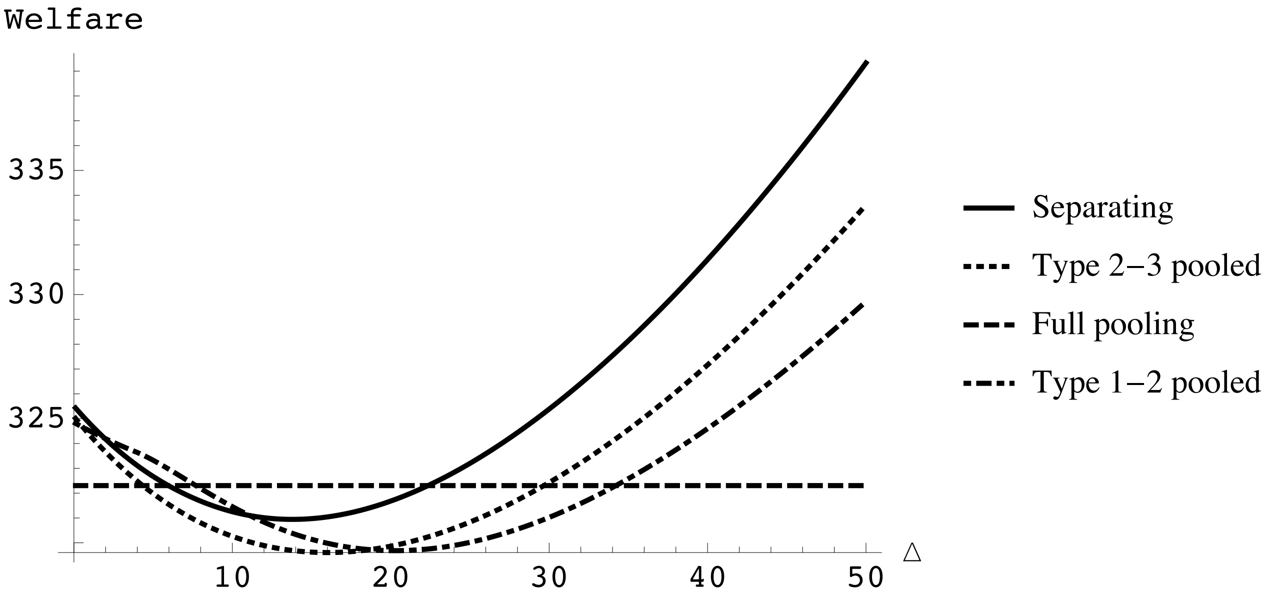

Figure 3: Three-type case with high degree of preference dispersion.

The first thing to notice about figures 2 and 3 is that in both cases we get the same qualitative results for the optimal tax regime that prevails as we change the variance in productivities as captured by Δ. In particular, we get that a separating tax regime dominates both when Δ is extremely small and when it is sufficiently large. For intermediate values of Δ, we get that the optimal tax regime is either a hybrid regime where workers of type 1 and 2 are pooled (when the variance is sufficiently low) or a full pooling regime.[35]

[35]The reason why in Figure 2 and 3 the pooling line is flat, meaning that under a full-pooling tax

Let us turn to analyze more closely the bundle offered to type 1 workers by the different tax regimes. When the hybrid regime which pools workers of type 1 and 2 dominates, both the consumption and labor supply of the low-skilled workers are higher than under a fully separating tax regime but lower than under a full pooling tax regime. Thus, such an intermediate tax regime allows combining redistribution via the tax-transfer channel (collecting taxes from workers of type 3 and transferring the revenue to workers of type 1 and 2) and redistribution via the wage channel (since pooling workers of type 1 and 2 implies wage redistribution from the latter to the former). This might prove to be socially desirable when a full pooling tax regime would imply a too large efficiency cost arising from the downward distortion imposed on the labor supply of the top-skilled workers. Given that this downward distortion is larger in the high-variance-in-*k*-scenario than in the low-variance-in-*k*-scenario, this also helps explaining why the interval where the hybrid regime dominates is larger in Figure 3 than in Figure 2.

Notice also that in our examples it is never the case that the optimal tax regime is a hybrid regime entailing pooling of workers of type 2 and 3. An intuition for this outcome rests on two observations. First of all, in such a hybrid tax regime wage redistribution does not directly benefit the lowest-skilled workers (since wage redistribution occurs between workers of type 2 and 3), whose utility the government tries to maximize.[36] Second, in our examples the hybrid regime pooling workers of type 2 and 3 is always (weakly) dominated by the fully separating regime. This appears to be due to the fact that, although both regimes separate the lowest skilled workers from the other types, the tax revenue transferred to type 1 workers is raised in a more efficient way under the fully separating tax regime. In particular, as compared with the fully separating tax regime, the hybrid regime pooling workers of type 2 and 3 never produces a smaller distortion on the labor supply of the top-skilled workers (and almost never implies a smaller distortion on the labor supply of type 2 workers).[37] Finally, as one can see comparing the two figures, increasing the dispersion in the distribu-

regime the social welfare stays constant as Δ varies, is that we consider a mean-preserving change in the distribution of productivities. Notice that this was not the case in Figure 1 where the average productivity was allowed to change.

[36] Instead, in the hybrid regime where workers of type 1 and 2 are pooled, and of course also in a full pooling regime, wage redistribution directly benefits the lowest-skilled workers.

[37] In the simulations used to build Figure 2, the fully-separating tax regime generates a (modest) upward distortion on the labor supply of the top-skilled workers only when $0.4 \leq \Delta \leq 20$ (i.e. when w^1 takes values between 80 and 99.6, with w^3 taking, therefore, values between 120 and 100.4). For this range of values the social welfare under a separating tax regime is the same as under the hybrid tax regime pooling workers of type 2 and 3; however, the full pooling tax regime constitutes in this case the socially optimal outcome. For values of Δ where the fully separating tax regime leaves undistorted the labor supply of the top-skilled workers, the hybrid regime pooling workers of type 2 and 3 generates a downward distortion on the labor supply of type 3 workers. In the simulations used to build Figure 3, the labor supply of the top-skilled workers is never distorted at a fully-separating tax regime, whereas it is always downward distorted under the hybrid regime pooling workers of type 2 and 3.

tion of labor-leisure preferences extends the range where a fully separating tax regime dominates.[38]

4 Model with a continuum of types

For tractability reasons, we have so far relied on a discrete-type model to highlight the result that, in labor markets characterized by asymmetric information between firms and workers, it may be optimal for the government to carry out redistribution also through the wage channel. However, one can show that the applicability of this qualitative result is more general. For this purpose, in this section we will provide a brief discussion of how the result generalizes to a continuous-type model.

Assume a population of workers of mass one and assume that skill levels (measured by the hourly productivity), denoted by w, are distributed according to some continuous CDF, denoted by F, with strictly positive densities, $f \equiv dF/dw$, over some bounded support, $[\underline{w}, \overline{w}]$ where $\underline{w} \geq 0$. The utility function of a type-w worker is once again given by $u(w, n, c) = c - g(w, n)$, where c denotes consumption, n denotes working hours, $\frac{\partial g}{\partial n} > 0$, $\frac{\partial^2 g}{\partial n^2} > 0$ and $\frac{\partial^2 g}{\partial n \partial w} < 0$.

Suppose that the tax function is given by $t(y)$. The consumption (net income) is then given by: $c(y) \equiv y - t(y)$. To stay in line with the previous analysis of the discrete-type case, we focus on the case of a *max-min* government (the qualitative results extend, by continuity considerations, to the case where the redistributive tastes of the government are sufficiently strong). Given the tax schedule in place, firms offer labor contracts summarized by the schedule $y(n)$, which specifies the compensation (gross income), y, for any amount of time worked, n. Notice that the compensation schedule is independent of the skill level w (that is, all types can choose any contract along the schedule). This reflects the fact that skill levels are private information, unobserved by either the firms or the government.

A type-w worker chooses labor supply by solving the following maximization program:

$$\max_n c[y(n)] - g(w, n). \tag{31}$$

Let $n^*(w)$ denote the solution to the program given in (31). Using primes to denote derivatives, and assuming that both $y(n)$ and $t(y)$ are continuously differentiable, the

[38] For the two scenarios with high and low variance in the labor-leisure preferences, we have also performed simulations keeping fixed the productivities of type 2 and 3 workers at, respectively, 100 and 110, and letting w^1, the productivity of the lowest skilled workers decrease from 90 to almost 0. In this case, we have found that the socially optimal tax regime is either a full pooling regime (for low and moderate values of the skewness of the distribution of productivities) or a fully separating regime (for high values of the skewness of the distribution of productivities). The results are available upon request.

first-order condition associated with the maximization program in (31) is given by:

$$c'\left(y\left(n^{*}(w)\right)\right)y'\left(n^{*}\left(w\right)\right)=\frac{\partial}{\partial n}g[w,n^{*}\left(w\right)]. \tag{32}$$

Given the tax schedule in place, a fully separating equilibrium satisfies the following properties:

(i) $\frac{\partial n^{*}(w)}{\partial w}>0$;

(ii) $y\left(n^{*}\left(w\right)\right)=n^{*}\left(w\right)\cdot w,\forall w$;

(iii) there exists no labor contract (y', n') that yields non-negative profits when offered in addition to the equilibrium compensation schedule, $y(n)$;

(iv) $\int_{\underline{w}}^{\overline{w}} t\left[y\left(n^{*}\left(w\right)\right)\right] dF(w)\geq 0$.

The first property is an immediate implication of the single-crossing property and defines full separation; namely, each type is offered a distinct bundle. The second property ensures that each labor contract offered in a fully separating equilibrium yields zero profits. The third condition ensures that the schedule $y(n)$ forms a *Nash* equilibrium in the sense that there exists no profitable deviation by a firm operating in the market. The final condition ensures that the government's budget is balanced.

A first result than can be established is that a fully separating equilibrium does not exist.

Proposition 3. *A fully separating equilibrium does not exist.*

Proof See appendix C □

The result stated in Proposition 3 is reminiscent of a similar result described by Riley (2001). There is however a subtle difference between Riley's (2001) framework and our setting. In Riley (2001) there is no tax in place and hence the fact that a fully separating equilibrium fails to exist implies that no equilibrium of any form exists. This is due to the ability of firms to engage in 'cream-skimming'; namely, to offer profitable contracts that attract, from any given pool of types, the higher skilled types only, and sort out their low-skilled counterparts. In our setting, instead, the ability of the government to levy taxes can enable the government to support an equilibrium by preventing firms' 'cream-skimming' strategies.

One simple example is a tax schedule that implements a complete pooling of wages, by taxing all income levels other than that associated with the pooling allocation at a confiscatory rate of 100 percent. Thus, an implication of Proposition 3, which shows that a fully separating equilibrium cannot be sustained even in the presence of government intervention, is that the government optimal re-distributive policy necessarily involves some wage pooling. Another implication, that can be derived as a corollary from the construction of the proof that we present in Appendix C, is that wage pooling necessarily occurs at the lower end of the skill distribution (even though wage pooling may potentially occur over other ranges of the skill distribution as well).

There are potentially numerous possibilities of combining income and wage redistribution. Without imposing more structure on the permissible forms of the income tax schedule, characterizing the optimal redistribution policy turns out to be a daunting task. However, one result that can be shown to hold in a continuous-type setting (and which differentiates the continuous-type case from the discrete-type case) is that complete pooling of wage rates can never be the optimal re-distributive policy; namely, a hybrid equilibrium in which some re-redistribution is done via the wage channel and some through the income channel necessarily constitutes the optimal solution. The fact that full pooling is never an optimum in a continuous-type setting is formally stated in the following Proposition.

Proposition 4. *Full pooling of wages does not form the optimal solution for the government's program.*

Proof See appendix D □

5 Concluding Remarks

There are two different channels via which concerns about inequity could be addressed by income taxation: one is by affecting the post-tax income distribution and the other is by affecting the underlying wage distribution. In the standard Mirrlees (1971) setting, labor markets are competitive and wage rates are exogenously given, as skills are perfect substitutes and perfectly observable by the firms. This leaves no scope for redistribution through the wage channel. Stiglitz (1982) and the subsequent literature challenged this prediction focusing on the role of complementarities across different skill types in the production technology.

In this paper we have employed a setting that maintains the Mirrlees (1971) assumption of perfect substitutability across skill types but allows for asymmetric information between firms and workers. We have demonstrated that in such a context the government can enhance redistribution by affecting the underlying wage distribution through an appropriate choice of the tax and transfer system. The latter limits the possibility for firms to engage in screening, and thereby allows implementing a (partial or full) pooling allocation that gives rise to redistributive gains via cross-subsidization of wage rates across skill types.

The major policy implication of wage pooling is the need for bunching, namely, having several types of workers (optimally) confined to the same income-consumption bundle. Thus, our theory calls for discontinuous jumps in the marginal tax rate at certain points and provides a potential rationalization for the multitude of kinks commonly observed in real-world tax schedules. Such kinks in the income tax schedule may restrict firms' ability to separate amongst workers with different abilities and labor-leisure preferences by means of a menu of non-linear compensation schemes. This mitigates (and potentially obviates) the need to use the income tax schedule for redistribution in response to the inequality in income levels across types which arises from the screening by firms in the labor market. Our analysis of the three-type model and the continuum case indicates that bunching may be desirable at the lower end of the skill distribution, but is less likely to be desirable at the higher end of the distribution.

The emphasis on bunching in this paper contrasts previous contributions in the optimal income tax literature which has considered bunching (in terms of income) to be more of a theoretical curiosity than an important economic result. In fact, a large number of papers in the literature rule out kinks in the optimal income tax schedule by restricting the optimal allocation through certain continutity assumptions.[39]

The message conveyed by our analysis is fairly general and relates to the role of income taxation as an instrument to attain redistribution through the wage channel by limiting the transmission of information between workers and firms. In this paper we have confined attention to one particular mechanism of information transmission; namely, screening by firms, but other channels, such as signaling by workers, may be considered as well.

References

Acemoglu, D. and Pischke, J.-S. (1998). Why do firms train? theory and evidence. *The Quarterly Journal of Economics*, 113(1):79–119.

Cremer, H., Pestieau, P., and Racionero, M. (2011). Unequal wages for equal utilities. *International Tax and Public Finance*, 18(4):383–398.

Diamond, P. A. (1998). Optimal income taxation: An example with a u-shaped pattern of optimal marginal tax rates. *The American Economic Review*, 88:83–95.

[39] An important exception is Ebert (1992) who considers bunching reflected in the pooling of incomes which arises due to the non-implementability of a continuous smooth tax schedule when second order sufficient conditions for optimality are violated (standard models follow a first-order approach and assume the second-order conditions are satisfied).

Ebert, U. (1992). A reexamination of the optimal nonlinear income tax. *Journal of Public Economics*, 49(1):47 – 73.

Gibbons, R. and Katz, L. F. (1991). Layoffs and lemons. *Journal of Labor Economics*, 9(4):pp. 351–380.

Kahn, L. B. (2013). Asymmetric information between employers. *American Economic Journal: Applied Economics*, 5(4):165–205.

Landers, R. M., Rebitzer, J. B., and Taylor, L. J. (1996). Rat race redux: Adverse selection in the determination of work hours in law firms. *The American Economic Review*, 86(3):pp. 329–348.

Mirrlees, J. A. (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 labor markets. *Bell Journal of Economics*, 8(2):394–418.

Riley, J. G. (2001). Silver signals: Twenty-five years of screening and signaling. *Journal of Economic Literature*, 39(2):pp. 432–478.

Rothschild, C. and Scheuer, F. (2013). Redistributive taxation in the Roy model. *The Quarterly Journal of Economics*, 128(2):623–668.

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(4):629–649.

Salanié, B. (2011). *The Economics of Taxation*. The MIT Press.

Spence, A. M. (1978). Product differentiation and performance in insurance markets. *Journal of Public Economics*, 10(3):427–447.

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

Stiglitz, J. E. (1987). Chapter 15 pareto efficient and optimal taxation and the new new welfare economics. In Auerbach, A. J. and Feldstein, M., editors, *Handbook of Public Economics*, volume 2 of *Handbook of Public Economics*, pages 991 – 1042. Elsevier.

Wilson, C. (1977). A model of insurance markets with incomplete information. *Journal of Economic Theory*, 16(2):167 – 207.

A Proof of Proposition 1

Consider the *max-min* case where the government assigns a zero weight to the utility of type-*h* workers, that is, it maximizes the utility of type-*l* workers. The result extends, by continuity considerations, to the case where the weight assigned to type-*l* workers is large enough. By invoking continuity considerations, it suffices to show that the separating allocation dominates the pooling allocation for $w^l = 0$ and $w^l = w^h$ [to establish part (i)], and that there exists some $0 < w^l < w^h$ for which the pooling allocation dominates the separating allocation [to establish part (ii)].

Let us formulate the optimal solution for the government problem under the two alternative regimes: (i) a separating equilibrium; and, (ii) a pooling equilibrium. The welfare-maximizing separating allocation is given by the solution to the following constrained optimization problem formulated in a *Lagrangean* form:

$$\begin{aligned} W^s\left(w^l, w^h, k\right) \equiv \max_{y^l, y^h, T, \lambda} \Bigg[& \left(y^l + T - kg\left(\frac{y^l}{w^l}\right)\right) + \\ & +\lambda\left(y^h - \frac{\gamma^l}{\gamma^h}\cdot T - g\left(\frac{y^h}{w^h}\right) - y^l - T + g\left(\frac{y^l}{\sum_i \gamma^i w^i}\right)\right) \\ & + \mu\left(y^l + T - kg\left(\frac{y^l}{w^l}\right) - y^h + \frac{\gamma^l}{\gamma^h}\cdot T + kg\left(\frac{y^h}{w^h}\right)\right)\Bigg]. \end{aligned} \tag{33}$$

Notice that (IC^h) is necessarily binding in the optimal solution (hence $\lambda > 0$), whereas (IC^l) may be slack (hence $\mu \geq 0$).[40] The welfare-maximizing pooling allocation is given by the solution to the following optimization problem:

$$W^p\left(w^l, w^h, k\right) \equiv \max_y \left(y - kg\left(\frac{y}{\sum_i \gamma^i w^i}\right)\right). \tag{34}$$

We turn next to establish part (i) of the Proposition. Consider first the case where $w^l = 0$. We first show that in this case (IC^l) is slack in the optimal solution for the separating regime [that is $\mu = 0$ in the *Lagrangean* expression in (33)]. To see this, first notice that when $w^l = 0$, both $n^l = 0$ and $y^l = 0$ [by virtue of the zero profit condition, (ZP^l)]. The constraint (IC^h) necessarily binds in the optimal solution. Suppose, then, by way of contradiction, that (IC^l) binds as well. Formulating the two (presumably

[40]Notice that the standard single crossing property doesn't apply, which means that if (IC^h) is binding in a separating allocation, this does not imply that (IC^l) is slack.

binding) incentive compatibility constraints yields:

$$y^h - \frac{\gamma^l}{\gamma^h} \cdot T - g\left(\frac{y^h}{w^h}\right) = T, \tag{35}$$

$$T = y^h - \frac{\gamma^l}{\gamma^h} \cdot T - kg\left(\frac{y^h}{w^h}\right). \tag{36}$$

Substituting for T from (35) in (36) and re-arranging yields:

$$(k-1)g\left(\frac{y^h}{w^h}\right) = 0.$$

As $k > 1$, it follows from (37) that $y^h = 0$, which violates the definition of a separating equilibrium (recalling that $y^l = 0$). We thus obtain a contradiction.

Employing the fact that (IC^l) slacks in the optimal solution for the separating regime, substituting for $w^l = 0$ into the expressions in (33) and (34) and further substituting for T from the binding (IC^h) constraint, yields upon re-arrangement:

$$W^s\left(0, w^h, k\right) \equiv \max_{y^h} \gamma^h \left(y^h - g\left(\frac{y^h}{w^h}\right)\right),$$
$$W^p\left(0, w^h, k\right) \equiv \max_{y} \left(y - kg\left(\frac{y}{\gamma^h w^h}\right)\right).$$

As $k > 1$, it follows that $W^p\left(0, w^h, k\right) < \max_y \left(y - g\left(\frac{y}{\gamma^h w^h}\right)\right)$. Hence, to establish that the separating allocation dominates the pooling allocation it suffices to show that

$$\frac{1}{\gamma^h} \max_y \left(y - g\left(\frac{y}{\gamma^h w^h}\right)\right) < \max_{y^h} \left(y^h - g\left(\frac{y^h}{w^h}\right)\right).$$

Letting $\hat{y} = \text{argmax}_y \left(y - g\left(\frac{y}{\gamma^h w^h}\right)\right)$, it follows:

$$\frac{1}{\gamma^h}\max_y \left(y - g\left(\frac{y}{\gamma^h w^h}\right)\right) = \frac{1}{\gamma^h}\left(\hat{y} - g\left(\frac{\hat{y}}{\gamma^h w^h}\right)\right)$$
$$< \frac{\hat{y}}{\gamma^h} - g\left(\frac{\hat{y}}{\gamma^h w^h}\right) \le \max_{y^h} \left(y^h - g\left(\frac{y^h}{w^h}\right)\right),$$

where the first (strict) inequality follows from $0 < \gamma^h < 1$ and the second (weak) inequality follows by construction of the maximization. Thus, we establish that for $w^l = 0$, the separating allocation dominates the pooling allocation.

We turn next to the case where $w^l = w^h = w$. Let $\hat{n}$ denote the welfare-maximizing number of hours worked by both types of agents under the pooling regime. Formally,

$$\hat{n} = \underset{n}{\operatorname{argmax}} \Big(nw - kg\left(n\right) \Big). \tag{37}$$

Consider the following separating allocation obtained as a small perturbation to the welfare-maximizing pooling regime: $n^h = \hat{n} + \epsilon$, $c^h = \hat{n}w + \epsilon g'\left(\hat{n}\right), n^l = \hat{n} - \epsilon \frac{\gamma^h}{\gamma^l}$ and $c^l = \hat{n}w - \epsilon \frac{\gamma^h}{\gamma^l} kg'\left(\hat{n}\right)$, where $\epsilon > 0$. Invoking a first-order approximation, notice that the suggested perturbation amounts to shifting the bundle of type h upwards and that of type l downwards, along their respective indifference curves (in the n-c space) going through the welfare-maximizing pooling allocation. As, by virtue of the fact that $k > 1$, the indifference curve associated with the low-skilled worker is steeper than that associated with the high-skilled worker, the suggested perturbation induces an allocation which is incentive compatible. Thus, to show that the separating allocation dominates the pooling allocation, it suffices to show that the suggested perturbation yields a fiscal surplus, as the latter could then be rebated in a lump-sum fashion to attain a Pareto improvement. The tax revenues associated with the suggested separating allocation are given by the following expression [substituting for $\left(n^i, c^i\right); i = l, h$, from the suggested perturbation]:

$$\begin{aligned} TR\left(\epsilon\right) &= \Big(\gamma^l n^l + \gamma^h n^h\Big) w - \Big(\gamma^l c^l + \gamma^h c^h\Big) = \\ &= \hat{n}w - \Big(\hat{n}w - \epsilon\gamma^h k g'\left(\hat{n}\right) + \epsilon\gamma^h g'\left(\hat{n}\right)\Big) = \epsilon\gamma^h\left(k-1\right) g'\left(\hat{n}\right) > 0, \end{aligned} \tag{38}$$

where the inequality follows as $k > 1$. Thus, we establish that for $w^l = w^h$, the separating allocation dominates the pooling allocation.

We turn next to establish part (ii) of the Proposition, by demonstrating that when the differences in the labor-leisure preferences between the two types of workers are sufficiently small, there exists some $0 < w^l < w^h$ for which the pooling allocation dominates the separating allocation. In particular, we will assume that $k = 1 + \epsilon$, with $\epsilon > 0$ and small, and show that the pooling allocation dominates the separating allocation for $w^l = w^h - \epsilon$.

We will prove the result under the assumption that (IC^l) is slack in the optimal solution for the separating regime. It suffices to do so, because, if the pooling equilibrium dominates the separating equilibrium when (IC^l) does not bind then this is clearly the case also when allowing for the possibility that (IC^l) binds (since that would imply that the welfare associated with the separating equilibrium would be even lower).

Notice that when $w^l = w^h$ and $k = 1$ there are no differences between the two types of workers (in term of productivities and/or labor-leisure preferences), hence, both regimes coincide, namely, $W^p = W^s$. Invoking a first order approximation, it follows that:

$$W^s\left(w^h-\epsilon, w^h, 1+\epsilon, 1\right) = W^s\left(w^h, w^h, 1\right) - \epsilon\frac{\partial W^s\left(w^h, w^h, 1\right)}{\partial w^l} + \epsilon\frac{\partial W^s\left(w^h, w^h, 1\right)}{\partial k}, \tag{39}$$

$$W^p\left(w^h-\epsilon, w^h, 1+\epsilon, 1\right) = W^p\left(w^h, w^h, 1\right) - \epsilon\frac{\partial W^p\left(w^h, w^h, 1\right)}{\partial w^l} + \epsilon\frac{\partial W^p\left(w^h, w^h, 1\right)}{\partial k}, \tag{40}$$

where $\epsilon > 0$ and small. Subtracting (39) from (40) yields:

$$W^p\left(w^h-\epsilon, w^h, 1+\epsilon, 1\right) - W^s\left(w^h-\epsilon, w^h, 1+\epsilon, 1\right) =$$
$$\epsilon\left(-\frac{\partial W^p\left(w^h, w^h, 1\right)}{\partial w^l} + \frac{\partial W^p\left(w^h, w^h, 1\right)}{\partial k} + \frac{\partial W^s\left(w^h, w^h, 1\right)}{\partial w^l} - \frac{\partial W^s\left(w^h, w^h, 1\right)}{\partial k}\right). \tag{41}$$

It suffices then to prove that the term within parenthesis in (41) is positive.

Assuming that (IC^l) is slack, differentiation of the expressions in (33) and (34) with respect to w^l, employing the envelope condition, yields:

$$\frac{\partial W^s\left(w^l, w^h, k\right)}{\partial w^l} = kg'\left(\frac{y^l}{w^l}\right)\frac{y^l}{w^{l2}} - \lambda g'\left(\frac{y^l}{\sum_i \gamma^i w^i}\right)\frac{\gamma^l y^l}{\left(\sum_i \gamma^i w^i\right)^2}, \tag{42}$$

$$\frac{\partial W^p\left(w^l, w^h, k\right)}{\partial w^l} = kg'\left(\frac{y}{\sum_i \gamma^i w^i}\right)\frac{\gamma^l y}{\left(\sum_i \gamma^i w^i\right)^2}. \tag{43}$$

Moreover, by differentiating (33) with respect to T and equating to zero it follows that $\lambda = \gamma^h$. Substituting for λ, w^l and k into the derivative expressions in (42) and (43), employing the fact that when $w^l = w^h$ and $k = 1$ it follows that $y^l = y = y^h$, yields upon rearrangement:

$$\frac{\partial W^s\left(w^h, w^h, 1\right)}{\partial w^l} = (1-\gamma^h\gamma^l)g'\left(\frac{y^h}{w^h}\right)\frac{y^h}{w^{h2}},$$
$$\frac{\partial W^p\left(w^h, w^h, 1\right)}{\partial w^l} = \gamma^l g'\left(\frac{y^h}{w^h}\right)\frac{y^h}{w^{h2}}.$$

Thus,

$$\frac{\partial W^s\left(w^h, w^h, 1\right)}{\partial w^l} - \frac{\partial W^p\left(w^h, w^h, 1\right)}{\partial w^l} = (1-\gamma^h\gamma^l-\gamma^l)g'\left(\frac{y^h}{w^h}\right)\frac{y^h}{w^{h2}} > 0, \tag{44}$$

where the inequality sign follows as $\gamma^h = 1-\gamma^l$ and $\gamma^l < 1$.

Differentiation of the expressions in (33) and (34) (assuming $\mu = 0$) with respect to

k, employing the envelope condition yields:

$$\frac{\partial W^s\left(w^l,w^h,k\right)}{\partial k}=-g\left(\frac{y^l}{w^l}\right), \tag{45}$$

$$\frac{\partial W^p\left(w^l,w^h,k\right)}{\partial k}=-g\left(\frac{y}{\sum_i \gamma^i w^i}\right). \tag{46}$$

Evaluating (45) and (46) at $w^l = w^h$ and $k = 1$, employing the fact that when $w^l = w^h$ and $k = 1$ it follows that $y^l = y = y^h$, yields:

$$\frac{\partial W^s\left(w^h,w^h,1\right)}{\partial k}=\frac{\partial W^p\left(w^h,w^h,1\right)}{\partial k}=-g\left(\frac{y^h}{w^h}\right). \tag{47}$$

Thus,

$$\frac{\partial W^p\left(w^h,w^h,1\right)}{\partial k}-\frac{\partial W^s\left(w^h,w^h,1\right)}{\partial k}=0. \tag{48}$$

Substituting from (48) and (44) into (41) then completes the proof of part (ii).

Finally, we turn to prove part (iii). By parts (i) and (ii) it follows that the second-best welfare frontier contains at least two regime switches (from separation to pooling and from pooling to separation). By continuity considerations, with any regime switch there exists an associated w^l for which the two regimes (separating and pooling) yield the same welfare level. Now, suppose by negation, that there exist more than two regime switches. Fix $k > 1$ and $w^h > 0$ and let the set of all w^l for which the two regimes (pooling and separating) yield the same welfare level be denoted by $G\left(k, w^h\right)$. Formally, $G\left(k, w^h\right) = \left\{w^l \left| W^p\left(w^l, w^h, k\right) = W^s\left(w^l, w^h, k\right)\right.\right\}$. By our presumption the set G contains at least 3 elements. Let the elements of the set G be arranged in an increasing order, namely: $w^l_1 < w^l_2 < ... < w^l_j$, where by our presumption $j \geq 3$. Now consider a pair of adjacent elements of the set G, w^l_i and w^l_{i+1}, with $1 \leq i \leq j-1$. By continuity considerations, it follows by Rolle's Theorem, that there exists some $w^l_i < w^l < w^l_{i+1}$, for which $\frac{\partial\left[W^p\left(w^l,w^h,k\right)-W^s\left(w^l,w^h,k\right)\right]}{\partial w^l} = 0$. As by presumption $j \geq 3$, it follows that there exist at least two levels of w^l for which the derivative $\frac{\partial\left[W^p\left(w^l,w^h,k\right)-W^s\left(w^l,w^h,k\right)\right]}{\partial w^l} = 0$. However, when $H(w^l, w^h, k) \equiv W^p\left(w^l, w^h, k\right) - W^s\left(w^l, w^h, k\right)$ is strictly concave with respect to w^l, there exists at most one level of w^l for which the derivative is zero. We thus obtain the desired contradiction. It follows that there exist exactly two regime switches and therefore $w^1 = w^2$ and $w^3 = w^4$. This completes the proof.

B Proof of Proposition 2

Consider the *max-min* case where the government assigns a zero weight to the utility of type-h workers, that is, it maximizes the utility of type-l workers. The result extends, by continuity considerations, to the case where the weight assigned to type-l workers is large enough.

We will first show that when the differences in productivities are sufficiently small, then the pooling allocation dominates the separating allocation when the labor-leisure differences are small enough. Formally, let $w^l = w^h - \epsilon$, with $\epsilon > 0$ and small. Suppose that $k = 1 + \epsilon$. By replicating the arguments in the proof of part (ii) of Proposition 1 (the arguments are hence omitted), it follows that the pooling allocation dominates the separating allocation.

We turn next to show that, when k is sufficiently large, the separating allocation dominates the pooling allocation.

We begin by noticing that as $k \to \infty$, both $n^l \to 0$ and $y^l \to 0$. Thus, replicating the arguments used in the proof of the part (i) of Proposition 1 (details are hence omitted), it follows by continuity considerations that when k is sufficiently large the incentive constraint (IC^l) is slack in the optimal solution for the separating regime. Recalling the Lagrangean expression for the welfare level of the separating equilibrium given by equation (33) (with $\mu = 0$) we consider the following (relaxed) maximization program:

$$W^s\left(w^l, w^h, k,\ y^l = 0\right) \equiv \max_{y^l=0, y^h, T, \lambda} \left(y^l + T - kg\left(\frac{y^l}{w^l}\right)\right) + \\ + \lambda\left(y^h - \frac{\gamma^l}{\gamma^h} \cdot T - g\left(\frac{y^h}{w^h}\right) - y^l - T + g\left(\frac{y^l}{\sum_i \gamma^i w^i}\right)\right). \quad (49)$$

The maximization program given in (49) is obtained as a restriction of the separating regime maximization to the case where $y^l = 0$. Notice that as $g(0) = 0$ and as the solution for the maximization in (49) entails redistribution from type-h to type-l workers $(T > 0)$, it follows that $W^s\left(w^l, w^h, k,\ y^l = 0\right) > 0$. Moreover, as by construction $y^l = 0$, it follows that $W^s\left(w^l, w^h, k,\ y^l = 0\right)$ is independent of k.

Now consider the pooling regime maximization given in (34). Notice that when $k \to \infty$ it follows that $y \to 0$ and the welfare level associated with the optimal pooling regime converges to zero (since $g(0) = 0$) . Hence, by continuity considerations, it follows that for a sufficiently high value of k, the welfare level associated with the optimal pooling regime is lower than that associated with the maximization program in (49), which is bounded away from zero and independent of k. As the welfare level associated with the maximization program in (49) is bounded from above by the optimal separating regime (where y^l is not constrained to be zero), we conclude that for k sufficiently large the separating allocation dominates the pooling allocation. By continuity considerations, and by virtue of the intermediate-value theorem, when the differences in productivities between the two types of workers are small, there exists some k for which the two regimes yield the same welfare level. Denote the value of k for which the two regimes yield the same welfare level by $\hat{k}$. We turn next to show that $\hat{k}$ is unique, which implies then by continuity considerations, that for any $k > \hat{k}$ the separating regime dominates, whereas, for any $k < \hat{k}$ the pooling regime prevails.

To establish the uniqueness of $\hat{k}$ it then suffices to show that:

$$\frac{\partial W^p\left(w^l, w^h, k\right)}{\partial k} - \frac{\partial W^s\left(w^l, w^h, k\right)}{\partial k} = -g\left(\frac{y}{\sum_i \gamma^i w^i}\right) + g\left(\frac{y^l}{w^l}\right) - \mu\left(g\left(\frac{y^h}{w^h}\right) - g\left(\frac{y^l}{w^l}\right)\right) < 0. \quad (50)$$

which follows from differentiation of the expressions (33) and (34) with respect to k and employing the envelope theorem. As g is strictly increasing and $\mu \geq 0$, to prove the condition in (50) it suffices to show that:

$$\text{(i)} \quad \frac{y}{\sum_i \gamma^i w^i} > \frac{y^l}{w^l} \quad \text{and} \quad \text{(ii)} \quad \frac{y^h}{w^h} > \frac{y^l}{w^l} \quad (51)$$

Formulating the first order conditions with respect to y^l and y^h by differentiating the *Lagrangean* expression in (33) yields:

$$(1+\mu)\left[1 - kg'\left(\frac{y^l}{w^l}\right)\frac{1}{w^l}\right] + \lambda\left[-1 + g'\left(\frac{y^l}{\sum_i \gamma^i w^i}\right)\frac{1}{\sum_i \gamma^i w^i}\right] = 0, \quad (52)$$

$$\lambda\left[1 - g'\left(\frac{y^h}{w^h}\right)\frac{1}{w^h}\right] - \mu\left[1 - kg'\left(\frac{y^h}{w^h}\right)\frac{1}{w^h}\right] = 0. \quad (53)$$

Formulating the first order condition with respect to y by differentiating the expression in (34) yields:

$$1 - kg'\left(\frac{y}{\sum_i \gamma^i w^i}\right)\frac{1}{\sum_i \gamma^i w^i} = 0. \quad (54)$$

Now, let $\tilde{y}$ denote the implicit solution to:

$$1 - kg'\left(\frac{\tilde{y}}{w^l}\right)\frac{1}{w^l} = 0. \quad (55)$$

Formulating the first-order condition with respect to T, by differentiating the *La-*

grangean expression in (33), yields:

$$1 - \frac{\lambda - \mu}{\gamma^h} = 0. \tag{56}$$

By virtue of condition (56) and as $0 < \gamma^h < 1$ it follows that $1 + \mu > \lambda \geq 0$. Therefore, by the strict convexity of g and as k>1 and $\sum_i \gamma^i w^i > w^l$, it follows from condition (52) that:

$$1 - kg'\left(\frac{y^l}{w^l}\right)\frac{1}{w^l} > 0. \tag{57}$$

Comparing the expressions in (55) and (57) it therefore follows, by virtue of the strict convexity of g that:

$$y^l < \tilde{y}. \tag{58}$$

Comparing the expressions in (54) and (55), it follows by the strict convexity of g and as $\sum_i \gamma^i w^i > w^l$ that:

$$\frac{y}{\sum_i \gamma^i w^i} > \frac{\tilde{y}}{w^l}. \tag{59}$$

Combining (58) and (59) establishes (i) in condition (51).

By virtue of condition (56) it follows that $\lambda > \mu \geq 0$. Therefore, as $k > 1$ and g is strictly increasing, it follows from condition (53) that:

$$1 - kg'\left(\frac{y^h}{w^h}\right)\frac{1}{w^h} < 0. \tag{60}$$

Comparing conditions (55) and (60) implies, by virtue of the strict convexity of g and as $w^h > w^l$, that:

$$\frac{y^h}{w^h} > \frac{\tilde{y}}{w^l}. \tag{61}$$

Combining conditions (58) and (61) establishes condition (ii) in (51). It follows that condition (50) holds.

Finally notice that by virtue of condition (50) it follows that the welfare gain associated with a switch from a separating to a pooling allocation is decreasing with respect to k. This concludes the proof.

C Proof of Proposition 3

Consider some worker-type $\hat{w}$ where $\hat{w} > \underline{w}$ and let $\hat{y}$ denote the gross level of income that satisfies the following equality:

$$c\left(y[n^*\left(\hat{w}\right)]\right) - g\left[\hat{w}, n^*\left(\hat{w}\right)\right] = c\left(\hat{y}\right) - g\left[\hat{w}, n^*\left(\underline{w}\right)\right]. \tag{62}$$

In words, the worker with ability $\hat{w}$ is just indifferent between his/her optimal contract along the (presumed) equilibrium compensation schedule $y(n)$ and the contract $\left(n^*\left(\underline{w}\right), \hat{y}\right)$. By virtue of the single-crossing property, any worker with ability type $w < \hat{w}$ will prefer the contract $\left(n^*\left(\underline{w}\right), \hat{y}\right)$ to his/her respective optimal contract along the schedule $y(n)$; on the other hand, any worker with ability type $w > \hat{w}$ will prefer his/her optimal contract along the (presumed) equilibrium compensation schedule to the contract $\left(n^*\left(\underline{w}\right), \hat{y}\right)$.

The profits associated with the contract $\left(n^*\left(\underline{w}\right), \hat{y}\right)$ are therefore given by:

$$\pi\left(\hat{w}\right) = n^*\left(\underline{w}\right) \cdot \int_{\underline{w}}^{\hat{w}} w dF\left(w\right) - F(\hat{w}) \cdot \hat{y}.$$

Let $\bar{\pi}(\hat{w})$ denote the per-worker profits associated with the contract $\left(n^*\left(\underline{w}\right), \hat{y}\right)$. Formally,

$$\bar{\pi}\left(\hat{w}\right) = n^*\left(\underline{w}\right) \cdot \bar{v}\left(\hat{w}\right) - \hat{y}, \tag{63}$$

where $\bar{\pi}\left(\hat{w}\right) = \pi\left(\hat{w}\right)/F(\hat{w})$ and $\bar{v}\left(\hat{w}\right) = \int_{\underline{w}}^{\hat{w}} w dF\left(w\right)/F(\hat{w})$.

Substituting for $\hat{y}$ from (62) into (63) yields:

$$\bar{\pi}\left(\hat{w}\right) = n^*\left(\underline{w}\right) \cdot \bar{v}\left(\hat{w}\right) - c^{-1}\left(c\left(y\left[n^*\left(\hat{w}\right)\right]\right) - g\left[\hat{w}, n^*\left(\hat{w}\right)\right] + g\left[\hat{w}, n^*\left(\underline{w}\right)\right]\right),$$

where c is invertible by virtue of the single-crossing property and the full separation property of the (presumed) equilibrium.

Differentiating $\bar{\pi}\left(\hat{w}\right)$ with respect to $\hat{w}$, and employing the envelope theorem, yields:

$$\partial\bar{\pi}\left(\hat{w}\right)/\partial\hat{w} = n^*\left(\underline{w}\right) \cdot \bar{v}'(\hat{w}) - c^{-1'}(\cdot) \cdot \left(-\frac{\partial}{\partial\hat{w}} g\left[\hat{w}, n^*\left(\hat{w}\right)\right] + \frac{\partial}{\partial\hat{w}} g\left[\hat{w}, n^*\left(\underline{w}\right)\right]\right)$$

Taking the limit when $\hat{w} \to \underline{w}$ yields:

$$\lim_{\hat{w}\to\underline{w}} \partial\bar{\pi}\left(\hat{w}\right)/\partial\hat{w} = n^*\left(\underline{w}\right) \cdot \bar{v}'(\underline{w}) > 0.$$

Moreover, by virtue of property (ii) in the definition of the fully separating equilibrium (the zero profit condition), it follows that:

$$\bar{\pi}\left(\underline{w}\right)=0.$$

Invoking a first-order approximation, it then follows that for w sufficiently close to $\underline{w}$, $\bar{\pi}\left(w\right)>0$. Thus, there exists a $\hat{y}$ sufficiently close to $y\left[n^{*}\left(\underline{w}\right)\right]$, such that the contract $\left(n^{*}\left(\underline{w}\right),\hat{y}\right)$ yields strictly positive profits. However, this violates property (iii) in the definition of the fully separating equilibrium. This concludes the proof.

D Proof of Proposition 4

Consider a tax schedule with only two available bundles featuring positive levels of after-tax income: $(\overline{y},\overline{t})$ and $(\underline{y},\underline{t})$, where y denotes the gross level of income and t the tax (transfer when negative) associated with y, and where $\overline{y}-\overline{t}>\underline{y}-\underline{t}$. Given the tax schedule in place, a RS separating equilibrium is given by a cutoff level of ability, $\hat{w}$, and work hours $\overline{n}$ and $\underline{n}$, $\overline{n}>\underline{n}$, associated respectively with $\overline{y}$ and $\underline{y}$, that satisfy the following set of properties:

$$\underline{y}-\underline{t}-g\left(\hat{w},\underline{n}\right)=\overline{y}-\overline{t}-g(\hat{w},\overline{n}); \tag{64}$$

$$\underline{y}=\underline{n}\cdot\int_{\underline{w}}^{\hat{w}}\frac{wdF\left(w\right)}{F\left(\hat{w}\right)},\overline{y}=\overline{n}\cdot\int_{\hat{w}}^{\overline{w}}\frac{wdF\left(w\right)}{\left[1-F\left(\hat{w}\right)\right]}; \tag{65}$$

$$F\left(\hat{w}\right)\cdot\underline{t}+\left[1-F\left(\hat{w}\right)\right]\cdot\overline{t}\geq 0; \tag{66}$$

$$n\int_{\underline{w}}^{w(n)}wdF\left(w\right)-\underline{y}F\left[w\left(n\right)\right]<0,\forall n<\underline{n}, \tag{67}$$

where $w(n)$ is implicitly given by: $\underline{y}-\underline{t}-g\left(w\left(n\right),n\right)=\overline{y}-\overline{t}-g(w(n),\overline{n})$.

By virtue of the single-crossing property, the cutoff level defined by condition (64) implies that all workers with ability level exceeding the threshold, $\hat{w}$, will choose the bundle $(\overline{y},\overline{t})$, whereas, all workers with ability level lower than the cutoff level, $\hat{w}$, will choose the bundle $(\underline{y},\underline{t})$. Condition (65) states that each of the two labor contracts offered in equilibrium $[(\underline{n},\ \underline{y})$ and $(\overline{n},\ \overline{y})]$ yields zero profits. Condition (66) states that the public budget is balanced. Finally, condition (67) ensures that any contract other than the two offered in equilibrium would yield negative profits. To see this, notice that by assumption the government cannot control directly the working hours specified by each labor contract. As firms can only offer a compensation level that coincides with one of the two levels of income associated with the tax schedule, the only margin of maneuver for firms is to reduce the working hours (associated with any given level of compensation). Condition (67) considers the only potentially profitable deviation, which is a reduction in the working hours associated with the compensation level $\underline{y}$ to some $n < \underline{n}$. Such a deviation will result in two effects: (i) a mechanical effect due to the decrease in the working hours (given the maintained level of compensation) that results in a reduction in profits; (ii) a behavioral effect due to the attraction of additional workers whose ability level lies within the interval $[\hat{w}, w(n)]$ (namely, workers with higher-than-average productivity levels) that results in an increase in profits.[41] Condition (67) states that the combined (mechanical and behavioral) effect results in a decrease in profits and, hence, renders the deviation unprofitable. Notice that we only consider a deviation at the lower level of compensation, since at the higher level of compensation the behavioral effect will induce a further decrease in profits by attracting workers whose ability level satisfies $w < \hat{w}$, namely, workers with lower-than-average ability levels.[42]

We turn next to show that full pooling does not constitute the optimal solution for the government's problem. The government is seeking to maximize the wellbeing of the least well-off individual, $\underline{w}$, given by:

$$u\left(\underline{w}, \underline{n}, \underline{y} - \underline{t}\right) = \underline{y} - \underline{t} - g(\underline{w}, \underline{n}), \tag{68}$$

subject to constraints (64)-(67), by setting: $\overline{y}, \underline{y}\ \overline{t},\ \underline{t}, \overline{n}, \underline{n}$, and $\hat{w}$.

We will henceforth assume that condition (67) is satisfied and focus on constraints (64)-(66).[43]

Notice that by setting $\hat{w} = \overline{w}$, the government implements a full wage-pooling allocation, in which all workers receive an hourly wage rate equaling the economy-wide average productivity and re-distribution is carried out exclusively via the wage channel (by virtue of condition (66), $\underline{t} = 0$, hence no income re-distribution takes place). In order to show that full wage-pooling does not constitute the optimal solution, it suffices to show that the welfare level associated with some cutoff level, $\hat{w} < \overline{w}$, exceeds the welfare level associated with the optimal full wage-pooling allocation.

One can reformulate the government's problem as an unconstrained maximization

[41]Recall that the average productivity associated with workers that choose the lower level of compensation is given by $\int_{\underline{w}}^{\hat{w}} \frac{wdF(w)}{F(\hat{w})}$.

[42]Recall that the average productivity associated with workers that choose the higher level of compensation is given by $\int_{\hat{w}}^{\overline{w}} \frac{wdF(w)}{[1-F(\hat{w})]}$.

[43]Notice that for our purposes, i.e. to prove Proposition 4, assuming that (67) is satisfied is not restrictive. The reason is that our proof relies on showing that, for $\hat{w}$ sufficiently close to $\overline{w}$, the welfare level attained by the optimal two-bundles tax-and-transfer system associated with the cutoff level $\hat{w}$ exceeds the welfare level associated with the optimal full wage-pooling allocation. The deviation considered in (67) entails a reduction in the working hours associated with the compensation level $\underline{y}$ to some $n < \underline{n}$. However, when $\hat{w}$ is close to $\overline{w}$, such a deviation generates a very large mechanical effect, resulting in a reduction in profits, given that the size of the workers' population whose labor supply is reduced is given by $F(\hat{w})$. On the other hand, the increase in profits associated with the behavioral effect is quite small, given that only few additional higher skilled agents (less than *1-F* $(\hat{w})$) can be attracted to the $(\underline{y}, \underline{t})$-bundle. Thus, when attention is restricted to the cases when $\hat{w}$ is sufficiently close to $\overline{w}$, one can safely disregard constraint (67).

program by eliminating constraints (64)-(66) through substitution. Substituting for $\underline{y}$, $\overline{y}$ and $\overline{t}$ from (65) and (66) into (64) yields upon re-arrangement:

$$\underline{t} = [1 - F(\hat{w})] \cdot \left[\underline{n} \cdot \int_{\underline{w}}^{\hat{w}} \frac{wdF(w)}{F(\hat{w})} - g(\hat{w}, \underline{n}) - \overline{n} \cdot \int_{\hat{w}}^{\overline{w}} \frac{wdF(w)}{[1 - F(\hat{w})]} + g(\hat{w}, \overline{n})\right]. \quad (69)$$

Substituting for $\underline{y}$ [from (65)] and for $\underline{t}$ [from (69)] into (68) yields:

$$V(\underline{w}, \underline{n}, \overline{n}, \hat{w}) \equiv \left\{\underline{n} \cdot \int_{\underline{w}}^{\hat{w}} \frac{wdF(w)}{F(\hat{w})} - [1 - F(\hat{w})] \cdot \left[\underline{n} \cdot \int_{\underline{w}}^{\hat{w}} \frac{wdF(w)}{F(\hat{w})} - g(\hat{w}, \underline{n}) \right.\right.$$
$$\left.\left. -\overline{n} \cdot \int_{\hat{w}}^{\overline{w}} \frac{wdF(w)}{[1 - F(\hat{w})]} + g(\hat{w}, \overline{n})\right] - g(\underline{w}, \underline{n})\right\} \quad (70)$$

The government's problem is thus given by the (unconstrained) maximization of (70) with respect to $\underline{n}, \overline{n}$ and $\hat{w}$.

Formulating the first-order conditions with respect to $\underline{n}$ and $\overline{n}$, and taking the limit where $\hat{w} \to \overline{w}$ yields, respectively:

$$\frac{\partial}{\partial \underline{n}} g(\underline{w}, \underline{n}) = \int_{\underline{w}}^{\overline{w}} wdF(w); \quad (71)$$
$$\frac{\partial}{\partial \overline{n}} g(\overline{w}, \overline{n}) = \overline{w}. \quad (72)$$

By virtue of the fact that $\frac{\partial^2 g}{\partial n^2} > 0$ and $\frac{\partial^2 g}{\partial n \partial w} < 0$ it follows from conditions (71) and (72) that $\overline{n} > \underline{n}$.

Differentiating the expression in (70) with respect to $\hat{w}$ and taking the limit where $\hat{w} \to \overline{w}$ yields upon re-arrangement:

$$\lim_{\hat{w} \to \overline{w}} \frac{\partial}{\partial \hat{w}} V(\underline{w}, \underline{n}, \overline{n}, \hat{w}) = F'(\overline{w}) \cdot [[\underline{n} \cdot \overline{w} - g(\overline{w}, \underline{n})] - [\overline{n} \cdot \overline{w} - g(\overline{w}, \overline{n})]] < 0. \quad (73)$$

The inequality sign follows from the fact that $\overline{n} > \underline{n}$ and by virtue of (72) which implies that $\overline{n} = \text{argmax}_n [n \cdot \overline{w} - g(\overline{w}, n)]$.

The inequality condition in (73) implies that for $\hat{w}$ sufficiently close to $\overline{w}$, the welfare level attained by the optimal two-bundles tax-and-transfer system associated with the cutoff level $\hat{w}$ exceeds the welfare level associated with the optimal full wage-pooling allocation. This concludes the proof.

E Proof of Lemma 1

Suppose by way of contradiction that a hybrid equilibrium in which types 1 and 3 are bunched together exists. Let the net income (consumption) levels in the bundles associated with types 1 and 3 (who are bunched together), and type 2, be respectively denoted by c^{13} and c^2. We need to examine two separate cases: (i) $c^{13} > c^2$, and, (ii) $c^{13} < c^2$.

Let the two labor contracts associated, in equilibrium, with types 1 and 3, and type 2, be respectively denoted by (n^{13}, y^{13}) and (n^2, y^2). Consider first case (i). In this case it must be that $n^{13} > n^2$, otherwise, all types would prefer the contract (n^{13}, y^{13}) which generates the bundle (n^{13}, c^{13}). However, by virtue of the single-crossing property, the fact that type 1 weakly prefers the bundle (n^{13}, c^{13}) to the bundle (n^2, c^2) implies that type 2 would strictly prefer (n^{13}, c^{13}) to (n^2, c^2). We therefore obtain a contradiction to the definition of the hybrid equilibrium.

We turn next to case (ii). In this case it must be that $n^{13} < n^2$, otherwise all types would prefer the contract (n^2, y^2) which generates the bundle (n^2, c^2). However, by virtue of the single-crossing property, the fact that type 2 weakly prefers the bundle (n^2, c^2) to the bundle (n^{13}, c^{13}) implies that type 3 would strictly prefer (n^2, c^2) to (n^{13}, c^{13}). We therefore obtain a contradiction to the definition of the hybrid equilibrium. This concludes the proof.