



The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
<http://ageconsearch.umn.edu>
aesearch@umn.edu

Papers downloaded from AgEcon Search may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.

No endorsement of AgEcon Search or its fundraising activities by the author(s) of the following work or their employer(s) is intended or implied.

Taxes and Employment Subsidies
in Optimal Redistribution Programs
(Revised Version)

Paul Beaudry, Charles Blackorby, and Dezso Szalay

No 779

WARWICK ECONOMIC RESEARCH PAPERS

DEPARTMENT OF ECONOMICS

THE UNIVERSITY OF
WARWICK

Taxes and Employment Subsidies in Optimal Redistribution Programs*

Paul Beaudry, Charles Blackorby, and Dezsö Szalay †

October 24, 2006

*We thank Tony Atkinson, Tim Besley, Richard Blundell, Craig Brett, Louis-André Gerard-Varet, Roger Guesnerie, Jon Kesselman, Thomas Piketty, Michel Poitevin, John Weymark and seminar participants at UBC, CIAR, CORE, Ecole Nationale des Ponts et Chaussées, The Institute for Fiscal Studies, Paris I, The University of Exeter, The University of Nottingham and The University of Warwick for comments. Financial support through a grant from the Social Sciences and Humanities Research Council of Canada is gratefully acknowledged. Finally we have benefited significantly from the remarks of three referees and the editor of this journal. Send correspondence to Charles Blackorby at Economics Department, University of Warwick, Gibbet Hill Road, Coventry, UK, CV4 7AL; c.blackorby@warwick.ac.uk.

†Paul Beaudry: University of British Columbia, and NBER. (Email: Beaudry@econ.ubc.ca)
Charles Blackorby: University of Warwick and GREQAM (Email: c.blackorby@warwick.ac.uk)
Dezsö Szalay: University of Warwick (Email: Dezso.Szalay@warwick.ac.uk)

Abstract

This paper explores how to optimally set tax and transfers when taxation authorities: (1) are uninformed about individuals' value of time in both market and non-market activities and (2) can observe both market-income and time allocated to market employment. We show that optimal redistribution in this environment involves distorting market employment upwards for low wage individuals through decreasing wage-contingent employment subsidies, and distorting employment downwards for high wage individuals through positive and increasing marginal income tax rates. In particular, we show that whether a person is taxed or subsidized depends primarily on his wage, that is, the optimal program involves a cut-off wage whereby workers above the cutoff are taxed as they increase their income, while workers earning a wage below the cutoff receive an income supplement (an earned income tax credit) as they increase their income. Finally, we show that the optimal program transfers zero income to individuals who choose not to work.

Journal of Economic Literature Classification Numbers: D82, H21, H 23.

Key words: Taxation, Redistribution, Wage Subsidies Screening.

1 Introduction

In most countries income redistribution is achieved through a variety of programs: these include direct income taxation, employment programs, welfare, unemployment insurance and pension schemes. Viewed as a whole, these programs create intricate incentives and complex redistribution patterns. Since the conditionality of these programs is quite varied, they generally result in a net tax-transfer system that depends not only on income but often depends on the extent of market participation as well. Reasoned economic policy should attempt to identify whether or not these programs are mutually consistent with the goal of redistribution.

The object of this paper is to explore the principles that should guide the evaluation of tax-transfer systems that depend on both market income and on quantity of time worked. In order to illustrate the types of issues we want to address, let us start with an example of an individual who pays taxes or receives transfers from a government depending on his or her interaction with three different systems: an income tax system, a social assistance system (welfare) and an unemployment insurance system.¹ The example is inspired by the Canadian social system, however it has been purposely simplified to clarify issues and therefore the numerical values should be viewed as mainly illustrative.

Let y represent an individual's market income, let h represent the number of weeks (≤ 50) worked by an individual over a year and let T represent total taxes (net of transfers) paid by the individual over a year.

The income tax system:

If $y \leq \$6000$, there is no income tax; on income above $\$6000$, a marginal income tax of 20% is applied (i.e., total income tax equals $\text{Max} [.2(y-6000),0]$).

The social assistance system (welfare):

If $y \leq \$6000$, the social assistance payment is $\$6000 - y$; if $y > \$6000$, there is no social assistance payment.

The unemployment insurance system:

Letting h be the number of weeks worked, if $h \leq 10$, the individual is not eligible for unemployment insurance; if $10 < h \leq 30$, then the individual is eligible for $h - 10$ weeks of unemployment insurance payments at 60% of weekly wages, up to a maximum payment of $\$400$ per week; if $30 < h < 50$, the individual is eligible for $50 - h$ weeks of unemployment insurance payments at 60% of weekly wages, up to a maximum payment of $\$400$ per week.

Consider the net tax implication of these three systems combined. The net amount of taxes paid (or transfer received) depends both on an individual's wage rate and on the number of weeks worked. Hence the pattern of tax rates faced by individuals varies with different market wage rates. In particular, consider the case where individual 1 earns $\$600$ per week worked, and individual 2 earns $\$1000$ per week. Then the net taxes-transfers, T , paid by individuals 1 and 2 as a function of annual income are given below where, in calculating these tax rates, we assume that an individual receives unemployment insurance payments for any eligible non-working weeks:

Tax function of individual 1:

If $y \leq 6000$, $T = y - 6000$ (marginal rate of 100%);

If $6000 < y \leq 18000$, $T = -.4(y - 6000)$ (marginal rate of -40%);

If $18000 < y$, $T = -4800 + .8(y - 18000)$ (marginal rate of 80%);

¹For simplicity, we have not included in the example the interaction with the pension system. However, the issues we address are also potentially relevant for pension systems since these programs have pay-outs that depend both on income earned and on amount worked.

Tax function of individual 2:

If $y \leq 6000$, $T = y - 6000$ (marginal rate of 100%);
If $6000 < y \leq 10000$, $T = .2(y - 6000)$ (marginal rate of 20%);
If $10000 < y \leq 30000$, $T = 800 - .2(y - 10000)$ (marginal rate of -20%);
If $30000 < y$, $T = -3200 + .6(y - 30000)$ (marginal rate of 60%).

There are three aspects to notice about this tax-transfer system. First, the tax rate depends not only on income but also depends on a worker's revealed market type, that is his or her wage rate. In particular, note that marginal tax rates are different at different income levels depending on a worker's wage rate. Second, the individuals face high marginal tax rates at both high and low income levels. Third, the individuals face negative marginal tax rates for intermediate income segments. Let us emphasize that all these features stand in stark contrast to the prescriptions one would derive from a Mirrlees' type optimal tax problem. (There is a fourth property: the marginal tax rates are neither monotonically increasing nor decreasing; this does not contradict Mirrlees, but is of interest to us.) However, given that the above example allows tax rates to be wage dependent, we immediately know that Mirrlees' analysis does not directly apply and hence an alternative framework is needed.

In this paper, we examine an optimal income tax problem in hope of providing guidance on how to design such a system. For example, we would like to know how to best set a tax and transfer system when the government can design the system to depend both on income and wage rates (or the number of weeks worked). Moreover, since we believe that one of the concerns of governments is to avoid transferring substantial income to individuals that simply do not want to engage in market employment, our analysis recognizes that individuals may have different valuations for their non-market time.

Our approach to the problem follows the optimal non-linear income taxation literature as pioneered by Mirrlees (1971),² that is, we approach redistribution as a welfare maximization problem constrained by informational asymmetries. However, we depart in two directions from Mirrlees' formulation. The first concerns the perceived need to target more effectively income transfers. For example, traditional welfare programs (or minimum revenue guarantees) are often criticized on the grounds that they transfer substantial income to individuals who value highly their non-market time, as opposed to transferring income only to the most needy. Although such a preoccupation is common, the literature is mostly mute on how to address this issue since the standard framework assumes that individuals value their non-market time identically. The second issue relates to the possibility of using work time requirements as a means of targeting transfers. Many social programs – such as most unemployment insurance programs or pension programs – employ information on time worked (either in years, weeks or hours) in order to determine eligibility; therefore it seems reasonable to allow for such a possibility when considering how best to redistribute income. Hence, the environment we examine includes (1) taxation authorities which are uninformed about individuals' potential value of time in market activities and about their potential value of time in non-market activities,³ and (2) income transfers that can be contingent on both earned (market) income and on the allocation of time to market employment - and as a result also on the wage rate. Under the above assumptions, our redistribution problem formally becomes a multidimensional screening problem with two dimensions of unobserved characteristics.⁴

²See also Mirrlees (1997).

³In our formulation, non-market activities can be interpreted as non-declared market activities.

⁴Screening problems with two-dimensions of unobserved characteristics are becoming more common in the literature. See Armstrong (1996), Rochet and Choné (1998) for the state of the art in this literature and a discussion of

Given the two-dimensional informational asymmetry, it is not surprising that the properties of the optimal redistribution program derived under our informational and observability assumptions are quite distinct from those found in the standard setup. More specifically, we show that optimal redistribution in our environment entails

- A cutoff wage, where individuals with wage above the cutoff are taxed and individuals with wages below the cutoff are subsidized.
- For individuals below the cutoff wage, their employment level is distorted upwards as they face wage-contingent income subsidies (or earned income tax credits) that decrease as income increases.
- For individuals above the cutoff wage, their employment level is distorted downwards as they face positive and increasing marginal tax rates as they increase their income.
- Individuals that choose not to work receive no income transfer.

The above results provide a stark contrast with those of the standard non-linear taxation literature in large measure because in that literature the informational asymmetry is restricted to the value of market time. Since his seminal contribution, Mirrlees' analysis has been extended in several directions. Many of the extensions of Mirrlees' original analysis involve giving more tools to the taxation authorities. For example, see Guesnerie and Roberts (1987) or Marceau and Broadway (1994).⁵ In a different vein, Boone and Bovenberg (2004) extend the Mirrlees' model by introducing search costs and frictions. The model generates voluntarily unemployed individuals, involuntarily unemployed individuals and employed ones with heterogeneous levels of productivity. Search gives rise to bunching at the low end of the productivity distribution. One surprising aspect of much of the traditional optimal taxation literature is that it conflicts with current policy debates which, *de facto*, tend to favor active employment programs such as employment subsidies (negative marginal taxation). More recent works by Saez (2004), Choné & Laroque (2005) and Laroque (2005)⁶ show that negative marginal tax rates can be optimal when one focuses on the extensive margin, that is, when labor supply is a zero-one decision. The current paper adds to this literature by highlighting why negative marginal tax rates can be optimal in an environment where individuals can adjust on both the intensive and the extensive margin. In particular, our approach prescribes a negative marginal tax rate on the margin where individuals choose their hours of work; an individual with sufficiently low wages experiences an increase in net income in response to an increase in his hours worked that is greater than the associated increase in market income. This we believe captures the margin that is at the core of many policy discussions about negative marginal tax rates.⁷ In Choné and Laroque (2005) and Laroque (2005), negative marginal tax rates arise when an increase in the wage holding hours fixed leads to a decrease in taxes. However, for an individual, the wage is not a choice variable, so individuals cannot try to improve their situation by taking advantage of the negative marginal rate.⁸

The paper is structured as follows. In Section 2 we present the constrained redistribution problem, discuss the first best allocation and derive simple properties of the optimal direct revelation

some of the difficulties associated with solving such problems.

⁵See also Besley and Coate (1995), Brett (1998), Cuff (2000), and Kanbur, Keen and Tuomola (1994).

⁶See also Diamond (1980) who first addressed the extensive margin in a Mirrlees' type model.

⁷The environment analyzed by Saez is substantially different from the one considered here and therefore a direct comparison is difficult. In particular, Saez (2004) examines an environment with costly job choice and moral hazard with respect to which jobs to choose. The framework does not allow agents to vary their market time, nor does it allow for difference across agents in the value of their market time, two aspects which are central to our work.

⁸Concurrently with this paper, Choné & Laroque (2006) have specified a model which can study both the intensive margin and the extensive margin as limiting cases. As our analysis considers intensive and extensive margins, our work is closely related to this paper. There are nevertheless several differences between the two papers. For example, when discussing the intensive margin Choné & Laroque (2006) have effectively only one dimension of heterogeneity and do not consider the intensive and extensive margins in the same model as we do here.

mechanism and the implied tax patterns. In section 3 we discuss the case where individuals' market productivities are known but their non-market productivities are not. In section 4 we analyze the case where both the valuations of market and non-market time are unknown. Since it appears to be impossible to get explicit solutions for our problem in the absence of any restrictions on the distributions of the value of market time and non-market time, we restrict attention to the case where the probability density function associated with the value of non-market time does not decrease too rapidly. For our analysis in section 4, we assume that market and non-market valuations are either independently distributed or, when there is some correlation between the two random variables, that the strength of affiliation between these valuations is non-decreasing in the level of market productivity. Finally, in Section 5, we discuss how the optimal solution can be implemented by a simple social policy that depends on wage rates and market income, as was the case in our initial example. All proofs are relegated to the appendix.

2 The Environment

The economy has two sectors—a formal market sector and an informal, non-market or household sector. An individual can work in the formal/market sector at a wage rate no greater than his or her intrinsic productivity. Income earned in the formal sector can be observed and hence taxed. The amount of time allocated to the formal sector can also be observed. Since the wage rate earned in the formal sector can be deduced from market income and time spent working in the market, the wage rate earned can be treated as effectively observable. However, the individuals' intrinsic market productivity, that is the highest possible wage they can earn, is unobservable. Besides working in the formal sector, an individual can also allocate time to production in the informal/household sector. Production in this sector is unobservable.⁹ Each agent is endowed with a fixed number of hours which we have normalized to one; if an individual works for $h \geq 0$ hours in the formal sector, he or she has $1 - h$ hours available for producing goods in the informal sector. Individuals have identical utility functions that are known and which depend upon the consumption of goods from both sectors of the economy. Individuals differ in their abilities and the ability level can vary across sectors. For example, one may be very productive in the formal/market sector but have low productivity in the informal sector or conversely.

Before describing our problem further, it is worth discussing our assumption regarding the observability of time worked, which could represent hours, weeks or years. This is particularly relevant since the more common assumption in the literature is that hours worked are not observable¹⁰ and that only income is observable. In practice hours or weeks worked are used in many countries to determine eligibility for social programs. For example, in Canada, one of the biggest social programs is unemployment insurance. Eligibility and payments from the Canadian unemployment insurance system depend explicitly on income and the amount of time worked (both in terms of weeks and hours per week). This is a clear example of a large program that exploits information on time worked to determine transfers. Problems with measuring time worked do not appear to be very important.¹¹ As another example, currently, in Canada, there is a large scale experiment

⁹Alternatively, production in the household sector can be viewed as leisure with individuals having different tastes for leisure.

¹⁰Dasgupta and Hammond (1980) and Maderner and Rochet (1995) also examine optimal redistribution in environments where taxation authorities can transfer income based on market-income and market allocation of time. However, in these papers there is only one dimension of unobserved characteristics. See also Kesselman (1973) and Bloomquist (1981) for a related literature.

¹¹There are obviously some groups in society for which it is very difficult to measure the amount of time worked,

aimed at encouraging welfare recipients to work; this program is called the self-sufficiency project (see Card and Robins (1996) for details). One particular aspect of this program is that it explicitly requires individuals to work 30 hours per week in order to be eligible for a transfer; recipients are required to mail in pay stubs showing their hours of work and earnings for the month. Again, this illustrates that social programs currently use information on time-worked and therefore it seems relevant to allow for such a possibility in our analysis. Obviously, working time is not observable for everyone. Nonetheless, we believe that it is useful to examine the case where we assume it is observable, and later we discuss how our results would need to be modified if time worked is not observable for high wage individuals.

Types are indexed by $i, j \in I \times J$, where $I = \{1, \dots, n\}$ and $J = \{1, \dots, m\}$. For each type there is a two-tuple (ω_i, θ_j) ; $\omega_i \in \{\omega_1, \dots, \omega_n\}$ is the productivity of an individual with type i in the formal/market sector and $\theta_j \in \{\theta_1, \dots, \theta_m\}$ is the productivity of an individual with type j in the informal/household sector. Higher indices correspond to higher productivities. We assume that everybody is productive to some extent, that is θ_1 and ω_1 are both positive. p_{ij} denotes the joint probability that an agent's productivities take values ω_i and θ_j , respectively. For the time being we impose no restrictions on the probability distribution. Assumptions are introduced below when needed.

The discreteness of the type space is useful in the sense that it allows us to derive our results in a simple way. However, we also want to have a simple comparison to the case where types are distributed on a continuous rectangle. Therefore, we assume that the productivities are measured in the same units, so that (for $i, j > 1$) $\omega_i - \omega_{i-1} = \theta_j - \theta_{j-1} = 1$ (these measures are arbitrary, so there is no good reason to have these differences different from each other). ¹² Moreover, we normalize types such that $\omega_1 = \theta_1 = 1$. We denote by $j_{FB}(\omega)$ the largest j for individuals with market productivity ω such that $\theta_{j_{FB}}(\omega) \leq \omega$.

Individuals evaluate their well being according to the utility function

$$U(hw_i + (1 - h)\theta_j - T) \quad (1)$$

where $w_i \leq \omega_i$ is the wage rate earned in the market sector and T is the amount of taxes paid to or subsidies received, respectively, from the government. We assume that $U(\cdot)$ is differentiable and concave. Note that the argument of the utility function is the net-income of the individual thus assuming that the individual consumes two goods that are perfect substitutes. The first good is bought from net market income; $c_i = hw_i - T$ is the amount consumed of this good. In addition, the individual consumes $c_j = (1 - h)\theta_j$ units of the good he produces in the informal sector. However, since the two goods are perfect substitutes, it is actually more convenient to work with (1) directly.

The government's objective is to maximize a utilitarian social welfare function. But the government is unable to implement a first-best optimum due to the asymmetry of information. In particular, the government cannot observe skill levels of individuals in either sector, that is, the government cannot observe either ω_i or θ_j . By the revelation principle, we can restrict attention to

for example the self-employed. Accordingly, these groups are often excluded from programs such as unemployment insurance. Moreover, if transfers are made contingent on time worked, this may create an incentive for firms and workers to collude to exploit the redistribution system. Although this is a possibility that should be kept in mind, we abstract from it in the current analysis since it does not appear to be a widespread concern in the actual implementation of programs which do depend on work time information.

¹²These assumptions are appealing because our results can be extended to the continuous case, where we simply replace sums by integrals.

direct, incentive compatible mechanisms, where individuals are asked to announce a type (ω_i, θ_j) and the government chooses an allocation of work-time between the sectors, $h(\omega_i, \theta_j)$, a tax to be paid by the individual, $T(\omega_i, \theta_j)$, and a job allocation such that the individual is presumably able to do this job, that is $w_i \leq \omega_i$. It is immediate that the job allocation decision is trivial. At any solution to the government's problem, every individual must work in his most productive job. Otherwise a Pareto improvement can be created.¹³ So, the government's problem can be written as follows

$$\max_{\{h(\omega_i, \theta_j), T(\omega_i, \theta_j)\}} \left\{ \sum_i \sum_j p_{ij} U(\theta_j + h(\omega_i, \theta_j)(\omega_i - \theta_j) - T(\omega_i, \theta_j)) \right\} \text{ s.t., for all } (i, j) : \quad (2)$$

$$U(\omega_i h(\omega_i, \theta_j) - T(\omega_i, \theta_j) + (1 - h(\omega_i, \theta_j))\theta_j) \geq U(\omega_{\hat{i}} h(\omega_{\hat{i}}, \theta_j) - T(\omega_{\hat{i}}, \theta_j) + (1 - h(\omega_{\hat{i}}, \theta_j))\theta_j) \quad \forall \hat{j}, \forall \hat{i} \leq i \quad (3)$$

$$U(\omega_i h(\omega_i, \theta_j) - T(\omega_i, \theta_j) + (1 - h(\omega_i, \theta_j))\theta_j) \geq U(\theta_j) \quad (4)$$

$$\sum_i \sum_j p_{ij} T(\omega_i, \theta_j) = 0 \quad \text{and} \quad 0 \leq h(\omega_i, \theta_j) \leq 1. \quad (5)$$

In the above problem, (3) represents the incentive compatibility constraints, and (4) represents the participation constraints. Constraint (5) represents the materials balance constraint. Since the incentive compatibility constraints in this problem are not standard, some clarification is in order. An individual can costlessly mimic any other individual who has a lower market productivity; that is, individual (ω_i, θ_j) can choose to be employed in any job paying a wage $w \leq \omega_i$. In effect, the incentive compatibility constraint (3) ensures that individual (ω_i, θ_j) finds his or her allocation at least as good as that of any agent employed at a wage no greater than his or her own market productivity ω_i . The participation constraints, (4), reflect our assumption that the government cannot impose (or collect) a positive tax on an individual with no market income, that is, the fruits of non-market activity are not transferable to the government. Under this assumption, any individual can guarantee a minimum level of utility by simply not working.

Notice that the incentive and participation constraints depend exclusively on the level of income and that total income can be written as the sum of opportunity cost of time θ_j and the after tax excess income (over and above home productivity) from market participation. We define this measure as

$$V(\omega_i, \theta_j, \theta_j) \equiv h(\omega_i, \theta_j)(\omega_i - \theta_j) - T(\omega_i, \theta_j) \quad (6)$$

and let

$$\hat{v}(\omega_i, \theta_j) \equiv V(\omega_i, \theta_j, \theta_j). \quad (7)$$

Our problem is a problem of multi-dimensional screening, and thus potentially extremely complex to solve. However, (3) and (6) reveal a crucial difference to the general problem of multi-dimensional screening. The after-tax excess gain from working depends only on the message sent about market productivity, but not on the market productivity itself. The dependence on the market productivity is only implicit in the sense that to each ω_i there is an upper bound which is equal to ω_i . To help understand the constraints imposed by the informational asymmetries, we begin by characterizing the first best outcome when both ω and θ are assumed to be observable.

¹³See Beaudry and Blackorby (2004) for the formal proof of this statement.

2.1 First Best

In the first-best situation, the government is assumed to know the productivities of each individual both in the market and in non-market employment. The problem is to find the optimal redistribution of income among individuals under the constraint that the redistribution is feasible and that individuals are willing to participate. Formally, the government's problem is

$$\begin{aligned} \max_{\hat{v}(\cdot, \cdot), h(\cdot, \cdot)} & \left\{ \sum_{i=1}^n \sum_{j=1}^m p_{ij} U(\theta_j + \hat{v}(\omega_i, \theta_j)) \right\} \quad s.t. \\ & \sum_{i=1}^n \sum_{j=1}^m p_{ij} \hat{v}(\omega_i, \theta_j) = \sum_{i=1}^n \sum_{j=1}^m p_{ij} h(\omega_i, \theta_j) (\omega_i - \theta_j) \\ & \hat{v}(\omega_i, \theta_j) \geq 0 \quad \text{and} \quad 0 \leq h(\omega_i, \theta_j) \leq 1 \quad \text{for all } i, j. \end{aligned}$$

The problem is strictly concave in the choice variables. The optimal allocation of working times are

$$h^*(\omega_i, \theta_j) = \begin{cases} 1 & \text{if } \omega_i \geq \theta_j \\ 0 & \text{else} \end{cases}$$

The first-order condition for $\hat{v}(\omega_i, \theta_j)$ is

$$U'(\theta_j + \hat{v}(\omega_i, \theta_j)) - \lambda \leq 0; \hat{v}(\omega_i, \theta_j) \geq 0; (U'(\theta_j + \hat{v}(\omega_i, \theta_j)) - \lambda) \hat{v}(\omega_i, \theta_j) = 0$$

where λ is the multiplier on the budget constraint. Thus, either $\hat{v}^*(\omega_i, \theta_j) = 0$ or, when $\hat{v}^*(\omega_i, \theta_j) > 0$, then the individual's marginal utility is set equal to the marginal utility of everyone who receives a strictly positive net excess income. Hence, utility for these individuals must be equalized, that is $\theta_j + \hat{v}^*(\omega_i, \theta_j) = c$ for all (ω_i, θ_j) such that $\hat{v}^*(\omega_i, \theta_j) > 0$. It follows that the net excess incomes at the optimum depend only on the opportunity cost of time but not on market productivity. An individual receives a strictly positive net excess income of $c - \theta_j$ if $\theta_j < c$ and the individual receives a zero net income if $\theta_j \geq c$.

Using these definitions and the optimal allocation of working time we can restate the government's budget constraint as

$$\sum_{i=1}^n \sum_{j=1}^m p_{ij} \max\{c - \theta_j, 0\} = \sum_{i=1}^n \sum_{j=1}^m p_{ij} \max\{\omega_i - \theta_j, 0\}$$

This determines the optimal level of c^* .

The first best is associated with the following tax/transfer scheme (a positive number is a tax). (1) If $\omega_i < \theta_j < c^*$, then $T = -(c^* - \theta_j)$. (2) If $\omega_i \geq \theta_j$ and $c^* \geq \theta_j$, then $T = \omega_i - c^*$. (3) If $\omega_i < \theta_j$ and $c^* \leq \theta_j$, then $T = 0$. (4) If $\omega_i \geq \theta_j > c^*$, $T = \omega_i - \theta_j$.

If an individual faces the above transfer scheme, and could lie about both dimensions of his type, he would want to claim that he or she has a low value of non-market time, and an even lower value of market time. This way the individual would receive the biggest transfer and enjoy the fruits of his non-market activity. The first best problem indicates that some of the information constraints in our screening problem are binding at the optimum. Which information constraints would bind if an individual could only lie about his non-market type θ ? In this case, an individual with $\omega \leq c^*$ and $\omega < \theta$ would want to claim to have the lowest possible value of θ subject to $\theta > \omega$, so as to receive a large transfer. In contrast, an individual with $\omega > c^*$ and $\omega \geq \theta$ would want

to claim to have the highest possible value of θ subject to $\theta < \omega$, so as to pay the minimum in taxes. The direction in which the information constraints bind at the optimum changes depending on an individual's market type. This property contributes to making the solution to our screening problem non-standard.

2.2 Reducing the information constraints

In our problem, an individual's excess income depends only implicitly on his market productivity type, in the sense that he cannot exaggerate his productivity but can do a job of a less qualified person. However, when he mimics a less qualified person with the same productivity in the informal sector, he obtains exactly the same excess income as this person obtains. This insight allows us to prove the following result:

Proposition 1 *The two-dimensional incentive constraint (3) is satisfied if and only if this pair of one-dimensional constraints is satisfied for all (ω_i, θ_j)*

$$h(\omega_i, \theta_j)(\omega_i - \theta_j) - T(\omega_i, \theta_j) \geq h(\omega_i, \theta_j)(\omega_i - \theta_j) - T(\omega_i, \theta_j) \quad \forall i \leq i \quad (8)$$

and

$$h(\omega_i, \theta_j)(\omega_i - \theta_j) - T(\omega_i, \theta_j) \geq h(\omega_i, \theta_j)(\omega_i - \theta_j) - T(\omega_i, \theta_j) \quad \forall j. \quad (9)$$

Proposition 1 is an important simplification, since it reduces the number of relevant incentive constraints dramatically. However, the problem remains rather complex and it remains difficult to obtain any closed form solutions. We therefore will place restrictions on the distribution of types in order to obtain clear and simple characterizations of the optimal redistribution scheme. Before introducing these restrictions, we note the following property that does not depend on the distribution of types.

Proposition 2 *If $T(\omega_i, \theta_j)$ and $h(\omega_i, \theta_j)$ are optimal direct mechanisms, then for any three values of θ , say $\theta_a, \theta_b, \theta_c$ with $\theta_a < \theta_b < \theta_c$,*

$$(i) h(\omega_i, \theta_c) \leq h(\omega_i, \theta_b) \leq h(\omega_i, \theta_a)$$

and

$$(ii) T(\omega_i, \theta_b) \leq \lambda T(\omega_i, \theta_a) + (1 - \lambda)T(\omega_i, \theta_c) \text{ where } \lambda \text{ is defined by } h(\omega_i, \theta_b) = \lambda h(\omega_i, \theta_a) + (1 - \lambda)h(\omega_i, \theta_c)$$

Proposition 2 indicates that, conditional on one's market wage, net taxes increase in a convex fashion with hours worked, (or equivalently with income). To see that Proposition 2 is restrictive, one can go back to the example presented in the introduction and notice that it does not satisfy the condition. Hence, the current framework suggests that the tax system in the example would not be optimal. In particular, Proposition 2 implies that it is not optimal to have a tax-transfer system where a worker faces, as hours worked are increased, first a positive marginal tax rate followed by a lower or negative marginal tax. This proposition will later be used in conjunction with other propositions to provide a comprehensive description of the tax system that implements the optimal allocation. In order to provide a complete characterization of the optimal redistribution problem, we now introduce restrictions on the distribution of θ and ω . Let $p_j(i)$ denote the probability density function of θ conditional on $\omega = \omega_i$ and let $P_j(i) = \sum_{k=1}^j p_k(i)$ denote the associated distribution function.

Assumption 1: For any i , $\frac{P_j(i)}{p_j(i)}$ is concave in j .

Monotonicity of inverse hazard rates is a standard restriction in much of the screening literature. For our purposes, concavity of the inverse hazard rate is more useful (so we do not assume monotonicity¹⁴). This rules out distributions where the density decreases rapidly as j increases. A simple distribution that satisfies this assumption is the uniform distribution. This is an attractive feature given that the uniform represents a diffuse prior, which we want to permit since we know very little about the actual distribution of θ . While Assumption 1 allows us to obtain simple and explicit solutions using standard methods of proof, it can be shown¹⁵ that many of our qualitative results can be derived without this assumption at the cost of substantially more analytical complications and less transparency.

While Assumption 1 is enough to solve our problem for the case of observable market types, we need additional assumptions for the analysis of the case where both types are unobservable. We introduce these assumptions when needed in Section 4 below. We place no restriction on the marginal distribution of ω throughout the paper.

It is worth immediately highlighting that our problem is interesting only if there are many market productivities. If there were only one market productivity ω then, absent government intervention, individuals would work full time in the formal sector if and only if $\omega \geq \theta_j$. The resulting incomes would be equal to $\max\{\omega, \theta_j\}$.

In order to solve our problem with unobservable ω and θ , it is useful to approach it in two steps. First, we analyze the problem assuming that market productivities ω are observable by the government, but the non-market productivities are not. This is done in Section 3. Then in Section 4, we examine how the results of Section 3 are affected by assuming that the government cannot observe either market or non-market productivities.

3 Case with Observable Market Characteristics

In the case where market characteristics are observable, the problem can be broken down into a few steps by treating individuals with the same market productivity as a group. For any such group, we first examine how best to proceed if we want to extract a total revenue of T from the group, and then we examine how best to proceed if we want to transfer to a group a total subsidy of S . Once this is known, we examine the problem of which group to tax and which group to subsidize, and by how much.

3.1 The Problem of Optimally Collecting Taxes

In this subsection, we consider how best to collect a total tax revenue T from a group of individuals with market productivity ω_i . Since we are treating ω_i as observable, and θ_j as unobservable, this problem is almost, but not quite a dual of Mirrlees' original problem.¹⁶ Individuals with a low opportunity cost of time are the natural candidates for taxation, because they have most to gain from market participation. As a result, an individual has an incentive to claim a higher

¹⁴Since we do not impose monotonicity of the inverse hazard rate, assumption 1 is neither stronger nor weaker than monotonicity of the inverse hazard rate.

¹⁵See the working paper under Beaudry & Blackorby (2004).

¹⁶In contrast to Mirrlees' analysis, which assumes that only income is observable, we assume that both the time worked and income are observable. In Mirrlees' model, the first-best would be implementable under our assumptions. However, in addition, we also have the individual's valuation of non-market time unobservable. For a model with observable hours worked, but unobservable income and no further source of asymmetric information, see Maskin and Riley (1985).

opportunity cost of time in order to be taxed less, and the government has to ensure that no individual exaggerates his/her opportunity cost of time. In order to describe the set of incentive compatible allocations first define the optimal excess income by

$$v(\omega_i, \theta_j) \equiv \max_{\theta_{\hat{j}}} V(\omega_i, \theta_{\hat{j}}, \theta_j). \quad (10)$$

Proposition 3 *With observable market productivities, an allocation for a market productivity group that is taxed is implementable if and only if*

$$v(\omega_i, \theta_j) = v(\omega_i, \theta_m) + \sum_{k=j+1}^m h(\omega_i, \theta_k), \quad (11)$$

$$h(\omega_i, \theta_j) \geq h(\omega_i, \theta_{j+1}) \text{ for } j \leq m-1 \quad \text{and} \quad v(\omega_i, \theta_m) \geq 0. \quad (12)$$

It may be the case that there is some type $(\omega_i, \theta_{j'})$ such that all types with weakly higher opportunity costs of time, $\theta_j \geq \theta_{j'}$ do not work in the formal sector. The monotonicity constraint (12) ensures that the set of types that works in the formal sector takes this threshold form.

We can recover the taxes collected from the allocation of work time and the indirect excess incomes using the relation $v(\omega_i, \theta_j) = h(\omega_i, \theta_j)(\omega_i - \theta_j) - T(\omega_i, \theta_j)$ and (11). We obtain

$$T(\omega_i, \theta_j) = h(\omega_i, \theta_j)(\omega_i - \theta_j) - v(\omega_i, \theta_m) - \sum_{k=j+1}^m h(\omega_i, \theta_k) \quad (13)$$

Recalling that $p_j(i)$ denotes the conditional probability of $\theta = \theta_j$ given that $\omega = \omega_i$, we can write the government's problem as

$$\begin{aligned} W(\omega_i, T) \equiv \max_{\{h(\omega_i, \cdot)\}, v(\omega_i, \theta_m)} & \sum_{j=1}^m p_j(i) U \left(\theta_j + v(\omega_i, \theta_m) + \sum_{k=j+1}^m h(\omega_i, \theta_k) \right) \text{ s.t.} \\ & \sum_{j=1}^m p_j(i) \left(h(\omega_i, \theta_j)(\omega_i - \theta_j) - v(\omega_i, \theta_m) - \sum_{k=j+1}^m h(\omega_i, \theta_k) \right) = T \end{aligned}$$

and in addition (5) and (12).

The solution to this problem depends on the total amount of tax collected T . If the amount of tax to be levied is very small, in particular if $T \leq T^{\min} \equiv (\omega_i - \theta_{j_{FB}-1}(\omega_i)) \sum_{j=1}^{j_{FB}-1} p_j(i)$, then solution to the problem is simple since the tax can be levied without changing the allocation of hours worked prescribed by the first best. In this case, all those with $\theta < \omega_i$ work full time and are taxed by the amount $\frac{T}{\sum_{j=1}^{j_{FB}-1} p_j}$. In contrast, if $T > T^{\max} \equiv \max_j \sum_{j'=1}^j p_{j'}(i) (\omega_i - \theta_{j'})$ then, there is no solution to the problem since it is not feasible to collect such a high level of taxes. Hence, the interesting case is when $T^{\min} < T \leq T^{\max}$. The solution for this case is given in the following proposition. Remembering that $P_j(i) \equiv \sum_{k=1}^j p_k(i)$ we have

Proposition 4 *If the government wants to levy a tax revenue T from a group of individuals with market productivity ω_i , where $T^{\min} < T \leq T^{\max}$, then the optimal allocation of working time and*

taxes take the form

$$h^*(\omega_i, \theta_j) = \begin{cases} 1 & \text{for } \theta_j < \theta_{j_t} \\ h_{j_t} & \text{for } \theta_j = \theta_{j_t} \\ 0 & \text{for } \theta_j > \theta_{j_t} \end{cases}$$

and

$$T(\omega_i, \theta_j) = \begin{cases} \omega_i - \theta_{j_t-1} - h_{j_t} & \text{for } \theta_j < \theta_{j_t} \\ h_{j_t}(\omega_i - \theta_{j_t}) & \text{for } \theta_j = \theta_{j_t} \\ 0 & \text{for } \theta_j > \theta_{j_t} \end{cases}$$

where j_t is the largest j such that $\theta_j < \omega_i$ and the equation

$$T = P_{j_t-1}(i)[(\omega_i - \theta_{j_t-1}) - h_{j_t}] + p_{j_t}(i)[h_{j_t}(\omega_i - \theta_{j_t})] \quad (14)$$

has a solution for some $h_{j_t} \in [0, 1]$. ¹⁷

This proposition indicates that there is a marginal type, j_t , who works generically less than full time in the formal sector and separates the remainder of his fellow types with the same market productivity into two classes. Those who have smaller opportunity costs than the marginal type work full time in the formal sector, those with higher opportunity cost of time stay in the informal sector. Note that in general j_t and h_{j_t} depend on ω_i and T . Here we subsume this dependence to simplify notation, but it will appear later when needed.

The intuition for the resulting allocation is rather simple. The government wishes to spread the burden of taxation among as many as possible people. Therefore j_t is as large as possible. Moreover, it wishes to equalize the incomes of those who work for it cannot equalize the incomes of those who fare better on their own, since participation in market activities is voluntary. By incentive compatibility, the time spent by all inframarginal types in the formal sector determines the income of their right-wards neighbor. If an agent works full time then the income is equalized between him and his left-wards neighbor. The time the marginal type spends in the formal sector does not influence the income of his right-wards neighbor, since this type consumes his endowment. However, the time the marginal agent spends in the formal sector plays an important role. By a chain of right-wards looking incentive constraints it determines the rents left to all the inframarginal types. The more hours the marginal type works the higher is the rent left to each inframarginal type and the smaller is the total tax collected. Therefore h_{j_t} adjusts to the level that is consistent with collecting the tax T .¹⁸ Using the insights that the incomes of all inframarginal types are equalized, the participation constraint of the marginal type is binding, and that the first-order condition holds for the hours worked by the marginal type; we can express the solution by the simple condition $U'(\theta_{j_t-1} + h_{j_t}) = \lambda_i \left(1 - \frac{p_{j_t}(i)}{P_{j_t-1}(i)} (\omega_i - \theta_{j_t})\right)$ where λ_i is the multiplier on the budget constraint of group i . The simplicity of this solution is due to Assumption 1; the effect on the resource constraint of an increase in hours worked by any type is monotonic in the opportunity costs. Note that when $T^{\min} < T \leq T^{\max}$, work allocations are distorted downwards relative to the first best. In particular, all types j with $j_t < j \leq j_{FB}$ no longer work in the formal sector as a result of taxes while they would have worked full time in the absence of taxes. For the individuals

¹⁷There is a straightforward algorithm to solve explicitly for j_t and h_{j_t} from (14) as a function of T and ω_i . This algorithm can be obtained from the following web page (Currently attached as a supplement to this document.).

¹⁸As T increases, h_{j_t} decreases until it is equal to zero. At this point we could call type j_{t-1} the marginal type and set $h_{j_{t-1}}$ equal to one without changing anything. If $T = T^{\max}$, then, decreasing the hours worked by j_{t-1} lowers total tax revenue.

with $j = j_t$, they also work less due to the taxes but nevertheless work a positive number of hours.

3.2 The Problem of Optimally Subsidizing

The incentive problems for a group that receives subsidies are diametrically opposed to the incentive problems of those who are taxed. Heuristically, the left-wards incentive constraints are more important than the right-wards constraints, because the government must remove the incentive to claim that the opportunity cost of time—the productivity in the informal sector—is lower than it in fact is. The government wishes to help those with a lower opportunity cost of time but it does not want to give its money to those who fare well on their own. In what follows we use the term “exclusion” as a synonym for not paying subsidies to an agent, and “inclusion” in the opposite sense.

The government’s problem can be understood as a combined problem of exclusion and redistribution. The government wishes to subsidize all agents with an opportunity cost of time less than or equal to θ_{j_s} , where j_s is a variable of its choice. Equivalently, the government wishes to exclude from subsidization all agents with opportunity cost of time higher or equal to θ_{j_s+1} . For any given j_s , the government’s problem is to distribute the available income to the agents that are included in the redistribution program. Henceforth, we call type (ω_i, θ_{j_s}) the marginal type. We begin again with a derivation of an alternative formulation of the incentive and participation constraints.

Proposition 5 *With observable market productivities, an allocation that includes all types with opportunity cost of time less or equal to θ_{j_s} for some j_s and excludes all other types, is implementable if and only if*

$$v(\omega_i, \theta_j) = \begin{cases} v(\omega_i, \theta_1) - \sum_{k=1}^{j-1} h(\omega_i, \theta_k) & \text{for } j \leq j_s \\ 0 & \text{otherwise,} \end{cases}$$

$$h(\omega_i, \theta_j) \geq h(\omega_i, \theta_{j+1}) \text{ for all } j < m-1,$$

$$v(\omega_i, \theta_1) - \sum_{k=1}^{j_s-1} h(\omega_i, \theta_k) \geq 0 \quad \text{and} \quad v(\omega_i, \theta_1) - \sum_{k=1}^{j_s} h(\omega_i, \theta_k) \leq 0$$

If all types up to type (ω_i, θ_{j_s}) are included, the excess incomes of these types satisfy $v(\omega_i, \theta_1) - \sum_{k=1}^{j-1} h(\omega_i, \theta_k) \geq 0$ for all $j \leq j_s$. Thus, the utilities of types with opportunity cost of time up to θ_{j_s} are linked by the left-wards adjacent incentive constraints. Types with opportunity cost of time of θ_{j_s+1} and higher get zero excess incomes. To make sure that all these types are excluded, the allocation must be such that the right-wards neighbor of the marginal type would obtain a non-positive excess income if he mimicked the marginal type. The allocation must satisfy the exclusion constraint $v(\omega_i, \theta_1) - \sum_{k=1}^{j_s} h(\omega_i, \theta_k) \leq 0$ ¹⁹.

¹⁹The exclusion and inclusion constraints make our model different from standard problems. For a general analysis of participation constraints in adverse selection models, see Jullien (2000).

We can write the government's problem as

$$\begin{aligned}\tilde{W}(\omega_i, S) \equiv \max_{\{h(\omega_i, \cdot)\}, v(\omega_i, \theta_1), j_s} & \sum_{j=1}^{j_s} p_j(i) U \left(\theta_j + v(\omega_i, \theta_1) - \sum_{k=1}^{j-1} h(\omega_i, \theta_k) \right) + \sum_{j=j_s+1}^m p_j(i) U(\theta_j) \text{ s.t.} \\ & \sum_{j=1}^{j_s} p_j(i) \left(h(\omega_i, \theta_j) (\omega_i - \theta_j) - \left(v(\omega_i, \theta_1) - \sum_{k=1}^{j-1} h(\omega_i, \theta_k) \right) \right) + S \geq 0, \\ & v(\omega_i, \theta_1) - \sum_{k=1}^{j_s-1} h(\omega_i, \theta_k) \geq 0 \quad \text{and} \quad v(\omega_i, \theta_1) - \sum_{k=1}^{j_s} h(\omega_i, \theta_k) \leq 0.\end{aligned}$$

The solution to the subsidization problem depends on the level of the total subsidy, with three possible configurations. A first case arises when subsidization is small ($S \leq S^{\min} \equiv P_{j_{FB}}(\theta_{j_{FB}+1} - \omega_i)$) and the first-best work allocation remains optimal; a second case arise when the total subsidy is sufficiently high so that the first-best allocation of work is no longer optimal but not everyone gets a subsidy; and the third case arises when the total subsidy S is so large that everyone gets a subsidy. This last case arises when $S \geq S^{\max} \equiv P_{m-1}(\theta_m - \omega_i)$. In the case where $S \leq S^{\min}$, the solution is simple as all types with $\theta_j < \omega_i$ work full time and receive a subsidy equal to $\frac{S}{P_{j_{FB}}}$.

Proposition 6 *If the government wants to distribute a total subsidy S to a group of individuals with market productivity ω_i , where $S^{\min} < S \leq S^{\max}$, then the optimal allocation of working time and subsidies take the form*

$$h^*(\omega_i, \theta_j) = \begin{cases} 1 & \text{for } \theta_j < \theta_{j_s} \\ h_{j_s} & \text{for } \theta_j = \theta_{j_s} \\ 0 & \text{for } \theta_j > \theta_{j_s} \end{cases}$$

and

$$-T(\omega_i, \theta_j) = \begin{cases} \theta_{j_s} - \omega_i + h_{j_s} & \text{for } \theta_j < \theta_{j_s} \\ h_{j_s} (\theta_{j_s+1} - \omega_i) & \text{for } \theta_j = \theta_{j_s} \\ 0 & \text{for } \theta_j > \theta_{j_s} \end{cases}$$

where j_s is the highest j such that the equation

$$P_{j_s-1}(i) [\theta_{j_s} - \omega_i + h_{j_s}] + p_{j_s}(i) h_{j_s} [\theta_{j_s+1} - \omega_i] = S \quad (15)$$

has a solution for some $h_{j_s} \in [0, 1]$. ²⁰

As we have explained above, the binding incentive constraints are “left-ward looking”. Moreover, the exclusion constraint must be binding. If this constraint were slack, then there would be no reason to put the marginal type to work in the formal sector, since $S > 0$ implies that the marginal type's productivity in the formal sector must be lower than his opportunity cost. But then the allocation cannot be incentive compatible, because all types with opportunity costs higher than the marginal type, who should be excluded from receiving subsidies, can claim the subsidy targeted at the marginal type without cost.

Again Assumption 1 implies that the trade-off between an individual's contribution to the resources available for redistribution and the rents left to inframarginal types changes monotonically

²⁰There is a straightforward algorithm to solve explicitly for j_s and h_{j_s} from (15) as a function of S and ω_i . This algorithm can be obtained from the following URL (Currently a supplement to this document.).

as we increase opportunity costs. Therefore, there is again a single marginal type, who divides his fellow types into two groups. Those who have lower opportunity costs work full time and those with higher opportunity costs of time do not work at all. All types who are subsidized receive the same amount of total income, so their marginal utilities are equalized. Using these insights, we can show that the optimal allocation is fully characterized by the simple first-order condition $U'(\theta_{j_s} + h_{j_s}) = \lambda_i \left(1 - \frac{p_{j_s}(i)}{P_{j_s}(i)} (\omega_i - \theta_{j_s})\right)$ where λ_i is the multiplier on the group's resource constraint. Work allocations are distorted upwards relative to first best allocations. In particular, all types j with $j_{FB} < j \leq j_s$ work more in the presence of subsidies than in the first best allocation. Individuals with $j_{FB} < j < j_s$ actually work full time in the presence of subsidies, while they would not work at all in their absence.

3.3 Optimal Redistribution with 2 Observable Market Types

We now consider the case where there are two market types: ω_1 and ω_n . Let p_1 denote the probability that $\omega = \omega_1$ and let p_n denote the probability that $\omega = \omega_n$. The government chooses T , the amount of taxes collected from one group, and S the amount to give to the other group, in order to maximize the expected utility of the entire population. In this case, it is obvious that it is optimal to levy a tax on the group with the higher market productivity, and subsidize the group with $\omega = \omega_1$. The budget constraint links T and S through the condition $p_n T = p_1 S$. We can write the government's problem as

$$\max_{T,S} \left\{ p_n W(\omega_n, T) + p_1 \tilde{W}(\omega_1, S) \right\} \quad \text{s.t.} \quad p_n T = p_1 S \quad \text{and} \quad T \leq T^{\max}.$$

The optimal levels of T and S , denoted T^* and S^* , satisfy either

$$-\frac{\partial W(\omega_n, T^*)}{\partial T} = \frac{\partial \tilde{W}(\omega_1, S^*)}{\partial S} \quad (16)$$

or $T^* = T^{\max}$, $S^* = \frac{p_n}{p_1} T^{\max}$ and

$$-\frac{\partial W(\omega_n, T^{\max})}{\partial T} < \frac{\partial \tilde{W}(\omega_1, \frac{p_n}{p_1} T^{\max})}{\partial S}$$

Recall from the previous propositions that the marginal type j_t and the hours this type works, h_{j_t} , are determined by the amount of total tax T to be collected from this group. Similarly the marginally subsidized individual j_s among types with market productivity ω_1 , and the hours he works h_{j_s} , are determined by the the amount of total subsidy S . Hence, it is useful to express condition (16) in terms of these variables. From the envelope conditions we obtain

Proposition 7 *The solution of the government's redistribution problem when there are two observable market productivities must satisfy either*

$$\frac{U'(\theta_{j_t-1} + h_{j_t})}{U'(\theta_{j_s} + h_{j_s})} = \frac{1 - \frac{p_{j_t}(n)}{P_{j_t-1}(n)} (\omega_n - \theta_{j_t})}{1 - \frac{p_{j_s}(1)}{P_{j_s}(1)} (\omega_1 - \theta_{j_s})}$$

where j_t, j_s, h_{j_t} and h_{j_s} satisfy the conditions of Propositions 4 and 6 or $T = \max_j \sum_{j'=1}^j p_{j'} (n) (\omega_n - \theta_{j'})$ and

$$\frac{U'(\theta_{j_t-1} + h_{j_t})}{U'(\theta_{j_s} + h_{j_s})} < \frac{1 - \frac{p_{j_t}(n)}{P_{j_t-1}(n)}(\omega_n - \theta_{j_t})}{1 - \frac{p_{j_s}(1)}{P_{j_s}(1)}(\omega_1 - \theta_{j_s})}$$

To understand the content of this proposition, it is helpful to recognize that the four objects j_t , j_s , h_{j_t} and h_{j_s} can all be thought as functions of the total tax to be levied on the high market productivity group. Hence, this proposition implicitly defines the optimal level T to levy on these types, and the government budget constraint indicates how much to subsidize in total the low market productivity group. Then, given the optimal levels of total taxes and total subsidies, Propositions 4 and 6 indicate the associated individual levels of taxes and the individual levels of subsidies that support the optimal allocation. In this sense, Proposition 7, in conjunction with Propositions 4 and 6, offer a complete description of the optimal redistribution problem with two observable market types.²¹

3.4 Many Observable Market Characteristics

We now generalize our findings to the case where there are many observed market characteristics. This problem can be reduced to finding a sequence of total taxes, $T_i \quad i = \{1, \dots, n\}$, where an element T_i represents the total tax levied on the group of individuals with market productivity ω_i , and a negative value of T_i represents a subsidy. For any two groups for which the maximal tax capacity is not attained, it must be that the marginal cost of taxation is equalized. In particular, if groups i and i' are subsidized (negative value of T_i), it must be the case that

$$\frac{U'(\theta_{j_{s_i}(\omega_i, T_i)-1} + h_{j_{s_i}}(\omega_i, T_i))}{\left(1 - \frac{p_{j_{s_i}(\omega_i, T_i)}(i)}{P_{j_{s_i}(\omega_i, T_i)-1}(i)}(\omega_i - \theta_{j_{s_i}(\omega_i, T_i)})\right)} = \frac{U'(\theta_{j_{s_{i'}}(\omega_{i'}, T_{i'})-1} + h_{j_{s_{i'}}}(\omega_{i'}, T_{i'}))}{\left(1 - \frac{p_{j_{s_{i'}}(\omega_{i'}, T_{i'})}(i')}{P_{j_{s_{i'}}(\omega_{i'}, T_{i'})-1}(i')}(\omega_{i'} - \theta_{j_{s_{i'}}(\omega_{i'}, T_{i'})})\right)}. \quad (17)$$

In (17), we make explicit the dependence of j_s and h_{j_s} on ω and the total tax paid by a group. Similarly, if group i has a positive value of total taxes, $T_i > 0$, and group i' is subsidized, it must be the case that either

$$\frac{U'(\theta_{j_{t_i}(\omega_i, T_i)-1} + h_{j_{t_i}}(\omega_i, T_i))}{\left(1 - \frac{p_{j_{t_i}(\omega_i, T_i)}(i)}{P_{j_{t_i}(\omega_i, T_i)-1}(i)}(\omega_i - \theta_{j_{t_i}(\omega_i, T_i)})\right)} = \frac{U'(\theta_{j_{s_{i'}}(\omega_{i'}, T_{i'})-1} + h_{j_{s_{i'}}}(\omega_{i'}, T_{i'}))}{\left(1 - \frac{p_{j_{s_{i'}}(\omega_{i'}, T_{i'})}(i')}{P_{j_{s_{i'}}(\omega_{i'}, T_{i'})-1}(i')}(\omega_{i'} - \theta_{j_{s_{i'}}(\omega_{i'}, T_{i'})})\right)} \quad (18)$$

or $T_i = T^{\max}(\omega_i) = \max_j \sum_{j'=1}^j p_{j'}(i)(\omega_i - \theta_{j'})$ and

$$\frac{U'(\theta_{j_{t_i}(\omega_i, T_i)-1} + h_{j_{t_i}}(\omega_i, T_i))}{\left(1 - \frac{p_{j_{t_i}(\omega_i, T_i)}(i)}{P_{j_{t_i}(\omega_i, T_i)-1}(i)}(\omega_i - \theta_{j_{t_i}(\omega_i, T_i)})\right)} < \frac{U'(\theta_{j_{s_{i'}}(\omega_{i'}, T_{i'})-1} + h_{j_{s_{i'}}}(\omega_{i'}, T_{i'}))}{\left(1 - \frac{p_{j_{s_{i'}}(\omega_{i'}, T_{i'})}(i')}{P_{j_{s_{i'}}(\omega_{i'}, T_{i'})-1}(i')}(\omega_{i'} - \theta_{j_{s_{i'}}(\omega_{i'}, T_{i'})})\right)} \quad (19)$$

A similar condition applies if two groups are taxed. These conditions, in addition to the government's budget constraint, $\sum_{i=1}^n p_i T_i = 0$, determine the optimal level of total taxes and subsidies

²¹In general our solution is second-best optimal, but it is easy to verify that in the special case where $\omega_n = \omega_1 + 1$, the first best allocation is actually implemented by this solution.

for each group, and then Propositions 4 and 6 can be used to determine the taxes that implement the optimal allocations. One interesting question is who gets taxed and who gets subsidized. The following proposition answers this question.

Proposition 8 *There is a critical market productivity $\bar{\omega}$ such that individuals are taxed only if their market productivity satisfies $\omega_i \geq \bar{\omega}$. Individuals are subsidized only if their market productivity satisfies $\omega_i < \bar{\omega}$.*

This proposition indicates that an optimal redistribution plan has the property that individuals are taxed or subsidized depending on whether their market productivity falls short of or exceeds a critical value. In other words, the most important determinant of whether an individual should be taxed or subsidized is not their market income but instead it is their market wage rate.

4 Case with Unobservable Market and Non-Market Productivities

Throughout the previous section we assumed that market productivities were observable by the government. In this section, we relax this assumption and consider the main case of interest where individuals can claim to have market productivities lower than that given by their innate ability. In the case of the first best allocation, we saw that this constraint could be relevant since in the first best allocation individuals had an incentive to lie about both their market and non-market productivity. Accordingly, it is important to ask how the addition of this informational problem affects the optimal redistribution program relative to the case where only non-market productivities are unobservable. Again we answer this question for distributions that satisfy certain regularity conditions. In particular, we impose

Assumption 2: For any j , $\frac{P_j(i)}{p_j(i)}$ is concave in i .

Note that this condition is similar to the condition in the one-dimensional case, but here we require the property to hold for the conditioning variable i . A simple and natural case that satisfies Assumption 2 is when market and non-market productivities are independent of each other. However, Assumption 2 holds more generally when—heuristically—market and non-market productivities are more strongly correlated the higher the level of market productivity. More precisely, we show in the appendix that the inverse hazard rate is concave in i if the strength of affiliation between market and non-market productivity is non-decreasing in the level of market productivity. Under these conditions, the answer to our question turns out to be very simple, as expressed in the following proposition.

Proposition 9 *For distributions that satisfy Assumptions 1 and 2, in particular, when productivities are independent, the optimal allocation with observable market characteristics remains incentive compatible when market characteristics are not observable.*

This proposition indicates that our characterization of the optimal redistribution program derived in Section 3 carries over to our main case of interest where innate market productivities are not observable. The intuition for this result is very simple. To capture this intuition, it is useful to start with the case where productivities are independent of each other, so that the *only* difference between two individuals with non-market productivity θ_j and market productivity ω_i and ω_{i+1} , respectively, is precisely that the marginal increase in resources available for redistribution is larger

when the latter works an instant longer. Because the trade-off between the increase in marginal utilities and the reduction in taxes collected from inframarginal individuals is the same in both groups, all individuals with market productivity ω_{i+1} will work at least as much in the formal sector as those do with market productivity ω_i . By incentive compatibility, the allocation of working times determines the net excess incomes of these individuals. Hence, if individuals with higher market productivity spend weakly more time in the formal sector, the resulting total incomes must be weakly higher in the group with the higher market productivity. But then, there is no incentive to mimic another individual with the same opportunity cost of time but a lower market productivity. By Proposition 1, this implies that there is no individual at all in any group with a lower market productivity that can be profitably mimicked. So, the optimal allocation with observed market productivities remains incentive compatible when market productivity is not observed.

The key to understanding the more general case of affiliated productivities is again the impact of a change in market productivity on the trade-off between the contribution to the resources available for redistribution when the working time of the marginal type is increased and the rents that have to be left to all inframarginal types. In addition to the difference in the productivity differentials $\omega_i - \theta_j$ and $\omega_{i+1} - \theta_j$, respectively, there is now also a difference in the ratios of types with opportunity costs of time equal to θ_j and types which have smaller opportunity costs of time. When the strength of affiliation between the productivities is increased as market productivity is increased, the conditional distributions of opportunity costs have more and more mass towards the higher levels of opportunity costs. For high productivity groups this rules out the government placing huge taxes on high productivity individuals. Since many high market productivity individuals have higher opportunity costs, these groups cannot be taxed heavily, because their members would simply stop working. Similarly, for groups that are subsidized, lower productivity groups have stochastically more mass towards the lower levels of opportunity costs than higher productivity groups have. This rules out the government subsidizing heavily a group of very low market productivity individuals who have very high opportunity costs.

Thus, we rule out cases where most of the low market productivity individuals have high opportunity costs of time and most of the high market productivity individuals have low opportunity costs of time, because this would allow the government to tax heavily the high wage earners and would give the government an incentive to subsidize heavily the low wage earners. In such a case, the vertical constraints might become binding. However, given our assumptions, it actually does not matter at all whether the government can or cannot observe the maximum market productivity of the individuals; the optimal allocation is in both cases the same.

5 The Structure of Income Taxes and Subsidies

Our analysis of the informationally constrained redistribution problem has allowed us to derive properties of an optimal tax system in the form of a direct revelation mechanism. However, in practice, tax systems do not take this form. Instead, tax systems are more akin to indirect revelation mechanism. For example, in the introduction, we discussed a simplified tax system that depended on one's income and one's wage rate. We can call such a system a wage contingent income tax system, and denote such a system by the function $\hat{T}(w, y)$ where $y = wh$ is income. In this section, we describe the properties of the wage contingent income tax system that implements

the solution to our optimal redistribution problem.²² In this case, we are assuming the government can observe both workers' incomes, and their wage rate. Obviously, this is equivalent to assuming the government can observe income and hours worked.

The first notable property, which follows directly from Proposition 8, is that there exists a critical wage \bar{w} , such that $\hat{T}(w, y) \geq 0$ for all y if $w > \bar{w}$, and $\hat{T}(w, y) \leq 0$ for all y if $w \leq \bar{w}$. This observation emphasizes that being taxed versus subsidized depends first and foremost on one's wage, not on income. The second striking property is that $\hat{T}(w, y) = 0$ if $y = 0$, that is, individuals that choose not to work do not get any subsidies. This result, which is implied by the nature of the direct tax functions derived in Propositions 4 and 6, implies the absence of welfare payments for employable individuals. This is in stark contrast to the traditional optimal tax literature which generally prescribes positive welfare payments to individuals who do not choose to work. In our setup, it is always better to use wage contingent employment subsidies to redistribute income since this allows the government to target workers with poor options both within and outside the market.

The third property, which follows from Proposition 2, relates to the nature of marginal tax rates and marginal subsidies. In particular, Proposition 2 implies that a wage contingent income tax system implements the optimal allocation only if it is convex in income over all levels of income that are achieved by some type in equilibrium. This indicates that an optimal wage-contingent income tax system has the property that as an individual increases his income (by increasing his hours worked), he faces either weakly increasing marginal tax rates if his wage rate is high, or alternatively faces weakly decreasing marginal income subsidies if his wage rate is low. In other words, negative marginal tax rates are weakly increasing as an individual increases his income. As an example, the following piece-wise linear tax schedule could be used to implement the optimal allocation.

For an individual being paid a wage above the critical level $w > \bar{w}$, and earning income y , then taxes are given by

$$\hat{T}(w, y) = \begin{cases} \left(1 - \frac{\theta_{j_t}}{w}\right) y & \text{for } y \leq \bar{y} \equiv h_{j_t} w \\ \left(1 - \frac{\theta_{j_t}}{w}\right) \bar{y} + (y - \bar{y}) \left(1 - \frac{\theta_{j_t-1}}{w}\right) & \text{for } y > \bar{y}. \end{cases}$$

For an individual begin paid a wage below the critical level $w \leq \bar{w}$, then subsidies are given by

$$-\hat{T}(w, y) = \begin{cases} \left(\frac{\theta_{j_s+1}}{w} - 1\right) y & \text{for } y \leq \underline{y} \equiv h_{j_s} w \\ \left(\frac{\theta_{j_s+1}}{w} - 1\right) \underline{y} + (y - \underline{y}) \left(\frac{\theta_{j_s}}{w} - 1\right) & \text{for } y > \underline{y} \end{cases}$$

In the above, the indices j_t, j_s and the work hours h_{j_t}, h_{j_s} are a function of the wage rate w and are determined as in Propositions 4 and 6, in conjunction with the conditions presented in Section 3.4. As can be seen, this tax schedule has the property that marginal taxes are weakly increasing for individuals with $w > \bar{w}$ since $\left(1 - \frac{\theta_{j_t}}{w}\right) < \left(1 - \frac{\theta_{j_t-1}}{w}\right)$. Similarly, marginal subsidies are decreasing for low wage individuals since $\left(\frac{\theta_{j_s+1}}{w} - 1\right) > \left(\frac{\theta_{j_s}}{w} - 1\right)$.

In summary, our analysis implies that a wage-contingent tax system has the following four properties: (1) the existence of a cutoff wage, where individuals with wages above the cutoff are taxed and individuals with wages below the cutoff are subsidized, (2) individuals below the cutoff wage face wage-contingent marginal income subsidies that decrease as income increases, (3) individuals above the cutoff wage face positive and increasing marginal tax rates as income

²²Since the function $h(\omega, \theta)$ which prescribes the optimal allocations is monotonic in θ (for a given ω), it is easy to verify that a wage-contingent income tax schedule can be used to implement the optimal redistribution problem.

increases, and (4) individuals that choose not to work receive no income transfer.

Throughout this analysis, we have been assuming that the government can observe both a worker's income and his wage rate (or hours worked). For many individuals this appears to be a reasonable assumption since many social programs in industrialized countries are based on such information and these programs appear to function properly. However, for some individuals, especially many high market productivity individuals, this assumption is unlikely to hold in practice. It is therefore relevant to ask how our results would need to be modified if governments could not observe hours worked for individuals paid at high wage rates. Without providing a full analysis here, such a modification would not change the flavor of our main results if the unobservability of hours or wages arose (mainly) for individuals with market productivity above the critical level associated with subsidization. In this case, the government could run a standard income tax system (based only on income) plus a separate earned-income subsidy system where individuals would need to have verifiable income and hours (or wages) statements to be eligible for a subsidy. While the tax system would be less efficient in the absence of information on hours worked, the subsidy system could still avoid transferring income to individuals with high value of time outside the market by requiring them to prove that they are working at low paying jobs. What is crucial for most of our results is the observability of hours (or wages) for potentially subsidized jobs; the observability of hours for high paying jobs is less critical.

6 Conclusion

The object of this paper is to explore the principles that govern the design of an optimal redistribution program in which taxation authorities have both reasons and tools to favor programs that target transfers more effectively than simple negative income tax schemes.²³ To this end we have analyzed a variant of the optimal taxation problem pioneered by Mirrlees. Our departure consists of allowing for a greater scope of unobserved heterogeneity in the population and allowing the government to transfer income based on both market income and market labor supply. Our main finding is that, in contrast to much of the optimal taxation literature, optimal redistribution in this environment is achieved using employment subsidies on low market performers, positive marginal tax rates on high market performers, and no transfers to non-working individuals .

How should these results be interpreted? In our view, these results are not a call for redesigning income tax systems to include a dependence on wages. Instead we view these results as supporting the potential relevance of certain active labor market programs as a complement to income tax as a means of redistributing income.²⁴ For example, these results provide potential support for programs, such as the US Earned Income Tax Credit and Canadian Self-Sufficiency Project, which supplement the income of low wage earners who choose to work. More generally, we view our results as suggesting the use of phased-out wage subsidies as a means of redistributing income to low earners, that is, wage subsidies that decrease in intensity as an individual chooses to supply more labor. Such phased-out subsidy programs, in effect, allow substantial transfers to the most needy in society without inciting either high market-value individuals or high non-market value individuals to take advantage of it.

²³Avenues of future research include examining the value of rendering some informal activities observable through monitoring, and rendering the acquisition of skill endogenous.

²⁴We also view these results as providing minimal guidelines of how such programs should interact with the income tax system in terms of the implied pattern of effective marginal tax rates.

7 Appendix

Proof of Proposition 1. The only if part is trivial. So, consider the sufficiency part. Suppose (8) and (9) are satisfied and consider type (ω_i, θ_j) which mimics type $(\omega_i, \theta_{\hat{i}})$ for $\hat{i} \leq i$ and arbitrary \hat{j} . The excess income he obtains this way is exactly the excess income that type (ω_i, θ_j) obtains from mimicking type $(\omega_i, \theta_{\hat{j}})$. But by (9) applied to type (ω_i, θ_j) , this implies that incentive compatibility is satisfied for an arbitrary \hat{j} and $\hat{i} \leq i$. ■

Proof of Proposition 2. The incentive compatibility constraints between types a, b and c directly imply that $h(\omega_i, \theta_c) \leq h(\omega_i, \theta_b) \leq h(\omega_i, \theta_a)$, which implies that for any $0 < \lambda < 1$,

$$h(\omega, \theta_b)(\omega - \theta_b) - T(\omega, \theta_b) \geq \lambda(h(\omega, \theta_a)(\omega - \theta_b) - T(\omega, \theta_a)) + (1 - \lambda)(h(\omega, \theta_c)(\omega - \theta_b) - T(\omega, \theta_c)).$$

If λ is further chosen such that $h(\omega_i, \theta_b) = \lambda h(\omega_i, \theta_a) + (1 - \lambda)h(\omega_i, \theta_c)$, this implies that $T(\omega_i, \theta_b) \leq \lambda T(\omega_i, \theta_a) + (1 - \lambda)T(\omega_i, \theta_c)$. ■

Proof of Proposition 3. We begin showing that the monotonicity condition (12) is necessary for incentive compatibility. Consider type (ω_i, θ_j) and apply (9) :

$$h(\omega_i, \theta_j)(\omega_i - \theta_j) - T(\omega_i, \theta_j) \geq h(\omega_i, \theta_{\hat{j}})(\omega_i - \theta_j) - T(\omega_i, \theta_{\hat{j}}).$$

Now, consider type $(\omega_i, \theta_{\hat{j}})$ (interchanging type and message) and a deviation to (ω_i, θ_j) .

$$h(\omega_i, \theta_j)(\omega_i - \theta_j) - T(\omega_i, \theta_j) \geq h(\omega_i, \theta_j)(\omega_i - \theta_j) - T(\omega_i, \theta_j).$$

Rearranging, we have

$$(h(\omega_i, \theta_j) - h(\omega_i, \theta_{\hat{j}}))(\theta_{\hat{j}} - \theta_j) \geq 0$$

which proves the claim.

Next we argue that $v(\omega_i, \theta_j) = v(\omega_i, \theta_m) + \sum_{k=j+1}^m h(\omega_i, \theta_k)$ and $h(\omega_i, \theta_j)$ non-increasing in θ_j are sufficient for incentive compatibility. The condition $v(\omega_i, \theta_j) = v(\omega_i, \theta_m) + \sum_{k=j+1}^m h(\omega_i, \theta_k)$ results from imposing the right-wards adjacent incentive constraints with equality and solving recursively. To see this, suppose the right-ward adjacent constraint holds with equality. Then,

$$\begin{aligned} h(\omega_i, \theta_j)(\omega_i - \theta_j) - T(\omega_i, \theta_j) &= h(\omega_i, \theta_{j+1})(\omega_i - \theta_j) - T(\omega_i, \theta_{j+1}) \\ &= h(\omega_i, \theta_{j+1})(\omega_i - \theta_{j+1}) - T(\omega_i, \theta_{j+1}) + h(\omega_i, \theta_{j+1}). \end{aligned}$$

So, $V(\omega_i, \theta_j, \theta_j) = V(\omega_i, \theta_{j+1}, \theta_{j+1}) + h(\omega_i, \theta_{j+1})$. Applying this logic repeatedly and solving recursively, gives expression (11). We wish to show that (11) and (12) jointly imply that any deviation from truth-telling is suboptimal. Notice that the excess income that type (ω_i, θ_j) obtains from mimicking type (ω_i, θ_l) is given by

$$\begin{aligned} V(\omega_i, \theta_l, \theta_j) &= h(\omega_i, \theta_l)(\omega_i - \theta_j) - T(\omega_i, \theta_l) \\ &= h(\omega_i, \theta_l)(\omega_i - \theta_l) - T(\omega_i, \theta_l) - (\theta_j - \theta_l)h(\omega_i, \theta_l) \\ &= V(\omega_i, \theta_l, \theta_l) - (\theta_j - \theta_l)h(\omega_i, \theta_l). \end{aligned}$$

Thus, $V(\omega_i, \theta_j, \theta_j) \geq V(\omega_i, \theta_l, \theta_j)$ for any l and j if

$$V(\omega_i, \theta_m, \theta_m) + \sum_{k=j+1}^m h(\omega_i, \theta_k) \geq V(\omega_i, \theta_m, \theta_m) + \sum_{k=l+1}^m h(\omega_i, \theta_k) - (\theta_j - \theta_l)h(\omega_i, \theta_l).$$

Consider first any $l > j$. We can write the comparison as

$$\begin{aligned} & V(\omega_i, \theta_m, \theta_m) + \sum_{k=l+1}^m h(\omega_i, \theta_k) + h(\omega_i, \theta_{j+1}) \dots + \dots h(\omega_i, \theta_l) \\ & \geq V(\omega_i, \theta_m, \theta_m) + \sum_{k=l+1}^m h(\omega_i, \theta_k) - (\theta_j - \theta_l) h(\omega_i, \theta_l) \end{aligned}$$

Cancelling equal terms on both sides we can simplify the condition to

$$h(\omega_i, \theta_{j+1}) \dots + \dots h(\omega_i, \theta_l) \geq (l - j) h(\omega_i, \theta_l).$$

Since the number of terms on each side is the same, and $h(\omega_i, \theta_j)$ is non-increasing in j , the inequality is satisfied. The proof for the case where $l < j$ is similar and therefore omitted.

Consider now the participation constraints. From the right-wards adjacent incentive constraints, $V(\omega_i, \theta_j, \theta_j) \geq V(\omega_i, \theta_{j+1}, \theta_j) \geq V(\omega_i, \theta_{j+1}, \theta_{j+1})$, and from the participation constraint of type (ω_i, θ_m) , $V(\omega_i, \theta_m, \theta_m) \geq 0$, all the participation constraints are satisfied. Finally, we show that all the incentive constraints must hold with equality. To see this, suppose there is a type (ω_i, θ_j) such that

$$V(\omega_i, \theta_j, \theta_j) > V(\omega_i, \theta_{j+1}, \theta_{j+1}) + h(\omega_i, \theta_{j+1}).$$

Then we can change the incentive system as follows. We can find $\varepsilon_1, \varepsilon_2 > 0$ to change the taxes to

$$\tilde{T}(\omega_i, \theta_j) = T(\omega_i, \theta_j) + \varepsilon_1 \quad \text{and} \quad \tilde{T}(\omega_i, \theta_{j+1}) = T(\omega_i, \theta_{j+1}) - \varepsilon_2.$$

The effect is to reduce type $(\omega_i, \theta_j)'$'s excess income and to increase type $(\omega_i, \theta_{j+1})'$'s excess net income. Let $p_j(i)$ denote the conditional probability that $\tilde{\theta} = \theta_j$ conditional on $\tilde{\omega} = \omega_i$. Since we do not change the allocation of types' (ω_i, θ_j) and (ω_i, θ_{j+1}) working time, we have to respect the condition $p_j(i)\varepsilon_1 = p_{j+1}(i)\varepsilon_2$. By construction, $\begin{pmatrix} V(\omega_i, \theta_j, \theta_j) \\ V(\omega_i, \theta_{j+1}, \theta_{j+1}) \end{pmatrix}$ can be viewed as generated from $\begin{pmatrix} \tilde{V}(\omega_i, \theta_j, \theta_j) \\ \tilde{V}(\omega_i, \theta_{j+1}, \theta_{j+1}) \end{pmatrix}$ by a mean-preserving spread. Since $U(\cdot)$ is concave, the latter gives the objective function a higher value. ■

Proof of Proposition 4. The proof is given in two parts. In the first part, we characterize the optimal allocation. In the second part, we use the structure of the optimal allocation to derive the budget constraint.

Part i: the structure of the allocation

The Lagrangian for our problem takes the form

$$\begin{aligned} L_i &= \sum_{j=1}^m p_j(i) U \left(\theta_j + v(\omega_i, \theta_m) + \sum_{k=j+1}^m h(\omega_i, \theta_k) \right) \\ &\quad + \lambda_i \left(\sum_{j=1}^m p_j(i) \left(h(\omega_i, \theta_j)(\omega_i - \theta_j) - v(\omega_i, \theta_m) - \sum_{k=j+1}^m h(\omega_i, \theta_k) \right) - T \right). \end{aligned}$$

For notational ease in this proof, let the marginal utility of type (ω_i, θ_j) be

$$u_i(\theta_j) \equiv U' \left(\theta_j + v(\omega_i, \theta_m) + \sum_{k=j+1}^m h(\omega_i, \theta_k) \right).$$

The derivative of L_i with respect to $h(\omega_i, \theta_1)$ is equal to

$$\frac{\partial L_i}{\partial h(\omega_i, \theta_1)} = \lambda_i (\omega_i - \theta_1) p_1(i)$$

which implies directly that $h^*(\omega_i, \theta_1) = 1$ since $\omega_i - \theta_1 > 0$.

The derivative of L_i with respect to $h(\omega_i, \theta_z)$ is equal to

$$\frac{\partial L_i}{\partial h(\omega_i, \theta_z)} = \sum_{j=1}^{z-1} p_j(i) u_i(\theta_j) + \lambda_i (p_z(i) (\omega_i - \theta_z) - P_{z-1}(i)).$$

In what follows, we will make repeated use of a convenient transformation. Define

$$E[u_i(\theta_j) | j \leq z-1] \equiv \sum_{j=1}^{z-1} \frac{p_j(i)}{P_{z-1}(i)} u_i(\theta_j).$$

We prove that our problem admits an interior solution for at most one $h(\omega_i, \theta_z)$. The derivative of L_i with respect to $h(\omega_i, \theta_z)$ for $z > 1$ is proportional to

$$\frac{\partial L_i}{\partial h(\omega_i, \theta_z)} = E[u_i(\theta_j) | j \leq z-1] + \lambda_i \left(\frac{p_z(i)}{P_{z-1}(i)} (\omega_i - \theta_z) - 1 \right). \quad (20)$$

Suppose (20) admits an interior solution for $z = j_{t_i}$, so the first-order condition holds:

$$E[u_i(\theta_j) | j \leq j_{t_i} - 1] = \lambda_i \left(1 - \frac{p_{j_{t_i}}(i)}{P_{j_{t_i}-1}(i)} (\omega_i - \theta_{j_{t_i}}) \right).$$

$E[u_i(\theta_j) | j \leq z-1]$ is non-increasing in z . To see this, note that incomes are non-decreasing in opportunity costs, since by incentive compatibility

$$\theta_{j+1} + \sum_{k=j+2}^m h(\omega_i, \theta_k) - \left(\theta_j + \sum_{k=j+1}^m h(\omega_i, \theta_k) \right) = 1 - h(\omega_i, \theta_{j+1}) \geq 0. \quad (21)$$

Hence, to prove our claim, it suffices to show that the expression $\frac{p_z(i)}{P_z(i)} (\omega_i - \theta_z)$ is decreasing in z , because that implies that the first-order condition cannot hold for $z > j_{t_i}$. So we want to show that

$$\frac{p_z(i)}{P_{z-1}(i)} (\omega_i - \theta_z) > \frac{p_{z+1}(i)}{P_z(i)} (\omega_i - \theta_{z+1}).$$

Define j_{FB} by the condition $\omega_i = \theta_{j_{FB}}$. Let $a \equiv j_{FB} - z$. With these definitions, we can write

$$\frac{p_z(i)}{P_{z-1}(i)} (\omega_i - \theta_z) = \frac{p_{j_{FB}-a}(i)}{P_{j_{FB}-a-1}(i)} (\omega_i - \theta_{j_{FB}-a}) = a \frac{p_{j_{FB}-a}(i)}{P_{j_{FB}-a-1}(i)}$$

and

$$\frac{p_{z+1}(i)}{P_z(i)} (\omega_i - \theta_{z+1}) = \frac{p_{j_{FB}-a+1}(i)}{P_{j_{FB}-a}(i)} (\omega_i - \theta_{j_{FB}-a+1}) = (a-1) \frac{p_{j_{FB}-a+1}(i)}{P_{j_{FB}-a}(i)}.$$

So the condition is equivalent to

$$a \frac{p_{j_{FB}-a}(i)}{P_{j_{FB}-a-1}(i)} > (a-1) \frac{p_{j_{FB}-a+1}(i)}{P_{j_{FB}-a}(i)}.$$

The condition is trivially satisfied for $a = 1$; so assume that $a > 1$. Multiplying both sides by $\frac{P_{j_{FB}-a}(i)}{p_{j_{FB}-a+1}(i)} \frac{P_{j_{FB}-a-1}(i)}{p_{j_{FB}-a}(i)}$ and rearranging we have the equivalent condition

$$\frac{P_{j_{FB}-a-1}(i)}{p_{j_{FB}-a}(i)} > a \left(\frac{P_{j_{FB}-a-1}(i)}{p_{j_{FB}-a}(i)} - \frac{P_{j_{FB}-a}(i)}{p_{j_{FB}-a+1}(i)} \right). \quad (22)$$

Notice that the condition is trivially satisfied if the inverse hazard rate is non-decreasing in j ,

because that makes the expression on the right-hand side become negative, while the expression on the left-hand side is strictly positive. However, suppose the inverse hazard rate is decreasing so that the right-hand side is strictly positive. In that case, (22) is still satisfied, provided that

$$\frac{P_{j_{FB}-a-1}(i)}{p_{j_{FB}-a}(i)} \geq \frac{P_{j_{FB}-1}(i)}{p_{j_{FB}}(i)} + a \left(\frac{P_{j_{FB}-a-1}(i)}{p_{j_{FB}-a}(i)} - \frac{P_{j_{FB}-a}(i)}{p_{j_{FB}-a+1}(i)} \right).$$

Rearranging, we have

$$\frac{P_{j_{FB}-a-1}(i)}{p_{j_{FB}-a}(i)} + a \left(\frac{P_{j_{FB}-a}(i)}{p_{j_{FB}-a+1}(i)} - \frac{P_{j_{FB}-a-1}(i)}{p_{j_{FB}-a}(i)} \right) \geq \frac{P_{j_{FB}-1}(i)}{p_{j_{FB}}(i)}. \quad (23)$$

Note that (23) is simply the definition for a decreasing function to be concave. Hence, the solution for $z > j_{t_i}$ is $h^*(\omega_i, \theta_z) = 0$.

Part ii: derivation of the resource constraint

Using the particular allocation, the tax paid by the marginal type satisfies $h_{j_t}(\omega_i - \theta_{j_t}) = T(\omega_i, \theta_{j_t})$ because this type's participation constraint is binding. The excess income of type $(\omega_i, \theta_{j_t-1})$ satisfies $V(\omega_i, \theta_{j_t-1}, \theta_{j_t-1}) = V(\omega_i, \theta_{j_t}, \theta_{j_t-1}) = h_{j_t}$, so his total income is equal to $\theta_{j_t-1} + h_{j_t}$. The taxes he pays satisfy the relation $\omega_i - \theta_{j_t-1} - T(\omega_i, \theta_{j_t-1}) = h_{j_t}$, so

$$T(\omega_i, \theta_{j_t-1}) = \omega_i - (\theta_{j_t-1} + h_{j_t}).$$

Since the marginal utilities of all inframarginal types are the same, all their incomes are the same, so the taxes paid by all inframarginal types are the same. Summing the taxes together we obtain

$$T = P_{j_t-1}(i)(\omega_i - (\theta_{j_t-1} + h_{j_t})) + p_{j_t}(i)h_{j_t}(\omega_i - \theta_{j_t})$$

which is the expression in the proposition. ■

Proof of Proposition 5. In contrast to proposition 3, the left-wards adjacent constraints must bind whenever the left-wards neighbor is working in the formal sector. Imposing these constraints and solving recursively, we find that

$$v(\omega_i, \theta_j) = v(\omega_i, \theta_1) - \sum_{k=1}^{j-1} h(\omega_i, \theta_k) \quad (24)$$

for any type who is supposed to be included in the redistribution program.

It can be shown that the left-ward adjacent incentive constraints plus monotonicity imply that there is no profitable deviation from truth-telling. Since this is standard, it is omitted. Second, following the same proof as in proposition 3, one can show that the adjacent constraints must be tight for all types that work in the formal sector. To avoid repetition, this step is omitted as well.

If the government wishes to include type (ω_i, θ_{j_s}) , then

$$v(\omega_i, \theta_{j_s}) = v(\omega_i, \theta_1) - \sum_{k=1}^{j_s-1} h(\omega_i, \theta_k) \geq 0.$$

The participation constraint of type (ω_i, θ_{j_s}) implies that all types (ω_i, θ_j) for $j < j_s$ also want to participate. On the other hand, the exclusion constraint for type $(\omega_1, \theta_{j_s+1})$,

$$v(\omega_i, \theta_{j_s+1}) = v(\omega_i, \theta_1) - \sum_{k=1}^{j_s} h(\omega_i, \theta_k) \leq 0$$

implies that all types (ω_i, θ_j) for $j > j_{s+1}$ are also excluded. In particular, if type $(\omega_i, \theta_{j_s+2})$

mimics the marginal type (ω_i, θ_{j_s}) he obtains a net excess income of

$$V(\omega_1, \theta_{j_s}, \theta_{j_s+2}) = V(\omega_i, \theta_1, \theta_1) - \sum_{k=1}^{j_s} h(\omega_i, \theta_k) - h(\omega_i, \theta_{j_s}) = -h(\omega_i, \theta_{j_s}) < 0$$

An analogous argument can be given for any type (ω_i, θ_j) for $j > j_s + 1$. ■

Proof of Proposition 6. The proof is given in two parts. In part i, we derive the structure of the allocation. In part ii) we use this structure to derive the representation of the resource constraint.

Part i: structure of the allocation

The Lagrangian function for our problem takes the form

$$\begin{aligned} L_i(s) = & \sum_{j=1}^{j_s} p_j(i) U \left(\theta_j + v(\omega_i, \theta_1) - \sum_{k=1}^{j-1} h(\omega_i, \theta_k) \right) + \sum_{j=j_s+1}^m p_j(i) U(\theta_j) \\ & + \lambda_i(s) \left(\sum_{j=1}^{j_s} p_j(i) \left(h(\omega_i, \theta_j)(\omega_i - \theta_j) - \left(v(\omega_i, \theta_1) - \sum_{k=1}^{j-1} h(\omega_i, \theta_k) \right) \right) + S \right) \\ & + \alpha \left(v(\omega_i, \theta_1) - \sum_{k=1}^{j_s-1} h(\omega_i, \theta_k) \right) - \beta \left(v(\omega_i, \theta_1) - \sum_{k=1}^{j_s} h(\omega_i, \theta_k) \right). \end{aligned}$$

To ease notation in this proof, we shall define the marginal utility of type (ω_i, θ_j) as

$$u_i(\theta_j) \equiv U' \left(\theta_j + v(\omega_i, \theta_1) - \sum_{k=1}^{j-1} h(\omega_i, \theta_k) \right).$$

We begin by stating the derivatives of the objective function with respect to the relevant choice variables. The derivative with respect to $h(\omega_i, \theta_z)$ for $z < j_s$ is equal to

$$\frac{\partial L_i(s)}{\partial h(\omega_i, \theta_z)} = - \sum_{j=z+1}^s p_j(i) u_i(\theta_j) + \lambda_i(s) \left(p_z(i)(\omega_i - \theta_z) + \sum_{j=z+1}^{j_s} p_j(i) \right) - \alpha + \beta. \quad (25)$$

The derivative with respect to $h(\omega_i, \theta_{j_s})$ is equal to

$$\frac{\partial L_i(s)}{\partial h(\omega_i, \theta_{j_s})} = \lambda_i(s) p_{j_s}(i) (\omega_i - \theta_{j_s}) + \beta. \quad (26)$$

The derivative with respect to $v(\omega_i, \theta_1)$ is equal to

$$\frac{\partial L_i(s)}{\partial v(\omega_i, \theta_1)} = \sum_{j=1}^{j_s} p_j(i) u_i(\theta_j) - \lambda_i(s) \sum_{j=1}^{j_s} p_j(i) + \alpha - \beta. \quad (27)$$

We analyze the case where $S^{\min} < S < S^{\max}$. The reason is that for $S \leq S^{\min}$, the first-best is feasible (and the solution is as described in the text) and the case $S > S^{\max}$ cannot be part of an overall optimum. For $S^{\min} < S < S^{\max}$, we must have $\omega_i - \theta_s < 0$. This implies by (26) that $\beta > 0$. To see this, suppose that $\beta = 0$. Then, by (26), we would have $h^*(\omega_i, \theta_{j_s}) = 0$. But then, type $(\omega_i, \theta_{j_s+1})$ is not excluded, so the allocation is not incentive compatible. Next, notice that $\alpha = 0$ at the optimum. If both β and α were strictly positive, then - since both constraints must hold with equality - we would have again that $h^*(\omega_i, \theta_{j_s}) = 0$, which means that effectively type $(\omega_i, \theta_{j_s-1})$ is the marginal type. Finally, at any optimum both $h(\omega_i, \theta_{j_s})$ and $v(\omega_i, \theta_1)$ must be at stationary points, so $\frac{\partial L_i(s)}{\partial h(\omega_i, \theta_{j_s})} = 0$ and $\frac{\partial L_i(s)}{\partial v(\omega_i, \theta_1)} = 0$. From (27) we have the first-order condition

for $v(\omega_i, \theta_1)$

$$\sum_{j=1}^{j_s} p_j(i) u_i(\theta_j) - \lambda_i(s) \sum_{j=1}^{j_s} p_j(i) = \beta. \quad (28)$$

Substituting (28) into the condition (26), we obtain

$$\frac{\partial L_i(s)}{\partial h(\omega_i, \theta_{j_s})} = \sum_{j=1}^{j_s} p_j(i) u_i(\theta_j) + \lambda_i(s) \left(p_s(i) (\omega_i - \theta_s) - \sum_{j=1}^{j_s} p_j(i) \right). \quad (29)$$

Substituting (28) into (25), we obtain

$$\begin{aligned} \frac{\partial L_i(s)}{\partial h(\omega_i, \theta_z)} &= - \sum_{j=z+1}^{j_s} p_j(i) u_i(\theta_j) + \lambda_i(s) \left(p_z(i) (\omega_i - \theta_z) + \sum_{j=z+1}^{j_s} p_j(i) \right) \\ &\quad + \sum_{j=1}^{j_s} p_j(i) u_i(\theta_j) - \lambda_i(s) \sum_{j=1}^{j_s} p_j(i) \\ &= \sum_{j=1}^z p_j(i) u_i(\theta_j) + \lambda_i(s) \left(p_z(i) (\omega_i - \theta_z) - \sum_{j=1}^z p_j(i) \right). \end{aligned} \quad (30)$$

Dividing by $P_z(i)$, we can write both (29) and (30) as

$$\frac{\partial L_i(s)}{\partial h(\omega_i, \theta_z)} = \sum_{j=1}^z \frac{p_j(i)}{P_z(i)} u_i(\theta_j) + \lambda_i(s) \left(\frac{p_z(i)}{P_z(i)} (\omega_i - \theta_z) - 1 \right) \quad (31)$$

for $z \leq j_s$. From our derivation above, the right-hand side of (31) is equal to zero at $z = j_s$, so

$$\sum_{j=1}^{j_s} \frac{p_j(i)}{P_{j_s}(i)} u_i(\theta_j) + \lambda_i(s) \left(\frac{p_{j_s}(i)}{P_{j_s}(i)} (\omega_i - \theta_{j_s}) - 1 \right) = 0. \quad (32)$$

To prove our proposition, it suffices to show that (32), in conjunction with Assumption 1 implies that

$$\sum_{j=1}^z \frac{p_j(i)}{P_z(i)} u_i(\theta_j) + \lambda_i(s) \left(\frac{p_z(i)}{P_z(i)} (\omega_i - \theta_z) - 1 \right) > 0$$

for all $z < j_s$. Letting $E[u_i(\theta_j) | j \leq z] \equiv \sum_{j=1}^z \frac{p_j(i)}{P_z(i)} u_i(\theta_j)$ this inequality can be written as

$$E[u_i(\theta_j) | j \leq z] > \lambda_i(s) \left(1 - \frac{p_z(i)}{P_z(i)} (\omega_i - \theta_z) \right).$$

We note that type (ω_i, θ_z) receives a weakly higher total income than type (ω_i, θ_{z-1}) , since

$$\theta_z + v(\omega_i, \theta_1) - \sum_{k=1}^{z-1} h(\omega_i, \theta_k) - \left(\theta_{z-1} + v(\omega_i, \theta_1) - \sum_{k=1}^{z-2} h(\omega_i, \theta_k) \right) = 1 - h(\omega_i, \theta_{z-1}) \geq 0.$$

Therefore, $E[u_i(\theta_j) | j \leq z]$ is non-increasing in z . Hence, $E[u_i(\theta_j) | j \leq z] \geq E[u_i(\theta_j) | j \leq j_s]$ for all $z < j_s$. To complete the argument, it suffices to show that $\left(1 - \frac{p_z(i)}{P_z(i)} (\omega_i - \theta_z) \right) < \left(1 - \frac{p_{j_s}(i)}{P_{j_s}(i)} (\omega_i - \theta_{j_s}) \right)$

for all $z < j_s$. This is equivalent to

$$\frac{p_z(i)}{P_z(i)}(\theta_z - \omega_i) < \frac{p_{j_s}(i)}{P_{j_s}(i)}(\theta_{j_s} - \omega_i). \quad (33)$$

It is easy to see that this condition is verified for all z such that $\theta_z \leq \omega_i$. We now prove that, under Assumption 1, the condition is also verified for any z such that $j_{FB} < z < j_s$.

In particular, we show that Assumption 1 implies that for all $z \leq j_s$

$$\frac{p_{z-1}(i)}{P_{z-1}(i)}(\theta_{z-1} - \omega_i) < \frac{p_z(i)}{P_z(i)}(\theta_z - \omega_i)$$

which in turn implies (33). To see this, it proves convenient to normalize this monotonicity condition around j_{FB} . Let $a \equiv z - j_{FB}$. With that definition, we can write

$$\frac{p_z(i)}{P_z(i)}(\theta_z - \omega_i) = \frac{p_{j_{FB}+a}(i)}{P_{j_{FB}+a}(i)}(\theta_{j_{FB}+a} - \omega_i) = a \frac{p_{j_{FB}+a}(i)}{P_{j_{FB}+a}(i)}$$

and

$$\frac{p_{z-1}(i)}{P_{z-1}(i)}(\theta_{z-1} - \omega_i) = \frac{p_{j_{FB}+a-1}(1)}{P_{j_{FB}+a-1}(1)}(\theta_{j_{FB}+a-1} - \omega_i) = (a-1) \frac{p_{j_{FB}+a-1}(i)}{P_{j_{FB}+a-1}(i)}.$$

So, we wish to show that

$$(a-1) \frac{p_{j_{FB}+a-1}(i)}{P_{j_{FB}+a-1}(i)} < a \frac{p_{j_{FB}+a}(i)}{P_{j_{FB}+a}(i)}.$$

This condition is trivially satisfied for $a = 1$. So consider the case where $a > 1$. Manipulating this condition the same way as we did in the case of taxation, we have the equivalent condition that

$$\frac{P_{j_{FB}+a}(i)}{p_{j_{FB}+a}(i)} > a \left(\frac{P_{j_{FB}+a}(i)}{p_{j_{FB}+a}(i)} - \frac{P_{j_{FB}+a-1}(i)}{p_{j_{FB}+a-1}(i)} \right). \quad (34)$$

Finally, notice that $\frac{P_z(i)}{p_z(i)}$ concave in z implies condition (34). To see this, observe simply that the definition of a concave function is that

$$\frac{P_{z+a}(i)}{p_{z+a}(i)} \geq \frac{P_z(i)}{p_z(i)} + a \left(\frac{P_{z+a}(i)}{p_{z+a}(i)} - \frac{P_{z+a-1}(i)}{p_{z+a-1}(i)} \right) \quad (35)$$

for any z and any $a \geq 0$. Since $\frac{P_z(i)}{p_z(i)} > 0$, (35) implies (34).

Part ii: Derivation of the Resource Constraint

With a binding exclusion constraint we have $v(\omega_i, \theta_1) - \sum_{k=1}^{j_s} h(\omega_i, \theta_k) = 0$. Therefore, the excess income of all types who receive subsidies are given by

$$v(\omega_i, \theta_j) = v(\omega_i, \theta_1) - \sum_{k=1}^{j-1} h(\omega_i, \theta_k) = \sum_{k=1}^{j_s} h(\omega_i, \theta_k) - \sum_{k=1}^{j-1} h(\omega_i, \theta_k) = \sum_{k=j}^{j_s} h(\omega_i, \theta_k).$$

We can calculate the individual subsidies, $S(\omega_i, \theta_j) = -T(\omega_i, \theta_j)$, using the relation

$$v(\omega_i, \theta_j) = h(\omega_i, \theta_j)(\omega_i - \theta_j) + S(\omega_i, \theta_j)$$

Hence,

$$S(\omega_i, \theta_j) = \sum_{k=j}^{j_s} h(\omega_i, \theta_k) - h(\omega_i, \theta_j)(\omega_i - \theta_j)$$

Using the structure of the allocation, we get

$$S(\omega_i, \theta_j) = \begin{cases} \theta_{j_s} - \omega_i + h_{j_s} & \text{for } \theta_j < \theta_{j_s} \\ h_{j_s}(\theta_{j_s+1} - \omega_i) & \text{for } \theta_j = \theta_{j_s} \\ 0 & \text{for } \theta_j > \theta_{j_s} \end{cases}$$

Summing these individual subsidies up we obtain

$$P_{j_s-1}(i)(\theta_{j_s} - \omega_i + h_{j_s}) + p_{j_s}(i)h_{j_s}(\theta_{j_s+1} - \omega_i) = S. \quad (36)$$

The marginal type is chosen optimally if j_s is as large as possible. If j_s can still be increased, this means that we can find a Pareto improvement as follows. By raising j_s , fewer types are excluded. All types that are included receive the same level of income. Hence, by raising j_s we raise all the incomes of all types that are included. The incomes of those who are and remain excluded are unchanged. ■

Proof of Proposition 8. Suppose the contrapositive were true and there were a productivity group ω_i that is subsidized and a productivity group ω_{i-1} that is taxed. Based on this assumption, we will construct a budget balanced, incentive compatible redistribution scheme between these two groups. It follows that the initial allocation was not optimal.

The idea of the redistribution scheme is as follows. Given $\omega_1 \geq \theta_1 > 0$, there is in each productivity group a set of individuals with low opportunity costs of time who will work full time at the optimal allocation. In groups that are taxed, the right-wards incentive constraints are tight. It follows that the marginal type, who works part time, has a strict preference for his own allocation relative to mimicking his left-wards neighbor who works full time. To see this formally, recall that the optimal allocation satisfies

$$V(\omega_{i-1}, \theta_{j_{t_{i-1}}-1}, \theta_{j_{t_{i-1}}-1}) = V(\omega_{i-1}, \theta_{j_{t_{i-1}}}, \theta_{j_{t_{i-1}}-1}) = V(\omega_{i-1}, \theta_{j_{t_{i-1}}}, \theta_{j_{t_{i-1}}}) + h_{j_{t_{i-1}}}.$$

Hence, we can write

$$V(\omega_{i-1}, \theta_{j_{t_{i-1}}}, \theta_{j_{t_{i-1}}}) = V(\omega_{i-1}, \theta_{j_{t_{i-1}}-1}, \theta_{j_{t_{i-1}}-1}) - h_{j_{t_{i-1}}}.$$

If the marginal type mimics his left-wards neighbor, he obtains excess income

$$V(\omega_{i-1}, \theta_{j_{t_{i-1}}-1}, \theta_{j_{t_{i-1}}}) = V(\omega_{i-1}, \theta_{j_{t_{i-1}}-1}, \theta_{j_{t_{i-1}}-1}) - 1.$$

But then it follows that

$$\begin{aligned} V(\omega_{i-1}, \theta_{j_{t_{i-1}}-1}, \theta_{j_{t_{i-1}}}) &= V(\omega_{i-1}, \theta_{j_{t_{i-1}}-1}, \theta_{j_{t_{i-1}}-1}) - 1 \\ &< V(\omega_{i-1}, \theta_{j_{t_{i-1}}-1}, \theta_{j_{t_{i-1}}-1}) - h_{j_{t_{i-1}}} = V(\omega_{i-1}, \theta_{j_{t_{i-1}}}, \theta_{j_{t_{i-1}}}). \end{aligned}$$

Hence, we can decrease the taxes paid by all individuals who work full time by an identical amount, say ε_{i-1} , without violating incentive compatibility.

In the group that is subsidized, we can decrease the subsidies paid to all individuals who work full time by an amount ε_i without affecting incentive compatibility and the exclusion constraint. To see this, recall that we have imposed the left-wards constraint for the marginal type so that

$$V(\omega_i, \theta_{j_{s_i}}, \theta_{j_{s_i}}) = V(\omega_i, \theta_{j_{s_i}-1}, \theta_{j_{s_i}}) = V(\omega_i, \theta_{j_{s_i}-1}, \theta_{j_{s_i}-1}) - 1$$

and we can write

$$V(\omega_i, \theta_{j_{s_i}-1}, \theta_{j_{s_i}-1}) = V(\omega_i, \theta_{j_{s_i}}, \theta_{j_{s_i}}) + 1.$$

If type $(\omega_i, \theta_{j_{s_i}-1})$ mimics his right-wards neighbor, then he would obtain an excess income of

$$V(\omega_i, \theta_{j_{s_i}}, \theta_{j_{s_i}-1}) = V(\omega_i, \theta_{j_{s_i}}, \theta_{j_{s_i}}) + h_{s_i}.$$

Hence, it follows that

$$V(\omega_i, \theta_{j_{s_{i-1}}-1}, \theta_{j_{s_{i-1}}-1}) = V(\omega_i, \theta_{j_{s_i}}, \theta_{j_{s_i}}) + 1 > V(\omega_i, \theta_{j_{s_i}}, \theta_{j_{s_i}}) + h_{s_i} = V(\omega_i, \theta_{j_{s_i}}, \theta_{j_{s_i}-1}).$$

Choose ε_i and ε_{i-1} such that

$$p_i P_{j_{s_i}-1}(i) \varepsilon_i + p_{i-1} P_{j_{s_i}-1}(i) \varepsilon_{i-1} = 0.$$

By construction, the new allocation and the initial allocation generate the same expected level of income for all groups together. However, the distributions differ by a mean preserving spread. Hence, the new allocation is preferred. ■

The following definitions and lemmas are used in the proof of proposition 8.

Definition *Market and non-market productivities are called affiliated if for any $j > j'$ and any integer $b > 0$*

$$\frac{p_{ij'}}{p_{ij}} - \frac{p_{i+b,j'}}{p_{i+b,j}} \geq 0 \quad (37)$$

and are called negatively affiliated if for any $j > j'$ and any integer $b > 0$

$$\frac{p_{ij'}}{p_{ij}} - \frac{p_{i+b,j'}}{p_{i+b,j}} \leq 0. \quad (38)$$

We say that the degree of affiliation is non-decreasing in i if for any $j > j'$ and any integer $b > 0$

$$\frac{p_{ij'}}{p_{ij}} - \frac{p_{i+b,j'}}{p_{i+b,j}} \geq \frac{p_{i'j'}}{p_{i'j}} - \frac{p_{i'+b,j'}}{p_{i'+b,j}} \text{ for any } i > i'. \quad (39)$$

Lemma 1 *If market and non-market productivities are affiliated (negatively affiliated), then $\frac{P_{j-1}(i)}{p_j(i)}$ is non-increasing (non-decreasing) in i .*

Proof. Take the definition of affiliation and rearrange it to get, for any $j > j'$ and any integer $b > 0$

$$p_{i+b,j} p_{ij'} \geq p_{ij} p_{i+b,j'}$$

mutliply both sides by $\frac{1}{p_{i+b}} \frac{1}{p_i}$ to obtain

$$p_j(i+b) p_{j'}(i) \geq p_j(i) p_{j'}(i+b) \text{ for } j > j'. \quad (40)$$

We can sum (40) over all $j' < j$ to obtain

$$p_j(i+b) \sum_{j'=1}^{j-1} p_{j'}(i) \geq p_j(i) \sum_{j'=1}^{j-1} p_{j'}(i+b) \text{ for } i > i'.$$

Computing the sum and rearranging, we have shown that

$$\frac{P_{j-1}(i)}{p_j(i)} \geq \frac{P_{j-1}(i+b)}{p_j(i+b)} \text{ for any } j.$$

In case of negative affiliation, all the inequalities are reversed. ■

Lemma 2 *If the degree of affiliation is non-decreasing in i then then $\frac{P_j(i)}{p_j(i)}$ is concave in i .*

Proof of Lemma. With $b = 1$ and $i' = i - 1$ we have from (39)

$$\frac{p_{ij'}}{p_{ij}} - \frac{p_{i+1,j'}}{p_{i+1,j}} \geq \frac{p_{i-1,j'}}{p_{i-1,j}} - \frac{p_{i,j'}}{p_{i,j}}.$$

Multiplying the first ratio by $\frac{p_i}{p_i}$, the second by $\frac{p_{i+1}}{p_{i+1}}$, and so on, we can write

$$\frac{p_i}{p_i} \frac{p_{ij'}}{p_{ij}} - \frac{p_{i+1}}{p_{i+1}} \frac{p_{i+1,j'}}{p_{i+1,j}} \geq \frac{p_{i-1}}{p_{i-1}} \frac{p_{i-1,j'}}{p_{i-1,j}} - \frac{p_i}{p_i} \frac{p_{ij'}}{p_{ij}}.$$

Substituting for $\frac{p_{ij'}}{p_{ij}} = p_{j'}(i)$, and for analogous terms in the remaining ratios, we have

$$\frac{p_{j'}(i)}{p_j(i)} - \frac{p_{j'}(i+1)}{p_j(i+1)} \geq \frac{p_{j'}(i-1)}{p_j(i-1)} - \frac{p_{j'}(i)}{p_j(i)}.$$

Summing for $j' = 1, \dots, j-1$, we can write

$$\sum_{j'=1}^{j-1} \frac{p_{j'}(i)}{p_j(i)} - \sum_{j'=1}^{j-1} \frac{p_{j'}(i+1)}{p_j(i+1)} \geq \sum_{j'=1}^{j-1} \frac{p_{j'}(i-1)}{p_j(i-1)} - \sum_{j'=1}^{j-1} \frac{p_{j'}(i)}{p_j(i)}.$$

Performing this summation, we have

$$\frac{P_{j-1}(i)}{p_j(i)} - \frac{P_{j-1}(i+1)}{p_j(i+1)} \geq \frac{P_{j-1}(i-1)}{p_j(i-1)} - \frac{P_{j-1}(i)}{p_j(i)}.$$

Adding $\frac{p_j(i)}{p_j(i)} - \frac{p_j(i+1)}{p_j(i+1)} = 0$ on the left-hand side and $\frac{p_j(i-1)}{p_j(i-1)} - \frac{p_j(i)}{p_j(i)} = 0$ on the right-hand side, we obtain

$$\frac{P_j(i)}{p_j(i)} - \frac{P_j(i+1)}{p_j(i+1)} \geq \frac{P_j(i-1)}{p_j(i-1)} - \frac{P_j(i)}{p_j(i)}.$$

Rearranging this condition, we have

$$\frac{P_j(i)}{p_j(i)} \geq \frac{1}{2} \frac{P_j(i-1)}{p_j(i-1)} + \frac{1}{2} \frac{P_j(i+1)}{p_j(i+1)},$$

or, in other words, that $\frac{P_j(i)}{p_j(i)}$ is concave in i . ■

Proof of Proposition 9. If $\frac{P_j(i)}{p_j(i)}$ is concave in i , three cases can arise. First, $\frac{P_j(i)}{p_j(i)}$ can be non-increasing in i for all i . Second, $\frac{P_j(i)}{p_j(i)}$ can be non-decreasing in i for small i and then non-increasing in i for higher i . Third, $\frac{P_j(i)}{p_j(i)}$ can be non-decreasing in i for all. The first case arises if productivities are affiliated at all levels of market productivity i ; the second case arises if productivities are negatively affiliated at low levels of i and positively affiliated at high levels of i ; and the third case arises if the productivities are negatively affiliated at all levels of i . We prove our result for the first in detail and show how it extends to the second case. The proof for the third case follows directly from these arguments and is therefore omitted. In each case, we begin proving that no individual in market productivity group ω_i , which is taxed, has an incentive to mimic his opportunity cost counterpart in the productivity group ω_{i-1} , which is also taxed. Then, we show that no individual in the least productive market productivity group ω_l , which is still taxed, has an incentive to mimic his counterpart in the productivity group ω_{l-1} , which receives subsidies. Finally, we show that no individual in a productivity group ω_i which is subsidized, has an incentive to mimic his counterpart in the productivity group ω_{i-1} , which also receives a subsidy ($i < l$). Finally, we repeat this analysis for the case where $\frac{P_j(i)}{p_j(i)}$ is first non-decreasing and then non-increasing in i , that is the case of negative affiliation for low levels of i .

Case 1: positive, non-decreasing affiliation.

Case 1a: $i > l$

We compare the marginal utilities of the inframarginal types in each productivity group and conclude that the identity of the marginal type is monotonic in the market productivity. Suppose the equation

$$U'(\theta_{z-1} + h) = \lambda \left(1 - \frac{p_z(i-1)}{P_{z-1}(i-1)} (\omega_{i-1} - \theta_z) \right)$$

has a solution at $z = j_{t_{i-1}}$ for some $h = h_{j_{t_{i-1}}} \in [0, 1]$, so

$$U'(\theta_{j_{t_{i-1}}-1} + h_{j_{t_{i-1}}}) = \lambda \left(1 - \frac{p_{j_{t_{i-1}}}(i-1)}{P_{j_{t_{i-1}}-1}(i-1)} (\omega_{i-1} - \theta_{j_{t_{i-1}}}) \right).$$

Consider now the derivative of the payoff function with respect to $h(\omega_i, \theta_{j_{t_{i-1}}})$. Fix the allocation for the group with market productivity ω_i at the optimal allocation for group ω_{i-1} , that is at $h(\omega_i, \theta_j) = 1$ for $j < j_{t_{i-1}}$ and $h(\omega_i, \theta_{j_{t_{i-1}}}) = h_{j_{t_{i-1}}}$, and consider the derivative of the payoff with respect to $h(\omega_i, \theta_{j_{t_{i-1}}})$, evaluated at this allocation. At this allocation, all individuals with opportunity costs less than or equal to $\theta_{j_{t_{i-1}}}$ receive the same total income; so the derivative of the payoff function is equal to

$$U'(\theta_{j_{t_{i-1}}-1} + h_{j_{t_{i-1}}}) - \lambda \left(1 - \frac{p_{j_{t_{i-1}}}(i)}{P_{j_{t_{i-1}}-1}(i)} (\omega_i - \theta_{j_{t_{i-1}}}) \right).$$

This expression is strictly positive if and only if

$$\lambda \left(1 - \frac{p_{j_{t_{i-1}}}(i)}{P_{j_{t_{i-1}}-1}(i)} (\omega_i - \theta_{j_{t_{i-1}}}) \right) < \lambda \left(1 - \frac{p_{j_{t_{i-1}}}(i-1)}{P_{j_{t_{i-1}}-1}(i-1)} (\omega_{i-1} - \theta_{j_{t_{i-1}}}) \right).$$

In turn, this condition is equivalent to

$$\frac{p_{j_{t_{i-1}}}(i)}{P_{j_{t_{i-1}}-1}(i)} (\omega_i - \theta_{j_{t_{i-1}}}) > \frac{p_{j_{t_{i-1}}}(i-1)}{P_{j_{t_{i-1}}-1}(i-1)} (\omega_{i-1} - \theta_{j_{t_{i-1}}}).$$

By the lemma, we have $\frac{P_{j-1}(i)}{p_j(i)} \geq \frac{P_{j-1}(i+b)}{p_j(i+b)}$ for any j , so

$$\frac{p_{j_{t_{i-1}}}(i)}{P_{j_{t_{i-1}}-1}(i)} \geq \frac{p_{j_{t_{i-1}}}(i-1)}{P_{j_{t_{i-1}}-1}(i-1)}.$$

Moreover, $\omega_i - \theta_{j_{t_{i-1}}} > \omega_{i-1} - \theta_{j_{t_{i-1}}} \geq 0$ where the last inequality follows from the fact that we want to tax individual $(\omega_{i-1}, \theta_{j_{t_{i-1}}})$. So, we have shown that

$$\begin{aligned} & U'(\theta_{j_{t_{i-1}}-1} + h_{j_{t_{i-1}}}) - \lambda \left(1 - \frac{p_{j_{t_{i-1}}}(i)}{P_{j_{t_{i-1}}-1}(i)} (\omega_i - \theta_{j_{t_{i-1}}}) \right) \\ & > U'(\theta_{j_{t_{i-1}}-1} + h_{j_{t_{i-1}}}) - \lambda \left(1 - \frac{p_{j_{t_{i-1}}}(i-1)}{P_{j_{t_{i-1}}-1}(i-1)} (\omega_{i-1} - \theta_{j_{t_{i-1}}}) \right) = 0. \end{aligned}$$

It follows that

$$\theta_{j_{t_i}} + h_{j_{t_i}} > \theta_{j_{t_{i-1}}} + h_{j_{t_{i-1}}}. \quad (41)$$

Therefore, the inframarginal types have no incentive to mimic their downward counterparts. Consider now the marginal types. The proof is trivial if $j_{t_i} > j_{t_{i-1}}$, because the marginal type in group i just receives his outside value $\theta_{j_{t_i}}$ when he mimics his downward counterpart. In case $j_{t_i} = j_{t_{i-1}}$, it is also easy to see that the marginal type has no incentive to mimic his downward counterpart,

because both obtain the value of their outside options. Finally, we show that the case $j_{t_i} < j_{t_{i-1}}$ can never arise at the optimum. Subtracting $\theta_{j_{t_i}}$ on both sides of inequality (41), we have

$$h_{j_{t_i}} > \theta_{j_{t_{i-1}}} - \theta_{j_{t_i}} + h_{j_{t_{i-1}}}.$$

But if $j_{t_i} < j_{t_{i-1}}$ then $\theta_{j_{t_{i-1}}} - \theta_{j_{t_i}} \geq 1$, so $\theta_{j_{t_{i-1}}} - \theta_{j_{t_i}} + h_{j_{t_{i-1}}} \geq 1 + h_{j_{t_{i-1}}}$. However, that can only hold if

$$h_{j_{t_i}} > 1 + h_{j_{t_{i-1}}}.$$

However, this contradicts the fact that both $h_{j_{t_i}}$ and $h_{j_{t_{i-1}}}$ belong to the unit interval.

Case 1b: $i = l$

For the group that is taxed, the marginal utility of the marginal type satisfies

$$U'(\theta_{j_{t_i}-1} + h_{j_{t_i}}) = \lambda \left(1 - \frac{p_{j_{t_i}}(i)}{P_{j_{t_i}-1}(i)} (\omega_i - \theta_{j_{t_i}}) \right).$$

For the group that is subsidized, the marginal utility of the marginal type satisfies

$$U'(\theta_{j_{s_{i-1}}} + h_{j_{s_{i-1}}}) = \lambda \left(1 - \frac{p_{j_{s_{i-1}}}(i-1)}{P_{j_{s_{i-1}}}(i-1)} (\omega_{i-1} - \theta_{j_{s_{i-1}}}) \right).$$

These solutions satisfy $U'(\theta_{j_{t_i}-1} + h_{j_{t_i}}) < U'(\theta_{j_{s_{i-1}}} + h_{j_{s_{i-1}}})$ if and only if

$$\lambda \left(1 - \frac{p_{j_{t_i}}(i)}{P_{j_{t_i}-1}(i)} (\omega_i - \theta_{j_{t_i}}) \right) < \lambda \left(1 - \frac{p_{j_{s_{i-1}}}(i-1)}{P_{j_{s_{i-1}}}(i-1)} (\omega_{i-1} - \theta_{j_{s_{i-1}}}) \right).$$

It is easy to see that this condition must always hold, since we have $\omega_i - \theta_{j_{t_i}} > 0$ and $\omega_{i-1} - \theta_{j_{s_{i-1}}} \leq 0$. The former property is necessary since group ω_i is taxed; the latter property is optimal since incomes of individuals in group ω_{i-1} are raised. It follows that the solution is incentive compatible among inframarginal individuals. It follows also that

$$\theta_{j_{t_i}-1} + h_{j_{t_i}} > \theta_{j_{s_{i-1}}} + h_{j_{s_{i-1}}}.$$

Consider the incentive of the marginal type to mimic his downward counterpart. If $j_{t_i} - 1 \geq j_{s_{i-1}}$, then the marginal type just receives the value of his outside option when he mimics his downward counterpart, so he has no incentive to do that. The case where $j_{t_i} - 1 < j_{s_{i-1}}$ cannot arise at the optimum, since this would imply that

$$h_{j_{t_i}} > \theta_{j_{s_{i-1}}} - \theta_{j_{t_i}-1} + h_{j_{s_{i-1}}} \geq 1 + h_{j_{s_{i-1}}}$$

but that implies that $h_{j_{t_i}}$ is larger than one.

Case 1c: $i < l$

Consider now the individuals in two adjacent productivity groups that are subsidized. Inframarginal individuals in the lower productivity group receive incomes equal to $\theta_{j_{s_{i-1}}} + h_{j_{s_{i-1}}}$; inframarginal individuals in the higher productivity group receive incomes equal to $\theta_{j_{s_i}} + h_{j_{s_i}}$. Suppose the equation

$$U'(\theta_z + h_{i-1}) = \lambda \left(1 - \frac{p_z(i-1)}{P_z(i-1)} (\omega_{i-1} - \theta_z) \right)$$

has a solution for $z = j_{s_{i-1}}$ and some $h_{j_{s_{i-1}}} \in (0, 1]$, so

$$U'(\theta_{j_{s_{i-1}}} + h_{j_{s_{i-1}}}) = \lambda \left(1 - \frac{p_{j_{s_{i-1}}}(i-1)}{P_{j_{s_{i-1}}}(i-1)} (\omega_{i-1} - \theta_{j_{s_{i-1}}}) \right).$$

As for case 1) we evaluate the derivative of the payoff function for the group with market productivity ω_i at the optimal allocation for group ω_{i-1} , to show that

$$U' \left(\theta_{j_{s_{i-1}}} + h_{j_{s_{i-1}}} \right) - \lambda \left(1 - \frac{p_{j_{s_{i-1}}}(i)}{P_{j_{s_{i-1}}}(i)} (\omega_i - \theta_{j_{s_{i-1}}}) \right) > 0.$$

This statement is true if and only if

$$\left(1 - \frac{p_{j_{s_{i-1}}}(i)}{P_{j_{s_{i-1}}}(i)} (\omega_i - \theta_{j_{s_{i-1}}}) \right) < \left(1 - \frac{p_{j_{s_{i-1}}}(i-1)}{P_{j_{s_{i-1}}}(i-1)} (\omega_{i-1} - \theta_{j_{s_{i-1}}}) \right)$$

or, equivalently, if and only if

$$\left(\frac{P_{j_{s_{i-1}}}(i-1)}{p_{j_{s_{i-1}}}(i-1)} - \frac{P_{j_{s_{i-1}}}(i)}{p_{j_{s_{i-1}}}(i)} \right) (\theta_{j_{s_{i-1}}} - \omega_i) < \frac{P_{j_{s_{i-1}}}(i)}{p_{j_{s_{i-1}}}(i)}. \quad (42)$$

We now prove that $\frac{P_j(i)}{p_j(i)}$ non-increasing and concave in i implies that (42) holds. $\frac{P_{j-1}(i)}{p_j(i)} \geq \frac{P_{j-1}(i+b)}{p_j(i+b)}$

From affiliation, the expression on the left-hand side is of (42) is non-negative. We can derive an upper bound for the expression on the left-hand side. It is tautologically true that

$$\theta_{j_{s_{i-1}}} - \omega_i = \theta_{1+j_{s_{i-1}}-1} - \omega_{1+i-1} = \theta_1 - \omega_1 + j_{s_{i-1}} - 1 - (i-1) = \theta_1 - \omega_1 + j_{s_{i-1}} - i.$$

By assumption, we have $\theta_1 - \omega_1 = 0$. Therefore, we can write

$$\theta_{j_{s_{i-1}}} - \omega_i = j_{s_{i-1}} - i \leq m - i,$$

where the last inequality follows from the fact that $j_{s_{i-1}} \leq m$. So, we can write

$$\left(\frac{P_{j_{s_{i-1}}}(i-1)}{p_{j_{s_{i-1}}}(i-1)} - \frac{P_{j_{s_{i-1}}}(i)}{p_{j_{s_{i-1}}}(i)} \right) (\theta_{j_{s_{i-1}}} - \omega_i) \leq \left(\frac{P_{j_{s_{i-1}}}(i-1)}{p_{j_{s_{i-1}}}(i-1)} - \frac{P_{j_{s_{i-1}}}(i)}{p_{j_{s_{i-1}}}(i)} \right) (m - i) + \frac{P_{j_{s_{i-1}}}(m)}{p_{j_{s_{i-1}}}(m)}.$$

Thus, our condition is satisfied if we have

$$\frac{P_{j_{s_{i-1}}}(m)}{p_{j_{s_{i-1}}}(m)} + \left(\frac{P_{j_{s_{i-1}}}(i-1)}{p_{j_{s_{i-1}}}(i-1)} - \frac{P_{j_{s_{i-1}}}(i)}{p_{j_{s_{i-1}}}(i)} \right) (m - i) < \frac{P_{j_{s_{i-1}}}(i)}{p_{j_{s_{i-1}}}(i)}.$$

Rearranging appropriately, we have

$$\frac{P_{j_{s_{i-1}}}(i)}{p_{j_{s_{i-1}}}(i)} + \left(\frac{P_{j_{s_{i-1}}}(i)}{p_{j_{s_{i-1}}}(i)} - \frac{P_{j_{s_{i-1}}}(i-1)}{p_{j_{s_{i-1}}}(i-1)} \right) (m - i) > \frac{P_{j_{s_{i-1}}}(m)}{p_{j_{s_{i-1}}}(m)}. \quad (43)$$

By construction, condition (43) is a sufficient condition for condition (42). To complete the proof, simply note that (43) is an equivalent way of stating that the function $\frac{P_z(i)}{p_z(i)}$ is concave in i . The proof follows then from the lemma. So, we have shown that $\theta_{j_{s_i}} + h_{j_{s_i}} > \theta_{j_{s_{i-1}}} + h_{j_{s_{i-1}}}$. By the now familiar arguments, this implies that no type, neither inframarginal nor marginal, has an incentive to mimic his downward neighbor.

Case 2: negative affiliation for low levels of i , positive affiliation for high levels of i . The proof of case 1b) does not rely on the affiliation assumptions, so it carries over unchanged. The proofs for the cases 1a) where $i > l$ and 1c) where $i < l$, can be interchanged; so we present them in reverse order. In each of the cases, a difference arises only on the part of the distribution where $\frac{P_j(i)}{p_j(i)}$ is increasing in i , so we focus on this part in what follows.

Case 2c) $i < l$:

Building on the arguments given above, the allocation is incentive compatible in the market

productivity dimension among the inframarginal types if it is true that

$$\frac{P_{j_{s_{i-1}}}(i-1)}{p_{j_{s_{i-1}}}(i-1)} (\theta_{j_{s_{i-1}}} - \omega_i) < \frac{P_{j_{s_{i-1}}}(i)}{p_{j_{s_{i-1}}}(i)} (\theta_{j_{s_{i-1}}} - \omega_{i-1}).$$

Using the second part of the lemma, we observe that this condition is verified, since $\frac{P_{j_{s_{i-1}}}(i)}{p_{j_{s_{i-1}}}(i)} \geq \frac{P_{j_{s_{i-1}}}(i-1)}{p_{j_{s_{i-1}}}(i-1)}$ and $\theta_{j_{s_{i-1}}} - \omega_{i-1} > \theta_{j_{s_{i-1}}} - \omega_i$. It follows from these arguments that at the optimum

$$\theta_{j_{s_i}} + h_{j_{s_i}} > \theta_{j_{s_{i-1}}} + h_{j_{s_{i-1}}}.$$

Therefore, the inframarginal types have no incentive to mimic their downward counterparts. Consider now the marginal types. The proof is trivial if $j_{s_i} > j_{s_{i-1}}$, because the marginal type in group i just receives his outside value $\theta_{j_{s_i}}$ when he mimics his downward counterpart. In case $j_{s_i} = j_{s_{i-1}}$, it is also easy to see that the marginal type has no incentive to mimic his downward counterpart, because both obtain the value of their outside options. Finally, we show that the case $j_{s_i} < j_{s_{i-1}}$ can never arise at the optimum. Subtracting $\theta_{j_{s_i}}$ on both sides of the inequality above, we have

$$h_{j_{s_i}} > \theta_{j_{s_{i-1}}} - \theta_{j_{s_i}} + h_{j_{s_{i-1}}}.$$

But if $j_{s_i} < j_{s_{i-1}}$ then $\theta_{j_{s_{i-1}}} - \theta_{j_{s_i}} \geq 1$, so $\theta_{j_{s_{i-1}}} - \theta_{j_{s_i}} + h_{j_{s_{i-1}}} \geq 1 + h_{j_{s_{i-1}}}$. However, that can only hold if

$$h_{j_{s_i}} > 1 + h_{j_{s_{i-1}}}.$$

However, this contradicts the fact that both $h_{j_{s_i}}$ and $h_{j_{s_{i-1}}}$ belong to the unit interval.

Case 2a) $i > l$

Recalling our previous demonstrations, we wish to show that

$$\frac{P_{j_{t_{i-1}}-1}(i-1)}{p_{j_{t_{i-1}}}(i-1)} (\omega_i - \theta_{j_{t_{i-1}}}) > \frac{P_{j_{t_{i-1}}-1}(i)}{p_{j_{t_{i-1}}}(i)} (\omega_{i-1} - \theta_{j_{t_{i-1}}}).$$

We can reformulate this expression to

$$\frac{P_{j_{t_{i-1}}-1}(i)}{p_{j_{t_{i-1}}}(i)} > \left(\frac{P_{j_{t_{i-1}}-1}(i)}{p_{j_{t_{i-1}}}(i)} - \frac{P_{j_{t_{i-1}}-1}(i-1)}{p_{j_{t_{i-1}}}(i-1)} \right) (\omega_i - \theta_{j_{t_{i-1}}}). \quad (44)$$

From negative affiliation, the right-hand side of this inequality is positive. We can bound the expression on the right-hand side, noting that

$$\omega_i - \theta_{j_{t_{i-1}}} \leq \omega_i - 1 = i - 1.$$

So, it is true that

$$\frac{P_{j_{t_{i-1}}-1}(1)}{p_{j_{t_{i-1}}}(1)} + \left(\frac{P_{j_{t_{i-1}}-1}(i)}{p_{j_{t_{i-1}}}(i)} - \frac{P_{j_{t_{i-1}}-1}(i-1)}{p_{j_{t_{i-1}}}(i-1)} \right) (i-1) > \left(\frac{P_{j_{t_{i-1}}-1}(i)}{p_{j_{t_{i-1}}}(i)} - \frac{P_{j_{t_{i-1}}-1}(i-1)}{p_{j_{t_{i-1}}}(i-1)} \right) (\omega_i - \theta_{j_{t_{i-1}}}).$$

Thus, a sufficient condition for (44) is

$$\frac{P_{j_{t_{i-1}}-1}(i)}{p_{j_{t_{i-1}}}(i)} \geq \frac{P_{j_{t_{i-1}}-1}(1)}{p_{j_{t_{i-1}}}(1)} + \left(\frac{P_{j_{t_{i-1}}-1}(i)}{p_{j_{t_{i-1}}}(i)} - \frac{P_{j_{t_{i-1}}-1}(i-1)}{p_{j_{t_{i-1}}}(i-1)} \right) (i-1).$$

This condition is just an equivalent way of writing the definition of a concave function when the function is increasing. The remainder of the proof is unchanged. ■

REFERENCES

Armstrong, M., Multiproduct Nonlinear Pricing, *Econometrica* **64**, 1996, 51-76.

Beaudry, P. and C. Blackorby, Taxes and Employment Subsidies in Optimal Redistribution Programs, The Warwick Economics Research Paper Series, 2004, 766.

Besley, T. and S. Coate, 1995, The Design of Income Maintenance Programmes, *The Review of Economic Studies*, **62**, 187-221.

Bloomquist, N.S., 1981, A Comparison of Tax Bases for a Personal Tax, *The Scandinavian Journal of Economics*, **83**, 390-407.

Brett, C., 1998, Who Should Be on Welfare? The Use of Work Requirement as Part of an Optimal Tax Mix, *Oxford Economic Papers* **50**, 607-622.

Card, D. and P.K. Robins, 1996, Do Financial Incentive Encourage Welfare Recipients to Work?, Social Research and Demonstration Corporation, Ottawa, Canada.

Choné, P. and G. Laroque, 2005, Optimal Incentives for Labor Force Participation, *Journal of Public Economics* **89**, 395-425.

Choné, P. and G. Laroque, 2006, Should low skilled work be subsidized?, INSEE-CREST.

Cuff, K., 2000, Optimality of Welfare with Heterogeneous Preferences *Canadian Journal of Economics* **33**, 149-174.

Dasgupta, P., and P. Hammond, 1980, Fully Progressive Taxation, *Journal of Public Economics*, **13**, 141-154.

Diamond, P., 1980, Income Taxation with Fixed Hours of Work, *Journal of Public Economics* **13**, 101-110.

Diamond, P., 1997, Optimal Income Taxation: an Example with a U-Shaped Pattern of Optimal Marginal Tax Rates, *American Economic Review* **88**, 1998, 83-95.

Guesnerie, R., and K. Roberts, 1987, Minimum Wage Legislation as a Second Best Policy, *European Economic Review*, **31**, 491-8

Guesnerie, R., and J. Seade, 1982, Non Linear Pricing in a Finite Economy, *Journal of Public Economics* **17**, 146-157.

Jullien, B., 2000, Participation Constraints in Adverse Selection Models, *Journal of Economic Theory* **93**, 1-47.

Kanbur, R., M. Keen, and M. Tuomola, 1994, Optimal Nonlinear Income Taxation and the Alleviation of Income-Poverty, *European Economic Review*, **38**, 1613-1633.

Kesselman, J., 1973, A Comprehensive Approach to Income Maintenance: SWIFT, *Journal of Public Economics*, **2**, 59-88.

Kesselman, J., 1977, Tax Instruments and Economic Equality, *Canadian Journal of Economics*, **10**, 289-299.

Laroque, G., 2005, Income Maintenance and labor force participation, *Econometrica* **73**, 341-376.

Maderner, N., and J.-C. Rochet, 1995, Is it Legitimate to Encourage Work Sharing?, *Scandinavian Journal of Economics*, 621-633.

Marceau, N. and R. Boadway, 1994, Minimum Wage Legislation and Unemployment Insurance as Instruments for Redistribution, *Scandinavian Journal of Economics*, **96**, 67-81.

Maskin and Riley, 1985, "Input versus Output Incentive Schemes", *Journal of Public Economics*, 28, 1-23.

Matthews, S. and Moore, J. H., 1997, Optimal Provision of Quality and Warranties: An Exploration in the Theory of Multidimensional Screening, *Econometrica*, **55**, 441-467.

Mirrlees, J.A., 1971, An exploration of the Theory of Optimal Taxation, *The Review of Economic Studies*, **38**, 175-208.

Mirrlees, J.A., 1997, Information and Incentives: The Economics of Carrots and Sticks, *Economic Journal*, **107**, 1311-1329.

Rochet, J.-C. and Choné, P., 1998, Ironing, Sweeping, and Multidimensional Screening, *Econometrica*, **66**, 783-826.