Tagging and income taxation: theory and an application

Helmuth Cremer
University of Toulouse
(IDEI and GREMAQ)

Firouz Gahvari
Department of Economics
University of Illinois at Urbana-Champaign

Jean-Marie Lozachmeur
University of Toulouse
(IDEI and GREMAQ-CNRS)

April 2008

Abstract

This paper studies the properties of tagging in an optimal income tax framework à la Mirrlees. Assuming quasi-linear preferences, a Rawlsian social welfare function, and a constant elasticity of labor supply, we derive the following results: (i) The marginal income tax rates in the two tagged groups, bracket the marginal tax rate when all individuals are pooled together. (ii) If the skills distribution in one group first-order stochastically dominates the other, tagging calls for redistribution from the former to the latter group. (iii) If the distribution of skills is lognormal and the hazard rates in the two tagged groups do not cross, *every* individual in the group with lower average skills would benefit from tagging and has a higher utility level than his counterpart in the group with higher average skills; members of the latter group may lose, as well as gain, from tagging. (iv) Calibrating the model to the population of US prime age workers grouped into young (25–34 years) plus old (55–64) and middle-aged (35–54) workers, tagging increases the welfare of the least well-off individuals in both groups by the equivalent of \$249 per year. Moreover, workers in the young-plus-old group gain from \$585 per year for a worker at the lowest decile to \$3,264 per year at the highest decile. Workers in the middle-aged group lose from \$22 at the lowest decile to \$2,752 at the highest decile.

Key words: Tagging, optimal income taxation, hazard rate, age.

JEL Classification: H21, H23.

1 Introduction

A generally accepted tenet of optimal tax theory is that redistribution should be based solely on taxpayers' earning potentials; other inherent characteristics, such as one's race, height, age and the like, are not considered relevant for redistributive purposes. In a classic paper, Akerlof (1978) argued that even if one accepts that such characteristics are not, in and out of themselves, pertinent for redistribution, they still have a role to play in designing optimal tax schemes provided that they are correlated with earning potentials. He considered a model in which high- and low-ability workers could be grouped into two categories on the basis of an exogenously observable characteristic at no cost. One category consisted of low-ability types only and the other of both low- and high-ability types.[1] He showed that, within his setup, conditioning taxes on income and a "tag" indicating the taxpayer's category increases social welfare (assumed to be utilitarian). The solution makes individuals in the group consisting only of low-ability types better off relative to a tax scheme that depends on income alone. It also implies that two low-ability persons end up with different utility levels, with the person in the group consisting only of low-ability types enjoying a higher utility level. The basic insight is that tagging reduces the cost of income redistribution, although it does violate the principle of horizontal equity.

After thirty years, the literature has produced very few specific results on how tagging might change the properties of an optimal general income tax schedule. Nor do we know much beyond Akerlof's original findings on who gains or loses as a result of tagging. Stern (1982) assumes that skills are imperfectly observable and compared lump-sum taxes based on skills with an optimum general income tax (in a two-group model). He shows that income taxation becomes the preferable scheme as classification errors increase and as the society's aversion to inequality increases (making errors more

[1]The other characteristic of Akerlof's model is the discrete labor supply decision: one either has a hard job or an easy job, with a fixed disutility associated with the hard job.

costly to the society). Bennett (1987) studies the properties of a categorical tax system where two characteristics of taxpayers are observed imperfectly, and all taxpayers with the same two characteristics pay the same *lump-sum* tax. Viard (2001a, b) studies tagging in an optimal linear income tax framework allowing the demogrants to differ across groups but not the income tax rates.

Immonen *et al.* (1998) follow a different track concentrating on the marginal income tax rate properties with tagging, in a model with a continuum of individuals (as in Mirrlees' (1971) original formulation). They do not derive any analytical results. Instead, they perform a number of simulations for two groups with one group (the "rich") having a higher mean income than the other (the "poor"). They find that while the marginal income tax rate is decreasing in income in the rich group, it is increasing in the poor group. The intuition they offer is based on an earlier informal argument by Dilnot *et al.* (1984). Everyone in the poor group is treated more favorably, including the high-ability persons. The increasing marginal income tax rate in the poor group is then meant to offset the favorable treatment of the high-ability persons in that group.

Hamilton and Pestieau (2004), while not framed in terms of tagging, is a related work. It studies the comparative statics of type proportions (high- and low-ability) in the context of a general income tax framework. They show that, assuming preferences that are quasi-linear in leisure and a social welfare function that is Rawlsian, the utility of low-ability persons is increasing, and the utility of high-ability persons is decreasing, in the proportion of high-ability persons in the population (when both individual types work).

Mankiew and Weinzierl (2007) study a model with many skill types who can be tagged on the basis of height. They too do not have much by way of analytical results (except for the traditional no distortion at the top and equalization of *average* social utility of income across different height groups). Using data from the National Longitudinal Survey of Youth, they show that earnings are positively correlated to heights.

This then calls for a higher *average* tax rate for the taller people within a utilitarian framework.[2]

Alesina *et al.* (2007) advocate tagging based on gender. However, their results rely on a different argument. Tagging is not due to any differences in the skills distribution between the sexes. Rather, it is motivated by their different behavioral responses to taxation. There are also papers that use the idea of tagging but not in conjunction with a general income tax. Parsons (1996) discusses the optimal benefit structure of an earnings insurance program when "eligibility requirements" are used as a tag to (imperfectly) identify those who are out of work. The result he finds is similar to Akerlof in that both able and disable individuals in the eligible group get a preferential treatment.

The closest papers to ours is Boadway and Pestieau (2006), analytically, and Kremer (2001), empirically. Boadway and Pestieau study a model with two ability types where the population can be separated into two groups, one of which has a higher proportion of low-ability individuals than the other (as opposed to Akerlof's formulation where one group consists only of the low-ability persons). In this way, they allow for a "two-sided error" where not only does a low-ability person in the not-favored group get a less favorable treatment than his counterpart in the favored group, but also that a high-ability person in the favored group gets a more favorable treatment than his counterpart in the not-favored group.

Boadway and Pestieau (2006) prove, assuming quasi-linear preferences, that tagging entails redistribution from the group with a higher proportion of high-ability persons to the group with a higher proportion of low-ability persons,[3] improves the social welfare within each group, and makes the high-ability individuals in the not-favored group—the group with a higher proportion of high-ability persons—worse off. They interpret this

[2] No categorical results are derived for *marginal* income tax rates.

[3] The result is derived for a Rawlsian social welfare function and, more generally, for any social welfare function that exhibits a constant absolute aversion to inequality.

last result as to imply that the tax system is more progressive in the tagged group with the highest proportion of high-ability types.[4] However, it is somewhat of a stretch to infer progressivity from a model with two groups of people.

A distinctive feature of our paper is that it derives a set of analytical results for a model with a continuum of individuals who can be divided into two groups with different ability distributions over the same support. The tagging is based on a publicly and costlessly observable exogenous characteristic. Assuming quasi-linear preferences, a Rawlsian social welfare function, and a constant and identical elasticity of labor supply within and across the tagged groups, we show that the marginal income tax rates in the two tagged groups bracket the marginal tax rate when all individuals are pooled together. This is the case at *all skill levels.* Secondly, we show that if the skills distribution in one group first-order stochastically dominates the other, tagging calls for redistribution from the former to the latter group. Third, we derive the solutions for income, consumption, and utility, for all individual types with and without tagging. Fourthly, assuming that distribution of skills is lognormal and the hazard rates in the two tagged groups do not cross, we show that *every* individual in the group with lower average skills would benefit from tagging (in terms of consumption and utility levels and compared to the pooled equilibrium solution). They will also consume more and have higher utility levels than their counterparts in the group with higher average skills. Members of the latter group may lose, as well as gain, from tagging.

Kremer (2001), unlike Boadway and Pestieau (2006), does not attempt to derive any analytical results. Based on earnings data from the Current Populations Surveys, it investigates the difference in earnings hazard rates between the very young workers (17–

[4] Boadway and Pestieau (2004) also assert that "Intuition suggests that if the density of ability distributions across pairs of groups crossed only once, and if all ability-types were in all groups, the analog of the above results should apply [to a case with n ability types (as opposed to two)]. Social welfare within each group should be increased by inter-group transfers and group-specific tax structures, and tax schedules should be more progressive in groups that have more skill intensive ability distributions" (p.16).

25) and the prime-age workers (31–64), with an eye to its implication for the optimal income tax rates in the two groups. He finds that the very young have much larger hazard rates. He also finds that these workers have more elastic labor supplies. Both of these factors suggest that the optimal marginal tax rate for the very young is smaller than for prime-age workers (higher hazard rates and higher elasticities imply higher excess burden).[5]

Tagging on the basis of age has the advantage of not violating the principle of horizontal equity as long as one adopts a lifetime perspective.[6] We thus also use age for the *application* of our theoretical model. However, in line with our analytical results, we concentrate on the implication of differences between *average* skills between groups for tagging.[7] Kremer (2001) concentrates on the very young workers who are not just *on average* of lower skills; they, almost, *all* have lower skills than prime-age workers. This makes them an easy group to target. Indeed, if all workers of a certain age were known to have a particular skill level, one could give them a non-distortionary tax and apply the income tax only to people that do not include this age group. For this reason, as well as the difference in the labor supply responses between the very young and the prime-age workers, we leave the very young workers altogether out of our simulations. Instead, we concentrate on tagging within the prime-age workers.

The prime age workers have typically a higher earning potential in their middle ages. This arises because workers gain experience over their working years. Moreover, being in their middle ages, they are not as yet afflicted by vagaries of the old age. The younger

[5] Kremer (2001) makes the same observations for the older workers between 65 to 70 years of age.

[6] Banks *et al.* (2008) discuss the issue of what can serve as an appropriate tag at length. They write that "Thus assertions about observability are not an adequate guide to the choice of a tax base for direct taxation. Complexity matters as well, as does a plausible sense of both political economy and public reactions." They also add "In contrast to height, age is used by actual tax structures, but very little apart from retirement-related rules. In the US there are distinctions for children (who can be dependents and so provide additional deductions) and those over 65, who may receive an additional deduction."

[7] Admittedly though the differences in mean incomes translate into differences in population responses to tax rates feeding into our numerical results.

workers suffer from a lack of experience while the older workers suffer from physical problems that develop with age. This observation is supported by the data. The US Census Bureau reports median incomes of \$45,797 for workers between the ages of 25 to 34 years, \$54,168 for workers between the ages of 35 to 44 years, \$58,968 for workers between the ages of 45 to 54 years, and \$46,593 for workers between the ages of 55 to 64 years of age in 2001. We thus divide the prime-age workers into two groups: The "less well-off" workers comprising age groups of 25–34 and 55–64, and the "well-off" workers consisting of those in the 35–54 years of age. We shall refer to the former as the "inexperienced/infirm" group, and to the latter as the "experienced" group.[8]

We calibrate our model to the population of US prime age workers such that it reflects the statistical facts mentioned above, while also assuming that all workers have a constant wage elasticity of labor supply equal to 0.5. We additionally assume that the distribution of skills is truncated lognormal. We find that tagging increases the welfare of the least well-off individuals in both groups of workers by \$249 per year. Moreover, workers in the inexperienced/infirm group gain anywhere from \$585 per year at the lowest decile to \$3,264 per year at the highest decile. Workers in the experienced group lose from \$22 at the lowest decile to \$2,752 at the highest decile. The comparison of gains and losses suggests that tagging by age can enhance the efficiency of the tax system considerably. Concerning tax rates, we derive, for all income brackets between \$10,000 and \$200,000, the values of the optimal marginal and average income tax rates for the two tagged groups as well as for the pooled population. The marginal tax rates are always decreasing, and the average tax rates decreasing, in income. The marginal rates vary between 80 percent to zero while the average rates remain always below 20%.

[8] We use these terms for labeling only; no other meaning—positive or negative—is implied. Indeed, judged by this labeling, one of the authors of this paper is infirm, one experienced, and one inexperienced!

2 The model

The economy is inhabited by a continuum of individuals who differ on the basis of one immutable and publicly observable characteristic, or "tag", such as age, as well as in skill levels. The average level of skills in one group is higher than in the other. We shall refer to the group with the higher average skills as group e, and to the other group as group i. Labor is the only factor of production and the aggregate production level is linear. The wage rates, which also represent the skill levels, are distributed according to the distribution function $F_e(w)$ in group e, and $F_i(w)$ in group i. The support of the two distributions are the same and given by $[\underline{w}, \overline{w}]$. The corresponding density functions, $f_j(w) = F_j'(w)$, $j = e, i$, are assumed to be strictly positive and differentiable for all $w \in [\underline{w}, \overline{w}]$. Individuals have identical preferences that depend on consumption, c, positively, and on labor supply, L, negatively. Preferences are represented by the quasi-linear utility function

$$u = c - h(L). \tag{1}$$

The social welfare criterion is Rawlsian (maxi-min); it is being implemented through a purely redistributive income tax system. The quasi-linearity of preferences and the maxi-min criterion make the problem analytically tractable. They also imply that the government is in effect maximizing the transfers to the least well-off individuals while keeping the excess burden of taxation at a minimum. Nothing else matters. We will further elaborate on these points later.

To understand how tagging may improve the redistributive power of the tax system, and how it affects the properties of the income tax structure, we derive the optimal income tax structure once ignoring the tag and a second time conditioned on it. In deriving the latter, we find it useful to follow a two-stage procedure, as in Immonen *et al.* (1998). In the first stage, one derives the optimal income tax structure for the two groups separately. This requires the government to achieve its objective function for each group separately, while assigning each its own separate tax revenue constraint. In the second stage, one determines the size of the surplus, or deficit, for each group so that the overall government's budget constraint is satisfied for the two groups taken together while simultaneously ensuring that the social welfare objective is attained for the entire population.

The two-stage procedure developed below is particularly simple, given our Rawlsian objective function and quasi-linear preferences. The two assumptions together imply that the size of the government's external revenue requirement enters the society's optimal income tax function additively. Any increase or decrease in the revenue requirement is met by a lump-sum tax, or rebate, levied *uniformly* on all taxpayers; no other aspects of the tax function is affected. As a consequence, the determination of the revenue requirements in the second stage, would leave the properties of the optimal tax schedule derived in the first stage intact.

2.1 The generic optimal income tax problem

Consider the determination of the optimal income tax structure for a given population conditioned only on income.[9] Let $c(w)$, $I(w) = wL(w)$, and $u(w)$ denote consumption, income, and utility of an individual of type w. Denote the government's revenue requirement by $\overline{R}$ and the net taxes collected from the population by

$$R \equiv \int_{\underline{w}}^{\overline{w}} [I(w) - c(w)]\, f(w) dw. \tag{2}$$

The optimal tax problem, given quasi-linear preferences and a Rawlsian objective function, is to derive $c(w)$, $I(w)$ in order to maximize the utility of the poorest individual when preferences are given by (1). The optimization is subject to the government's

[9] The material in this subsection follows that in Boadway and Jacquet (2008) who study optimal taxation under the twin assumptions of quasi-linearity and a maxi-min criterion (but without tagging) in some detail.

budget constraint $R \geqq \overline{R}$ and the local incentive compatibility constraint

$$\frac{du}{dw} = \frac{I(w)}{w^2} h' \left(\frac{I(w)}{w} \right). \tag{3}$$

Integrating (3), we have

$$u(w) = \underline{u} + \int_{\underline{w}}^{w} \frac{I(s)}{s^2} h' \left(\frac{I(s)}{s} \right) ds, \tag{4}$$

where $\underline{u} = u(\underline{w})$ is the utility of the poorest individual. Observe that the second term in the right-hand side of (4) is the informational rent that one has to leave for an individual with wage $w > \underline{w}$ to reveal his type. Given $u(w)$, we can determine $c(w)$ from equation (1) as

$$c(w) = u(w) + h \left(\frac{I(w)}{w} \right). \tag{5}$$

Substituting for $c(w)$ from (5) into (2), we find, after some algebraic manipulations shown in the Appendix, that the expression for R is simplified to

$$R = -\underline{u} + \Psi(F), \tag{6}$$

where

$$\begin{aligned} \Psi(F) \quad \equiv \quad & \int_{\underline{w}}^{\overline{w}} \left[I(w) - h \left(\frac{I(w)}{w} \right) \right] f(w) dw - \\ & \int_{\underline{w}}^{\overline{w}} \frac{I(w)}{w^2} h' \left(\frac{I(w)}{w} \right) [1 - F(w)] dw. \end{aligned} \tag{7}$$

Observe that expression (7) measures the maximum tax revenue the government can, in the second best, extract from a population with the distribution function $F(\cdot)$. The first term is the total surplus; that is, sum of all individuals' utilities if they pay no taxes to the government. The second term is the sum of informational rents noted above (aggregated over all individuals with wages $w > \underline{w}$). Note that $\Psi(F)$ depends solely on individuals' preferences and the distribution of productivities. This follows because with the maxi-min criterion, the government is indifferent between the utilities of all individuals with $w > \underline{w}$, regardless of the size of their w.

With $R = \overline{R}$, then, the optimal income tax problem is reduced to that of the choice of $I(w)$ to maximize

$$\underline{u}\left(\overline{R}; F\right) = -\overline{R} + \Psi(F), \tag{8}$$

or, equivalently, to maximize $\Psi(F)$. This is thus the same as minimizing the "aggregate excess burden" of the tax system. One should not be surprised by this in light of our previous observation that our maxi-min social welfare criterion assigns a weight of zero to the utility of all individuals with $w > \underline{w}$. Thus differentiate (7) with respect to $I(w)$ and set the resulting equation equal to zero. This yields the first-order condition

$$\left[1 - \frac{1}{w} h'\left(\frac{I(w)}{w}\right)\right] f(w) - \frac{1 - F(w)}{w^2}\left[h'\left(\frac{I(w)}{w}\right) + \frac{I(w)}{w} h''\left(\frac{I(w)}{w}\right)\right] = 0.$$

Rearranging the above equation, and denoting the optimized value of a variable with a "star" over it, we find $I^*(w; F)$ as the "solution" to

$$\frac{1 - \frac{1}{w} h'\left(I(w)/w\right)}{\frac{1}{w} h'\left(I(w)/w\right)} = \left[\frac{1 - F(w)}{w f(w)}\right]\left[1 + \frac{I(w)}{w}\frac{h''\left(I(w)/w\right)}{h'\left(I(w)/w\right)}\right]. \tag{9}$$

Writing $\Psi^*(F)$ for the maximized value of $\Psi(F)$, from (8) we write the expression for $\underline{u}\left(\overline{R}; F\right)$, optimized over $I(w)$, as

$$\underline{u}^*\left(\overline{R}, F\right) = -\overline{R} + \Psi^*(F). \tag{10}$$

Then, from (4) and (5),

$$u^*\left(w; \overline{R}, F\right) = \underline{u}^*\left(\overline{R}, F\right) + \int_{\underline{w}}^{w} \frac{I^*(s; F)(s)}{s^2} h'\left(\frac{I^*(s; F)}{s}\right) ds, \tag{11}$$

$$c^*\left(w; \overline{R}, F\right) = u^*\left(w; \overline{R}, F\right) + h\left(\frac{I^*(w; F)}{w}\right). \tag{12}$$

Equations (11)–(12) tell us that once the optimal value of $I(w)$—which is independent of the value of $\overline{R}$—is chosen, a one unit increase (or decrease) in $\overline{R}$ lowers (or increases) the optimal values of $\underline{u}\left(\overline{R}, F\right)$, $u\left(w; \overline{R}, F\right)$, and $c\left(w; \overline{R}, F\right)$ by one unit.

2.2 The implementing income tax function

Introduce $T\left(I^*(w;F)\right)$ to denote the implementing tax function. We assume $T(\cdot)$, whose shape we want to determine, to be differentiable. Consider the behavior of an individual of type w who maximizes $u = c - h\left(I/w\right)$ subject to the budget constraint $c = I - T\left(I\right)$. The individual's first-order condition for this problem yields: $1 - T'\left(I\right) - h'\left(I/w\right)/w = 0$. Consequently, the implementing tax function is characterized by its marginal income tax rate and determined according to

$$T'\left(I^*(w;F)\right) = 1 - \frac{h'\left(I^*(w;F)/w\right)}{w}, \tag{13}$$

where $I^*(w;F)$ satisfies equation (9). To determine the properties of $T'(\cdot)$, rearrange (13) and rewrite it as

$$\begin{aligned}
\frac{T'}{1-T'} &= \frac{1 - \frac{h'(I^*(w;F)/w)}{w}}{\frac{h'(I^*(w;F)/w)}{w}} \\
&= \left[1 + \frac{L^*(w;F)h''(L^*(w;F))}{h'(L^*(w;F))}\right]\left[\frac{1-F(w)}{wf(w)}\right],
\end{aligned} \tag{14}$$

where the second equality in (14) follows from (9), and we have substituted $L^*(w;F)$ for $I^*(w;F)/w$ and dropped the argument of $T'(\cdot)$. To get an intuition for this formula, assume the government imposes a tax on the marginal income of people with ability w. This tax affects only the people with ability levels w and above. There are thus $f(w)$ persons with the aggregate earning potential of $wf(w)$, whose incentives are affected. Now observe that $L^*(\cdot)\ h''(\cdot)/\ h'(\cdot)$ in (14) is equal to the inverse of the wage elasticity of labor supply for an individual of type w. Thus when $wf(w)$ is multiplied by the first bracketed expression on the right-hand side of (14), it gives us a measure of the incentive costs of the tax. As to the benefits, the marginal income tax on w generates a revenue from all people with ability levels above w and does it with no incentive cost because the tax is infra-marginal for them. The optimal marginal income tax strikes a balance between the costs and benefits of taxation as measured by its revenue. Note that our Rawlsian specification for the social welfare function makes this calculus a simple exercise by not counting the utility loss of taxes that the people above w experience.

Observe that with $h'(\cdot) > 0$ and $h''(\cdot) > 0$, equation (14) implies that $T'(\cdot) > 0$ as long as $F(w) < 1$, and $T'(\cdot) = 0$ at $w = \overline{w}$. Now define

$$H(w; F) \equiv 1 + \frac{L^*(w; F)h''(L^*(w; F))}{h'(L^*(w; F))}, \tag{15}$$

$$\Omega(w) \equiv \frac{1 - F(w)}{wf(w)}. \tag{16}$$

We can then rewrite equation (14) as

$$\frac{T'}{1 - T'} = H(w; F) \times \Omega(w). \tag{17}$$

Observe that in (17), taxpayers' preferences enter only through the $H(w; F)$ term. Equation (17) thus tells us the more elastic is the labor supply of the w-type, the lower must be the marginal income tax rate they face. On the other hand, $\Omega(w)$ depends solely on the properties of the distribution function, $F(w)$. Specifically, $\Omega(w)$ is equal to the inverse of the hazard rate (the Mills' ratio) divided by w. With the hazard rate here indicating the proportion of type w individuals in the population of people with wages w and above, equation (17) indicates that the higher is the proportion the lower must be the marginal tax rate. Assuming a constant wage elasticity, the marginal income tax rate is determined entirely by the shape of the hazard rate and the size of the wage rate. Pareto distribution has an increasing hazard rate leading to a constant marginal income tax rate (an increasing hazard rate and an increasing w result in a constant $\Omega(w)$). The hyperbolic function, $f(w) = \lambda e^{-\lambda w}$, has a constant hazard rate resulting in a marginal income tax rate that is always decreasing in w ($\Omega(w)$ is decreasing because w is increasing). The hazard rate for lognormal distribution, has an initially increasing segment followed by a decreasing part; see Jorgenson *et al.* (1967). The corresponding marginal income tax rate must then be initially decreasing in w, but it may eventually become increasing. Salanié (2003) notes that, given quasi-linearity and the Rawlsian social welfare function, the second-order condition for the optimal income tax problem is satisfied if and only if $\Omega(w)$ is nonincreasing in w (p 95).

3 Optimal taxation with and without tagging

We start with the characterization of optimal tax solution when the information on experience (age) is discarded. Then, we proceed to examine how conditioning the tax on experience, changes the properties of the optimal tax solution. Our emphasis in this section is simply to find out the implications of the added information (i.e., the knowledge of the distributions function for the skills in the two groups as opposed to knowing only the distribution function for the skills in the entire population). We are not interested here in exploring the nature of the added information (i.e., the fact that there is a positive correlation between being of low-ability and belonging to group i), and its implications for how different people in different groups fare as a result of tagging. We will discuss the latter in the following section.

3.1 Optimal solution without tagging

Denote the distribution function for the entire population by $F_{ei}(w)$ and its corresponding density by $f_{ei}(w)$. The optimal tax structure here is precisely the same as the one derived under the generic optimal tax problem, with $F_{ei}(w)$ and $f_{ei}(w)$ replacing $F(\cdot)$ and $f(\cdot)$. Denote the optimal solution without tagging by superscript N, so that $\underline{u}^N = \underline{u}^*(0; F_{ei}), u^N(w) = u^*(w; 0, F_{ei})$, and $c^N(w) = c^*(w; 0, F_{ei})$. To be consistent, we denote the optimal values of $I(w)$ and $\Psi(F_{ei})$, which are independent of $\overline{R}$, by $I^N(w) = I^*(w; F_{ei})$ and $\Psi^*(F_{ei})$. The optimal solutions are thus characterized by

$$\underline{u}^N = \Psi^*(F_{ei}), \tag{18}$$

$$u^N(w) = \underline{u}^N + \int_{\underline{w}}^{w} \frac{I^N(s)}{s^2} h'\left(\frac{I^N(s)}{s}\right) ds, \tag{19}$$

$$c^N(w) = u^N(w) + h\left(\frac{I^N(w)}{w}\right). \tag{20}$$

Finally, denoting $H(w; F_{ei})$ by $H^N(w)$ and using equation (17), we can write the marginal income tax rate as

$$\frac{T'^N}{1-T'^N} = H^N(w) \times \Omega_{ei}(w). \tag{21}$$

3.2 Optimal solution under tagging

Given the Rawlsian social welfare function, we continue to be concerned with the maximization of the welfare of the least well-off individuals in the entire population. With tagging, we can write this as maximization of $\min[\underline{u}_e, \underline{u}_i]$ As observed earlier, it is simpler to derive the tagged solution in a two-stage manner. The first stage corresponds to the generic optimal income tax problem of Subsection 2.1 wherein $I_j(w), c_j(w)$ are derived to maximize $\underline{u}_j$ under the constraint that $R_j \geqq \overline{R}_j$, $j = e, i$, treating $\overline{R}_j$ as fixed. This yields $I_j^*(w) = I^*(w; F_j)$ with a maximal $\underline{u}_j$ equal to

$$\underline{u}_j^*\left(\overline{R}_j\right) \equiv \underline{u}^*\left(\overline{R}_j; F_j\right) = -\overline{R}_j + \Psi^*(F_j).$$

In the second stage, we choose $\overline{R}_e, \overline{R}_i$ to maximize $\min[\underline{u}_e^*\left(\overline{R}_e\right), \underline{u}_i^*\left(\overline{R}_i\right)]$ subject to $\overline{R}_e + \overline{R}_i = 0$. This is of course equivalent to wanting to equalize $\underline{u}_e^*\left(\overline{R}_e\right)$ and $\underline{u}_i^*\left(\overline{R}_i\right)$. Using (10), $\overline{R}_e, \overline{R}_i$ are then found as the solution to

$$\begin{aligned} -\overline{R}_e + \Psi^*(F_e) &= -\overline{R}_i + \Psi^*(F_i), \\ \overline{R}_e + \overline{R}_i &= 0. \end{aligned}$$

Denoting the optimal solution with tagging by superscript G, We have, from above,

$$\overline{R}_e^G = \frac{\Psi^*(F_e) - \Psi^*(F_i)}{2}, \tag{22}$$

$$\overline{R}_i^G = \frac{\Psi^*(F_i) - \Psi^*(F_e)}{2}. \tag{23}$$

These in turn imply, using (10) again,

$$\underline{u}^G = \frac{\Psi^*(F_e) + \Psi^*(F_i)}{2}. \tag{24}$$

Additionally, from equations (11)–(12), we have

$$u_j^G(w) = \underline{u}^G + \int_{\underline{w}}^{w} \frac{I_j^*(s)}{s^2} h'\left(\frac{I_j^*(s)}{s}\right) ds, \tag{25}$$

$$c_j^G(w) = u_j^G(w) + h\left(\frac{I_j^*(w)}{w}\right). \tag{26}$$

Finally, denoting $H(w; F_j)$ by $H_j^G(w)$ and using equation (17), we can write the marginal income tax rate for an individual of type w, when conditioned on experience, $j = e, i$, as

$$\frac{T_j'^G}{1 - T_j'^G} = H_j^G(w) \times \Omega_j(w). \tag{27}$$

The procedure described above highlights a number of interesting features of the optimal tax problem under tagging. First, the interaction between groups occurs through the transfers only. Second, the marginal tax rate for any given individual depends only on the characteristics of the group to which he belongs (specifically on the elasticity of the labor supply and hazard rate).[10] Third, *the profile of the marginal tax rates* within any given group, does not depend on the characteristics of the second group with which it is combined. Specifically, it does not matter whether this second group is "richer" or "poorer". Observe, however, that this property applies only to the marginal tax rates and *not* the average tax rates. Finally, if one of the tagged groups has a degenerate skills distribution with a single wage level, as in the ordinal Akerlof's (1978) example, this group should be subjected to a lump sum tax and face a zero marginal tax rate (regardless of how high or low the wage level is).

3.3 Tagged versus non-tagged solutions

We now turn to the question of the comparison between the two types of solutions. To this end, we first present a lemma, stating the relationship between the hazard rates of the two groups and the entire population.

[10] With our quasi-linear specification, external revenue requirements do not affect the values of the marginal tax rates. With more general preferences, this will not be the case. Thus marginal tax rates may depend on the presence of other groups, but only through the revenue requirement.

Lemma 1 *Assume there are two groups of individuals, e and i, whose populations are of the same size, and whose skill levels are distributed over the same support $[\underline{w}, \overline{w}]$ according to the distribution functions $F_e(w)$ and $F_i(w)$. Then, the inverse of the hazard rate for an individual of type w in the entire population is bracketed by the inverse of the individual's hazard rates in the two groups.*

Proof. Given our assumptions, the following relationships hold between the distribution functions, and the density functions, for the entire population and the two groups that comprise it:

$$F_{ei}(w) = \frac{F_e(w) + F_i(w)}{2}, \tag{28}$$

$$f_{ei}(w) = \frac{f_e(w) + f_i(w)}{2}. \tag{29}$$

It then follows from these relationships and the definition of $\Omega(w)$ that

$$\begin{aligned} \Omega_{ei}(w) &= \frac{1 - F_{ei}(w)}{w f_{ei}(w)} = \frac{1 - [F_e(w) + F_i(w)]/2}{w\,[f_e(w) + f_i(w)]/2} \\ &= \frac{f_e(w)}{f_e(w) + f_i(w)}\Omega_e(w) + \frac{f_i(w)}{f_e(w) + f_i(w)}\Omega_i(w). \end{aligned} \tag{30}$$

■

Now if, for any given w, the elasticity of the labor supply does not vary across the tagged groups, then it will follow from (30) that the marginal income tax rate in the pooled equilibrium solution is a weighted average of the marginal income tax rates in the two tagged groups. Specifically, the following relationship holds between T_e', T_i', and T_{ei}':

$$\frac{T_{ei}'}{1 - T_{ei}'} = \frac{f_e(w)}{f_e(w) + f_i(w)} \frac{T_e'}{1 - T_e'} + \frac{f_i(w)}{f_e(w) + f_i(w)} \frac{T_i'}{1 - T_i'}. \tag{31}$$

That is, the marginal income tax rate in the pooled equilibrium case is a weighted average of the marginal tax rates for the two tagged groups and thus bracketed by them.

To go beyond this, and to derive closed-form solutions, we make one other assumption requiring labor supply to exhibit a constant wage elasticity. Thus set $h(L) = L^{1+1/\varepsilon}$, where ε is the labor supply elasticity.[11] Observe that the strict convexity of $h(L)$ implies $\varepsilon > 0$. One can then immediately derive a closed-form solution for optimal incomes. We show in the Appendix that $I_j^*(w)$, $j = e, i, ei$, is equal to

$$I_j^*(w) = w^{1+\varepsilon}\left(\frac{\varepsilon}{1+\varepsilon}\right)^{\varepsilon}\left[1+\left(1+\frac{1}{\varepsilon}\right)\Omega_j(w)\right]^{-\varepsilon}. \tag{32}$$

Then, given that $\Omega_{ei}(w)$ is a weighted average of $\Omega_e(w)$ and $\Omega_i(w)$, equation (32) tells us that $I_{ei}^*(w)$ always lies between $I_e^*(w)$ and $I_i^*(w)$. Moreover, $I_e^*(w) \gtreqless I_i^*(w) \Leftrightarrow \Omega_e(w) \lesseqgtr \Omega_i(w)$. That is, $I_e^*(w) \gtreqless I_i^*(w)$ if and only if the hazard rate for the w-type in group e is greater than/equal to/smaller than the hazard rate for the w-type in group i. This makes perfect sense. With the hazard rate indicating the "proportion" of individuals of type w in the population of people with wages w and above, a higher hazard rate implies that we want to assign the person a higher income level. The following proposition summarizes these results.

Proposition 1 *Assume preferences are quasi-linear, the elasticity of labor supply for a given w-type is the same in the two groups, and the social welfare function is Rawlsian. There are two groups of individuals of equal size, e and i, each with a continuum of skills distributed over the same support. Experience is perfectly represented by age and publicly observable. Then:*

(i) The optimal marginal income tax rates faced by the w-type individuals in the two groups under the tagging solution brackets the optimal marginal income tax rate faced by the w-type individuals under the no-tagging solution.

(ii) Assuming a constant labor supply elasticity within and across groups, the income of w-type workers in e and i groups under the tagged solution bracket the income that they will earn if the entire population is pooled.

[11] There is no need to assign a coefficient to $L^{1+1/\varepsilon}$ as one can always adjust the unit of measurement for c.

For future reference, observe that with a constant elasticity of labor supply, from (17),the marginal income tax rates are simplified to

$$\frac{T'_j}{1-T'_j} = \left(1+\frac{1}{\varepsilon}\right)\Omega_j(w), \quad j = e, i, ei. \tag{33}$$

4 Welfare and redistribution

Policy discussions often center around the welfare of a particular group considered—"tagged"—for special treatment (i.e. given categorical assistance). The favored group is the one referred to generally as the "tagged group," although both groups are effectively tagged. For this reason, and to avoid confusion with the case that no tagging is done, we do not speak of tagged and untagged groups. Instead, we continue to use the favored and not-favored labeling. Of course, to the extent that both groups contain less able and more able individuals, one wants to know what happens to the welfare of less able persons, and the more able groups, in the favored and not-favored groups. There are in fact three interrelated welfare questions. One is the effect of tagging on social welfare in general; the second concerns redistribution across types within a group, and the third is that of redistribution across groups. Note that the latter two are inter-related. In particular, redistribution across groups would also affect redistribution within a group. Note also that redistribution from the not-favored to the favored group is essentially a redistribution from the total population to the favored group.

The answer to the first question is a trivially obvious one. It is plain that one can always offer the no-tagging solution to the two groups. That is, $I^N(w)$ and the corresponding, $\underline{u}^N, u^N(w), c^N(w)$, constitute a feasible allocation to the optimal tax problem with tagging. Consequently, if the procedure outlined above yields a solution differently from $I^N(w)$, it must be a better allocation (i.e., $\underline{u}^G > \underline{u}^N$). Observe also that the first-stage problems yield a solution $I_j^*(w) = I^N(w)$ if, and generally only if, $F_e(w) = F_i(w)$. Put differently, when the two groups have different distribution of skills, one can always increase social welfare by offering the two groups different tax schedules.

The answers to the second and third questions are more involved. Given that the Rawlsian social welfare is represented by the utility of the least well-off persons of the society as a whole, tagging in our setup redistributes to the least able individuals in both groups, favored as well as the not-favored group, leaving them equally well off.[12] However, one cannot a priori determine what happens to the welfare of the more able individuals in either group. All one can say at this stage is that tagging improves aggregate output. That is, we have

$$\int_{\underline{w}}^{\overline{w}} I_e^*(w)\ f_e(w)dw + \int_{\underline{w}}^{\overline{w}} I_i^*(w)\ f_i(w)dw \geqq 2\int_{\underline{w}}^{\overline{w}} I_{ei}^*(w)\ f_{ei}(w)dw. \tag{34}$$

This result follows from the definition of $\Omega_j(w)$, $j = e, i, ei$, and the relationship between $F_e(w), F_i(w)$, and $F_{ei}(w)$; see the Appendix.

To shed more light on the redistribution questions, we have to know more about the skills distributions beyond the fact that they are different. For example, tagging is often rationalized because one group happens to contain relatively more disadvantaged people whom the government wants to help. Akerlof's (1978) original example was based on the assumption that one could tag a particular group which consisted of only low-ability individuals. He was then able to show that the low-ability persons in the tagged group receive more transfers and are better off under the tagged solution as compared to the pooled equilibrium (untagged) solution. One could not determine the impact that tagging would have on welfare of the low-ability persons in the not-favored group. As noted earlier, this ambiguity does not arise with a Rawlsian social welfare function as the least able individuals would get the same treatment in the favored and not-favored groups.

With the tagged group in Akerlof (1978) consisting of only the low-ability persons, it also follows from his result that there is a redistribution from the not-favored group (or

[12] Of course, with other types of social welfare function, say utilitarian, the utility of less able individuals in the tagged and untagged groups will be different.

the population as a whole) to the favored group. A generalization of this result is that tagging makes the favored group better off if the group contains a higher proportion of low ability persons than the not-favored group. This is what we prove in Proposition 2, while recasting the concept of containing a higher proportion of low-ability individuals in terms of stochastic dominance.

Proposition 2 *Assume preferences are quasi-linear and the social welfare function is Rawlsian. There are two groups of individuals of equal size, e and i, each with a continuum of skills distributed according to the distribution functions $F_e(w)$ and $F_i(w)$ over the support $w \in [\underline{w}, \overline{w}]$. Group e first-order stochastically dominates group i so that $F_e(w) \leq F_i(w)$ for all $w \in [\underline{w}, \overline{w}]$ and $F_e(w) < F_i(w)$ for some w. Then, in the second best, the government can extract more tax revenues from group e than from group i. That is, $\Psi^*(F_e) > \Psi^*(F_i)$ so that $\overline{R}_e^G > 0$ and $\overline{R}_i^G < 0$.*

Proof. First, observe that incentive compatibility does not depend on the distribution of types so that the allocation $(I_i^*(w), c_i^G(w))$ is incentive compatible in e as well as in i. Moreover, one can choose $\overline{R}_i$ such that assigning $(I_i^*(w), c_i^G(w))$ to the individuals in e results in the same value for $\underline{u}$ as assigning $(I_e^*(w), c_e^G(w))$ to them; namely $\underline{u}_e^*\left(\overline{R}_e^G\right)$. Now, given that $(I_e^*(w), c_e^G(w))$ is the optimal solution in e, we must have

$$\int_{\underline{w}}^{\overline{w}} \left[I_e^*(w) - c_e^G(w)\right] dF_e(w) \geqq \int_{\underline{w}}^{\overline{w}} \left[I_i^*(w) - c_i^G(w)\right] dF_e(w). \tag{35}$$

Next, denote the taxes paid by a person of type w by $t(w) \equiv T\left(I(w)\right)$. Differentiating $t(w)$ with respect to w yields

$$\frac{dt(w)}{dw} = T'(I)\frac{dI(w)}{dw} \geqq 0,$$

where the sign follows from the signs of $T'(I)$ and $dI(w)/dw$ in Mirrlees' optimal tax problem. With $t(w)$ being non-decreasing in w, it follows from the definition of first-order stochastic dominance that

$$\int_{\underline{w}}^{\overline{w}} \left[I_i^*(w) - c_i^G(w)\right] dF_e(w) > \int_{\underline{w}}^{\overline{w}} \left[I_i^*(w) - c_i^G(w)\right] dF_i(w). \tag{36}$$

Inequalities (35)–(36) then imply that

$$\int_{\underline{w}}^{\overline{w}} \left[I_e^*(w) - c_e^G(w)\right] dF_e(w) > \int_{\underline{w}}^{\overline{w}} \left[I_i^*(w) - c_i^G(w)\right] dF_i(w). \tag{37}$$

■

Proposition 2 thus shows that tagging entails redistribution from group e to group i. Observe that the key property which allows this proof to go through is the independence of the optimal allocations $\left(I_e^*(w), c_e^G(w)\right)$ and $\left(I_i^*(w), c_i^G(w)\right)$ from revenue constraints. In the absence of this property, inequality (36) may not hold. That it holds here is due to the assumptions of quasi-linearity of preferences and the maxi-min criterion.

To make further deductions about welfare of individuals beyond the poorest, one has to make assumptions on the distribution of skills. We assume that skills have a lognormal distribution (the commonly assumed distribution for incomes). The only assumption we make on the difference between the two groups is that group e has a higher mean wage than group i. The two lognormal distributions have identical other moments. Given the shape of lognormal distributions, it follows that the hazard rate for an individual of type w will be higher in group i for initially low values of w. Thus initially $\Omega_i(w) < \Omega_{ei}(w) < \Omega_e(w)$. Now as w increases, $\Omega(w)$ decreases for lognormal distributions (at least initially). Two possibilities may arise. First, $\Omega_e(w)$ remains always above $\Omega_i(w)$ Note that at high values of w, $f_i(w) < f_e(w)$ but $F_i(w) > F_e(w)$ so that the hazard rate in group e can remain always below the hazard rate in group i. In this case the $\Omega_i(w) < \Omega_{ei}(w) < \Omega_e(w)$ property is always satisfied. Second, it is *possible* that as w increases, at some point, $\Omega_e(w)$ equates $\Omega_i(w)$ and then follows below it. We show in the Appendix that in this case $\Omega_e(w)$ cannot cross $\Omega_i(w)$ more than once (assuming that $\Omega(w)$ is always decreasing in w so that the second-order condition for the optimal income tax problem is satisfied). Denote the value of w at which $\Omega_i(w) = \Omega_e(w)$, if this point exists, by w^D. We will then have $\Omega_i(w) < \Omega_{ei}(w) < \Omega_e(w)$ for all $w \in [\underline{w}, w^D)$, and $\Omega_i(w) > \Omega_{ei}(w) > \Omega_e(w)$ for all $w \in (w^D, \overline{w}]$.

First, observe that the lognormal assumption has implications for the marginal tax rates of the two groups. Specifically, we have that $T_i'(I) < T_{ei}'(I) < T_e'(I)$ holds either for all w, or for all $w \in [\underline{w}, w^D)$. In the latter case, we will have $T_i'(I) > T_{ei}'(I) > T_e'(I)$ for all $w \in (w^D, \overline{w}]$.

Next, turning to the question of welfare, consider the expressions for $u^N(w)$, $u_j^G(w)$, $c^N(w)$, and $c_j^G(w)$, derived in the Appendix:

$$u^N(w) = \underline{u}^N + \left(\frac{\varepsilon}{1+\varepsilon}\right)^{\varepsilon} \int_{\underline{w}}^{w} s^{\varepsilon} \left[1 + \left(1 + \frac{1}{\varepsilon}\right) \Omega_{ei}(s)\right]^{-1-\varepsilon} ds, \tag{38}$$

$$u_j^G(w) = \underline{u}^G + \left(\frac{\varepsilon}{1+\varepsilon}\right)^{\varepsilon} \int_{\underline{w}}^{w} s^{\varepsilon} \left[1 + \left(1 + \frac{1}{\varepsilon}\right) \Omega_j(s)\right]^{-1-\varepsilon} ds, \tag{39}$$

$$c^N(w) = u^N(w) + \left(\frac{\varepsilon}{1+\varepsilon}\right)^{1+\varepsilon} w^{1+\varepsilon} \left[1 + \left(1 + \frac{1}{\varepsilon}\right) \Omega_{ei}(w)\right]^{-1-\varepsilon}, \tag{40}$$

$$c_j^G(w) = u_j^G(w) + \left(\frac{\varepsilon}{1+\varepsilon}\right)^{1+\varepsilon} w^{1+\varepsilon} \left[1 + \left(1 + \frac{1}{\varepsilon}\right) \Omega_j(w)\right]^{-1-\varepsilon}. \tag{41}$$

Equations (38)–(39) tell us that, as long as $\Omega_i(w) < \Omega_{ei}(w) < \Omega_e(w)$, which either always holds or is satisfied for $w \in [\underline{w}, w^D)$, then $u_i^G(w) - u^N(w) > 0$, and $u_i^G(w) - u_e^G(w) > 0$. These results are quite comforting in that they tell us it is not just the poorest persons who become better off with tagging. Either all or the relatively poorer people in the tagged group i also become better off. Observe, however, that one does not know how $u_e^G(w)$ and $u^N(w)$ compare because while $\underline{u}^G > \underline{u}^N$, the second expression in $u_e^G(w)$ is smaller than the second expression in $u^N(w)$. This means that it is possible for a person with a low w in group e to become worse off as a result of tagging. Similarly, we have from (40)–(41), $c_i^G(w) - c^N(w) > 0$, and $c_i^G(w) - c_e^G(w) > 0$; but that the comparison between $c_e^G(w)$ and $c^N(w)$ is ambiguous.

If the hazard rates intersect at $w^D < \overline{w}$, equations (38)–(41) imply that for all $w \in (w^D, \overline{w}]$, $u_e^G(w) - u^N(w) > 0$, $u_e^G(w) - u_i^G(w) > 0, c_e^G(w) - c^N(w) > 0$, and $c_e^G(w) - c_i^G(w) > 0$; but we do not know how $u_i^G(w)$ compares with $u^N(w)$ and $c_i^G(w)$ with $c^N(w)$. That is, tagging may have the unintended consequence of making even the richer individuals in the not-favored group e better off. The following Proposition summarizes our results.

Proposition 3 *In addition to the assumptions as of Proposition 1, assume that the distribution of skills in the two groups is lognormal with group e having a higher mean than group i. Then:*

(i) The pattern of marginal income tax rates are $T_i'(I) < T_{ei}'(I) < T_e'(I)$ *for all* w*, or for all* $w \in [\underline{w}, w^D)$*. In the latter case,* $T_i'(I) > T_{ei}'(I) > T_e'(I)$ *for all* $w \in (w^D, \overline{w}]$.

(ii) Every poor individual $w \in [\underline{w}, w^D)$ *in group i, or all individuals in this group, benefit from tagging (in terms of consumption and utility levels and compared to the pooled equilibrium solution). These individuals will also consume more and have higher utility levels than corresponding w-type persons in group e. Similar individuals in group e may lose, as well as gain.*

(iii) Assume there exists a $w^D < \overline{w}$ *at which the hazard rate for group i becomes equal to the hazard rate for group e. Then, every individual* $w \in (w^D, \overline{w}]$ *in group e will benefit from tagging. These individuals will also consume more and have higher utility levels than corresponding w-type persons in group i. Similar individuals in group i may lose, as well as gain.*

5 Simulations

This section applies our theoretical model to the taxation of the US prime-age workers using age as the tag. We divide the working population in two groups of equal size on the basis of age. The age is construed as a proxy for "experience". We shall refer to all workers between the ages of 25–34 and 55–64 as the "inexperienced/infirm" group and those between the ages of 35 to 54 as the "experienced" group.[13] Specifically, we assume that the more experienced people tend to be employed in the more productive occupations.

Assume that the distribution of skills in each of the two groups of experienced (the

[13] Population growth, disability, early retirement, early death etc. imply that different age groups may contain different numbers of people. We shall ignore these empirical niceties here.

ages 35 to 54) and inexperienced/infirm (the ages 25 to 34 and 55 to 64) workers is truncated lognormal over the same support. Assume also that the variance of the distribution is the same in the two groups. Set this—the mean logarithmic deviation of income—at 0.37, which is the commonly used figure in the literature; see, e.g., Tuomala (1990) and Saez (2001). Configure the means of the two skills distributions in such a way that the median gross income will be \$55,000 in the experienced group and \$45,000 in the inexperienced/infirm group. These are very close to their observed values for the year 2001, as reported by the US Census Bureau (and cited in the Introduction). The pooled population also has a truncated lognormal distribution with the same variance. Its mean is configured to correspond to a median income of \$50,000. Finally, assume that the wage elasticity of labor supply is constant everywhere and identical for the two populations; setting its value at 0.5 which is, again, the commonly used figure in this literature.

Figure 1 presents the graphs of the optimal marginal income tax rates for the experienced and inexperienced/infirm groups along with the optimal marginal income tax rates for the pooled population (as a function of income). Observe that, in line with our theoretical result, the two groups' marginal tax rates bracket the pooled population's tax rate at every income level, with the experienced group facing the higher rate. Observe also that the marginal tax rates are everywhere decreasing in income. They start at the very high rate of 80 percent for an income level of \$10,000 and decrease all the way to zero for the top income level of \$200,000.

Turning to the average income tax rates, their graphs are presented in Figure 2. Unlike the marginal income tax rates, average tax rates are increasing in income, indicating the progressivity of the three tax schedules. Observe that the average tax rates are very close across the groups up to an income level of \$30,000, after which the tax rate for the pooled population is bracketed by those of the experienced group (at a higher rate) and the inexperienced/infirm group (at a lower rate).

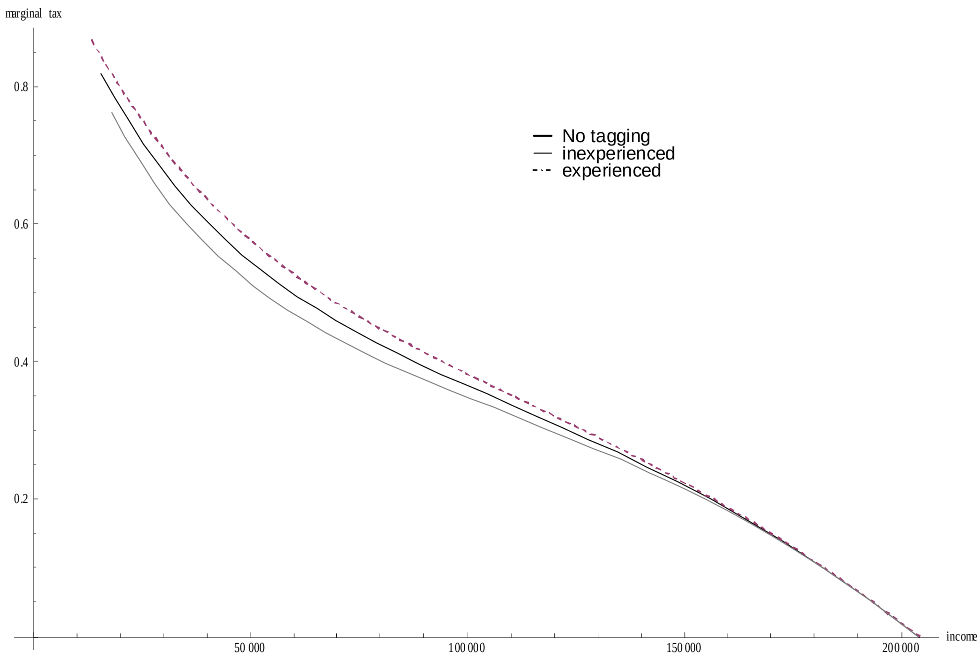

Figure 1: Optimal marginal income tax rates.

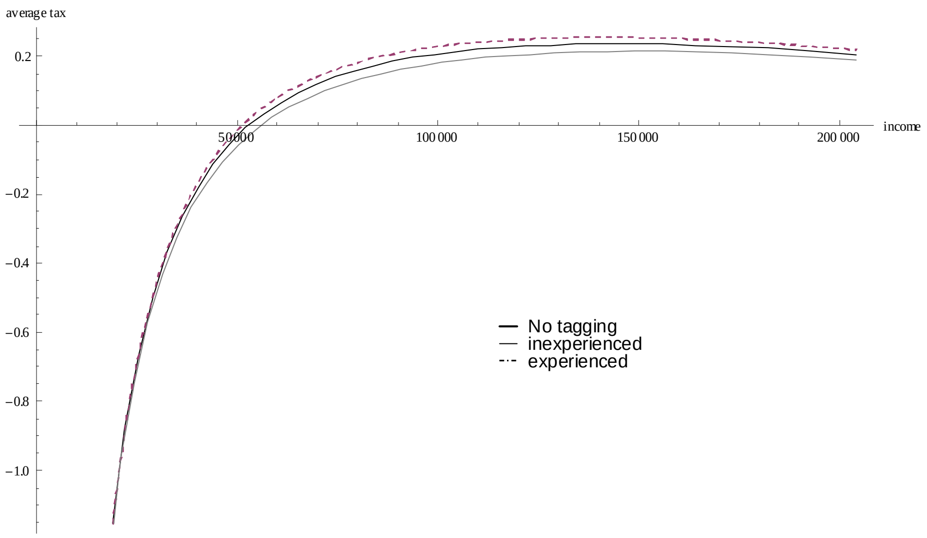

Figure 2: Optimal average income tax rates.

Table 1. Welfare implications of tagging as compared to the pooled population solution
(monetary figures in dollars per year)

	Inexperienced/infirm		Experienced	
	Utility change %	Compensating variation	Utility change %	Compensating variation
Least well-off	0.6	249	0.6	249
Skills deciles				
1	1.4	585	0	-22
2	2.5	$1,022$	-1.0	-429
3	3.3	$1,483$	-2.0	-889
4	4.0	$1,922$	-2.8	$-1,344$
5	4.4	$2,320$	-3.3	$-1,762$
6	4.6	$2,662$	-3.6	$-2,123$
7	4.5	$2,937$	-3.7	$-2,412$
8	4.3	$3,133$	-3.6	$-2,617$
9	3.9	$3,241$	-3.3	$-2,728$
10	3.4	$3,264$	-2.9	$-2,752$

The higher average tax rates faced by workers in the experienced group shows that they are the losers in the tagging procedure. To get a clearer picture of these welfare effects, Table 1 reports how workers in the two groups are affected at different decile levels of skills. Observe, first, that tagging improves the welfare of the least well-off individuals in both groups by \$249 per year. This is the Compensating Variation measure of the welfare change (which, given our quasi-linear specification for the preferences, is also equal to the Equivalent Variation measure). In terms of utility, this amounts to an increase of 0.6%. Turning to individuals with higher wages, workers in the inexperienced/infirm group gain and workers in the experienced group lose at all deciles. The gains range from \$585 per year for a worker at the lowest decile to \$3,264 per year at the highest decile. These are matched by losses of \$22 to \$2,752.

Given our Rawlsian perspective, the gains and losses to workers at different deciles are not part of the calculus of social welfare. The improvement in social welfare is the \$249 per year gain enjoyed by the least well-off workers in *both* groups. However, that the gains at every decile outweigh the losses by a considerable sum attests to the "efficiency-enhancing" power of tagging. Leaving premature deaths and illnesses aside, all workers go through the same life cycle. Consequently, although one ends up losing during some years of one's working life (ages 35 to 54), this is more than compensated by one's gain at other years (ages 25 to 34 and 55 to 64).

6 Conclusion

The optimal marginal income tax rate in Mirrlees depends on many factors. This makes the study of tagging somewhat difficult. This paper has restricted the analysis to quasi-linear preferences, and a maxi-min social welfare criterion, to reduce the determinants of the tax rate to the labor supply elasticity and the hazard rate of skills. Considering an economy that can be tagged into two different groups, each consisting of individuals with a continuum of skills over the same support, with every person having the same constant wage elasticity of labor supply, we have been able to show that the optimal marginal income tax rates in the two tagged groups bracket the optimal marginal tax rate in the pooled population at *all skill levels.* This result does not rest on any particular assumption concerning the skills distributions.

Secondly, we have shown that if one group first-order stochastically dominates the other, tagging will entail redistribution from that group to the latter. Third, we have derived a number of closed-form solutions for income, consumption, and utility of individuals with varying skill levels, when they are tagged into two groups as well as when the whole population is pooled. Fourth, we have shown that if the mean of skills distribution is higher in one group, assuming that the distribution of skills is lognormal and that the hazard rates in the two tagged groups do not cross, every individual in the group with the lower mean will benefit from tagging (in terms of consumption and utility levels and compared to the pooled equilibrium solution). These individuals will also consume more and have higher utility levels than their counterparts with the same ability in the group with the higher mean. The individuals in the higher mean group may lose, as well as gain.

Finally, as an application of our model, we have considered tagging on the basis of age. Tagging is often considered objectionable because it violates the principle of horizontal equity. However, if one adopts a lifetime perspective on this issue, tagging on the basis of age escapes this objection. Taxing a person harsher during some years of his life, while treating him more favorably during other years, applies to every worker and thus, viewed from a lifetime perspective, does not violate the principle of horizontal equity.

We have calibrated our model to the population of prime age workers in the US, and have derived the optimal marginal and average income tax rates for the tagged solutions as well as for the pooled population for all income brackets between $10,000 and $200,000 (assuming all workers have the same wage elasticity of labor supply equal to 0.5). The marginal tax rates start at the very high rate of 80 percent and decline all the way to zero, with the tax rates for the experienced and inexperienced/infirm groups bracketing the tax rate for pooled population. The average tax rates are always increasing in income indicating the progressivity of all three tax schedules. They are very close across the groups up to the income level of $30,000, after which the tax rate for the pooled population is bracketed by a higher rate for the experienced group and a lower rate for the inexperienced/infirm group.

In terms of welfare, we have shown that tagging improves the welfare of the least well-off individuals in both groups by $249 per year. Given our Rawlsian perspective, this also measures the gain in social welfare. Nevertheless, we have also calculated the gains and losses to workers at all deciles of skills distribution. Workers in the inexperienced/infirm group gain anywhere from $585 per year at the lowest decile to

\$3,264 per year at the highest decile. Workers of the same ability type in the experienced group, on the other hand, lose from \$22 at the lowest decile to \$2,752 at the highest decile. These numbers are, of course, quite tentative. Yet that the gains outweigh the losses by quite a bit, suggests that tagging by age can enhance the efficiency of the tax system considerably. This merits further investigation.

Appendix

Derivation of equation (6): Substitute for $c(w)$ from (5) into (2) to arrive at

$$R = \int_{\underline{w}}^{\overline{w}} \left[I(w) - h\left(\frac{I(w)}{w}\right)\right] f(w)dw - \underline{u} \\ - \int_{\underline{w}}^{\overline{w}} \left[\int_{\underline{w}}^{w} \frac{I(s)}{s^2} h'\left(\frac{I(s)}{s}\right) ds \right] f(w)dw. \tag{A1}$$

Integrating the last expression in the right-hand side of (A1) by parts, we get

$$\int_{\underline{w}}^{\overline{w}} \left[\int_{\underline{w}}^{w} \frac{I(s)}{s^2} h'\left(\frac{I(s)}{s}\right) ds \right] f(w)dw = $$

$$\left[\int_{\underline{w}}^{w} \frac{I(s)}{s^2} h'\left(\frac{I(s)}{s}\right) dsF(w) \right]_{\underline{w}}^{\overline{w}} - \int_{\underline{w}}^{\overline{w}} F(w)\frac{I(w)}{w^2} h'\left(\frac{I(w)}{w}\right) dw = $$

$$\int_{\underline{w}}^{\overline{w}} \frac{I(w)}{w^2} h'\left(\frac{I(w)}{w}\right) [1 - F(w)]dw.$$

Substituting this expression into (A1) and simplifying results in equation (6).

Derivation of the expression for $I_j^*(w)$: Setting $h(L) = L^{1+\frac{1}{\varepsilon}}$ simplifies the expression for $\Psi(F_j)$ in (7), $j = e, i, ei$, to

$$\Psi(F_j) = \int_{w^-}^{w^+} \left[I_j(w) - \left(\frac{I_j(w)}{w}\right)^{1+\frac{1}{\varepsilon}} \right] f_j(w)dw - \\ \int_{w^-}^{w^+} \left(1 + \frac{1}{\varepsilon}\right) \left(\frac{I_j(w)}{w}\right)^{1+\frac{1}{\varepsilon}} \frac{[1 - F_j(w)]}{w} dw \\ = \int_{w^-}^{w^+} \left\{ I_j(w) - \left[1 + \left(1 + \frac{1}{\varepsilon}\right)\Omega_j(w)\right] \left(\frac{I_j(w)}{w}\right)^{1+\frac{1}{\varepsilon}} \right\} f_j(w)dw. \tag{A2}$$

The first-order condition for the maximization of $\Psi(F_j)$ with respect to $I_j(w)$ then is

$$1 - \left[1 + \left(1 + \frac{1}{\varepsilon}\right)\Omega_j(w)\right] \frac{1 + \frac{1}{\varepsilon}}{w} \left(\frac{I_j(w)}{w}\right)^{\frac{1}{\varepsilon}} = 0.$$

Solving this equation for $I_j(w)$ yields

$$I_j^*(w) = wL_j^*(w) = w^{1+\varepsilon} \left(\frac{\varepsilon}{1+\varepsilon}\right)^{\varepsilon} \left[1 + \left(1 + \frac{1}{\varepsilon}\right)\Omega_j(w)\right]^{-\varepsilon}, \quad j = e, i, ei. \tag{A3}$$

Proof of the claim that $\Omega_e(w)$ cannot cross $\Omega_i(w)$ more than once: When the skills distribution is lognormal, the expression for $\Omega(w)$ is given by

$$\Omega(w) = \frac{\sigma\sqrt{2\pi}}{2}\left[1 - \frac{2}{\sqrt{\pi}}\int_0^{\frac{\ln w-\mu}{\sigma\sqrt{2}}} e^{-t^2}dt\right] e^{\frac{(\ln w-\mu)^2}{2\sigma^2}},$$

where μ and σ^2 are the mean and variance of $\ln w$. Differentiating $\Omega(w)$ with respect to w yields

$$\begin{aligned}\frac{d\Omega}{dw} &= -\frac{1}{w} + \frac{\sqrt{2\pi}}{2\sigma w}(\ln w - \mu)\left[1 - \frac{2}{\sqrt{\pi}}\int_0^{\frac{\ln w-\mu}{\sigma\sqrt{2}}} e^{-t^2}dt\right] e^{\frac{(\ln w-\mu)^2}{2\sigma^2}} \\ &= \frac{1}{w}\left[(\ln w - \mu)\frac{\Omega}{\sigma^2} - 1\right]. \end{aligned} \tag{A4}$$

Now denote the expression for Ω by Ω_e when $\mu = \mu_e$ and by Ω_i when $\mu = \mu_i$, where, by assumption, $\mu_e > \mu_i$. It then follows from equation (A4) that

$$\frac{d}{dw}(\Omega_e - \Omega_i) = \frac{1}{w\sigma^2}\left[(\ln w - \mu_e)\,\Omega_e - (\ln w - \mu_i)\,\Omega_i\right]. \tag{A5}$$

Equation (A5) tells us that whenever $\Omega_e = \Omega_i (= \Omega)$,

$$\frac{d}{dw}(\Omega_e - \Omega_i) = \frac{\Omega}{w\sigma^2}\left[\mu_i - \mu_e\right] < 0.$$

But this can not happen two consecutive times if Ω_e and Ω_i are always decreasing in w.

Derivation of equations (38)–(41): With $h(L) = L^{1+\frac{1}{\varepsilon}}$, we have

$$\begin{aligned}\int_{\underline{w}}^{w} \frac{I_j^*(s)}{s^2} h'\left(\frac{I_j^*(s)}{s}\right) ds &= \int_{\underline{w}}^{w}\left(1+\frac{1}{\varepsilon}\right)\frac{L_j^*(s)}{s}\; L_j^*(s)^{\frac{1}{\varepsilon}} ds \\ &= \left(1+\frac{1}{\varepsilon}\right)\int_{\underline{w}}^{w} s^{-1}\; L_j^*(s)^{1+\frac{1}{\varepsilon}} ds.\end{aligned}$$

Substitute for $L_j^*(w)$ from (A3) into above. This gives

$$\int_{\underline{w}}^{w} \frac{I_j^*(s)}{s^2} h'\left(\frac{I_j^*(s)}{s}\right) ds = \left(\frac{\varepsilon}{1+\varepsilon}\right)^{\varepsilon}\int_{\underline{w}}^{w} s^{\varepsilon}\left[1+\left(1+\frac{1}{\varepsilon}\right)\Omega_j(s)\right]^{-1-\varepsilon} ds.$$

Given this expression, one can simplify the expressions for $u^N(w)$ and $u_j^G(w)$, $j=e,i$, derived in (19) and (25), to (38)–(39). Similarly, one simplifies the expressions for $c^N(w)$ and $c_j^G(w)$, $j=e,i$, derived in (20) and (26), to (40)–(41).

Derivation of the expressions for $\underline{u}^N$ and $\underline{u}^G$: First, substitute the expression for $I_j^*(w)$ from (A3) into (A2). This gives the optimized value of $\Psi(F_j)$, $j=e,i,ei$, as

$$\begin{aligned}\Psi^*(F_j) &= \int_{\underline{w}}^{\overline{w}} L_j^*(w)\left\{w-\left[1+\left(1+\frac{1}{\varepsilon}\right)\Omega_j(w)\right]\; L_j^*(w)^{\frac{1}{\varepsilon}}\right\} f_j(w)dw,\\ &= \frac{1}{1+\varepsilon}\int_{\underline{w}}^{\overline{w}} wL_j^*(w)f_j(w)dw,\\ &= \varepsilon^{\varepsilon}\left(1+\varepsilon\right)^{-1-\varepsilon}\int_{\underline{w}}^{\overline{w}} w^{1+\varepsilon}\left[1+\left(1+\frac{1}{\varepsilon}\right)\Omega_j(w)\right]^{-\varepsilon} f_j(w)dw. \qquad \text{(A6)}\end{aligned}$$

Next, given the expression for $\Psi^*(F_{ei})$ and using (18), we have

$$\underline{u}^N = \Psi^*(F_{ei}) = \varepsilon^{\varepsilon}\left(1+\varepsilon\right)^{-1-\varepsilon}\int_{\underline{w}}^{\overline{w}} w^{1+\varepsilon}\left[1+\left(1+\frac{1}{\varepsilon}\right)\Omega_{ei}(w)\right]^{-\varepsilon} f_{ei}(w)dw,$$

Finally, given the expression for $\Psi^*(F_j)$, $j=e,i$, and using (24), we have

$$\begin{aligned}\underline{u}^G &= \frac{\Psi^*(F_e)+\Psi^*(F_i)}{2}\\ &= \frac{1}{2}\varepsilon^{\varepsilon}\left(1+\varepsilon\right)^{-1-\varepsilon}\left\{\int_{\underline{w}}^{\overline{w}} w^{1+\varepsilon}\left[1+\left(1+\frac{1}{\varepsilon}\right)\Omega_e(w)\right]^{-\varepsilon} f_e(w)dw+\right.\\ &\qquad \left.\int_{\underline{w}}^{\overline{w}} w^{1+\varepsilon}\left[1+\left(1+\frac{1}{\varepsilon}\right)\Omega_i(w)\right]^{-\varepsilon} f_i(w)dw\right\}.\end{aligned}$$

Proof of inequality (34): Let

$$\Delta \equiv \int_{\underline{w}}^{\overline{w}} I_e^*(w)\; f_e(w)dw + \int_{\underline{w}}^{\overline{w}} I_i^*(w)\; f_i(w)dw - 2\int_{\underline{w}}^{\overline{w}} I_{ei}^*(w)\; f_{ei}(w)dw.$$

Rewrite this expression as

$$\Delta = \int_{\underline{w}}^{\overline{w}} \left[I_e^*(w)-I_{ei}^*(w)\right] f_e(w)dw + \int_{\underline{w}}^{\overline{w}} \left[I_i^*(w)-I_{ei}^*(w)\right] f_i(w)dw.$$

Substitute for $I_i^*(w)$, $j = e, i, ei$, from (A3) in above and simplify. We have

$$\begin{aligned}
\Delta &= \left(\frac{\varepsilon}{1+\varepsilon}\right)^{\varepsilon} \int_{\underline{w}}^{\overline{w}} w^{1+\varepsilon} \Bigg\{ \left[1 + \left(1 + \frac{1}{\varepsilon}\right)\Omega_e(w)\right]^{-\varepsilon} - \\
&\left[1 + \left(1 + \frac{1}{\varepsilon}\right)\Omega_{ei}(w)\right]^{-\varepsilon} \Bigg\} f_e(w)dw + \left(\frac{\varepsilon}{1+\varepsilon}\right)^{\varepsilon} \int_{\underline{w}}^{\overline{w}} w^{1+\varepsilon} \times \\
&\left\{ \left[1 + \left(1 + \frac{1}{\varepsilon}\right)\Omega_i(w)\right]^{-\varepsilon} - \left[1 + \left(1 + \frac{1}{\varepsilon}\right)\Omega_{ei}(w)\right]^{-\varepsilon} \right\} f_i(w)dw. \qquad \text{(A7)}
\end{aligned}$$

Next, observe that from the expressions for $\underline{u}^G$ and $\underline{u}^N$, we have

$$\begin{aligned}
\underline{u}^G - \underline{u}^N &= \frac{1}{2}\varepsilon^{\varepsilon}(1+\varepsilon)^{-1-\varepsilon} \left\{ \int_{\underline{w}}^{\overline{w}} w^{1+\varepsilon} \left[1 + \left(1 + \frac{1}{\varepsilon}\right)\Omega_e(w)\right]^{-\varepsilon} f_e(w)dw + \right. \\
&\left. \int_{\underline{w}}^{\overline{w}} w^{1+\varepsilon} \left[1 + \left(1 + \frac{1}{\varepsilon}\right)\Omega_i(w)\right]^{-\varepsilon} f_i(w)dw \right\} - \\
&\varepsilon^{\varepsilon}(1+\varepsilon)^{-1-\varepsilon} \int_{\underline{w}}^{\overline{w}} w^{1+\varepsilon} \left[1 + \left(1 + \frac{1}{\varepsilon}\right)\Omega_{ei}(w)\right]^{-\varepsilon} f_{ei}(w)dw \\
&= \frac{1}{2}\varepsilon^{\varepsilon}(1+\varepsilon)^{-1-\varepsilon} \left\{ \int_{\underline{w}}^{\overline{w}} w^{1+\varepsilon} \left[1 + \left(1 + \frac{1}{\varepsilon}\right)\Omega_e(w)\right]^{-\varepsilon} f_e(w)dw - \right. \\
&\left. \int_{\underline{w}}^{\overline{w}} w^{1+\varepsilon} \left[1 + \left(1 + \frac{1}{\varepsilon}\right)\Omega_{ei}(w)\right]^{-\varepsilon} f_{ei}(w)dw \right\} + \\
&\frac{1}{2}\varepsilon^{\varepsilon}(1+\varepsilon)^{-1-\varepsilon} \left\{ \int_{\underline{w}}^{\overline{w}} w^{1+\varepsilon} \left[1 + \left(1 + \frac{1}{\varepsilon}\right)\Omega_i(w)\right]^{-\varepsilon} f_i(w)dw - \right. \\
&\left. \int_{\underline{w}}^{\overline{w}} w^{1+\varepsilon} \left[1 + \left(1 + \frac{1}{\varepsilon}\right)\Omega_{ei}(w)\right]^{-\varepsilon} f_{ei}(w)dw \right\} \\
&= \frac{1}{2}\varepsilon^{\varepsilon}(1+\varepsilon)^{-1-\varepsilon} \left\{ \int_{\underline{w}}^{\overline{w}} w^{1+\varepsilon} \left[1 + \left(1 + \frac{1}{\varepsilon}\right)\Omega_e(w)\right]^{-\varepsilon} f_e(w)dw - \right. \\
&\left. \frac{1}{2}\int_{\underline{w}}^{\overline{w}} w^{1+\varepsilon} \left[1 + \left(1 + \frac{1}{\varepsilon}\right)\Omega_{ei}(w)\right]^{-\varepsilon} \left[f_e(w) + f_i(w)\right] dw \right\} +
\end{aligned}$$

$$
\begin{aligned}
& \frac{1}{2}\varepsilon^{\varepsilon}(1+\varepsilon)^{-1-\varepsilon}\left\{\int_{\underline{w}}^{\overline{w}} w^{1+\varepsilon}\left[1+\left(1+\frac{1}{\varepsilon}\right)\Omega_i(w)\right]^{-\varepsilon} f_i(w)dw - \right. \\
& \left. \frac{1}{2}\int_{\underline{w}}^{\overline{w}} w^{1+\varepsilon}\left[1+\left(1+\frac{1}{\varepsilon}\right)\Omega_{ei}(w)\right]^{-\varepsilon}\left[f_e(w)+f_i(w)\right]dw\right\} \\
= \; & \frac{1}{2}\varepsilon^{\varepsilon}(1+\varepsilon)^{-1-\varepsilon}\left\{\int_{\underline{w}}^{\overline{w}} w^{1+\varepsilon}\left[1+\left(1+\frac{1}{\varepsilon}\right)\Omega_e(w)\right]^{-\varepsilon} f_e(w)dw - \right. \\
& \left. \int_{\underline{w}}^{\overline{w}} w^{1+\varepsilon}\left[1+\left(1+\frac{1}{\varepsilon}\right)\Omega_{ei}(w)\right]^{-\varepsilon} f_e(w)dw\right\} + \\
& \frac{1}{2}\varepsilon^{\varepsilon}(1+\varepsilon)^{-1-\varepsilon}\left\{\int_{\underline{w}}^{\overline{w}} w^{1+\varepsilon}\left[1+\left(1+\frac{1}{\varepsilon}\right)\Omega_i(w)\right]^{-\varepsilon} f_i(w)dw - \right. \\
& \left. \int_{\underline{w}}^{\overline{w}} w^{1+\varepsilon}\left[1+\left(1+\frac{1}{\varepsilon}\right)\Omega_{ei}(w)\right]^{-\varepsilon} f_i(w)dw\right\} \\
= \; & \frac{1}{2}\varepsilon^{\varepsilon}(1+\varepsilon)^{-1-\varepsilon}\int_{\underline{w}}^{\overline{w}} w^{1+\varepsilon}\left\{\left[1+\left(1+\frac{1}{\varepsilon}\right)\Omega_e(w)\right]^{-\varepsilon} - \right. \\
& \left.\left[1+\left(1+\frac{1}{\varepsilon}\right)\Omega_{ei}(w)\right]^{-\varepsilon}\right\} f_e(w)dw + \frac{1}{2}\varepsilon^{\varepsilon}(1+\varepsilon)^{-1-\varepsilon}\int_{\underline{w}}^{\overline{w}} w^{1+\varepsilon} \times \\
& \left\{\left[1+\left(1+\frac{1}{\varepsilon}\right)\Omega_i(w)\right]^{-\varepsilon} - \left[1+\left(1+\frac{1}{\varepsilon}\right)\Omega_{ei}(w)\right]^{-\varepsilon}\right\} f_i(w)dw. \qquad \text{(A8)}
\end{aligned}
$$

Substituting from (A8) into (A7), we get

$$
\Delta = 2(1+\varepsilon)\left(\underline{u}^G - \underline{u}^N\right) \geqq 0.
$$

References

[1] Akerlof, George A. 1978. The economics of "tagging" as applied to the Optimal income tax, welfare programs, and manpower planning. American Economic Review, vol 68, 8–19.

[2] Alesina, Alberto, Ichino, Andrea, and Loukas Karabarbounis. 2007. Gender based taxation and the division of family chores. Unpublished manuscript, Harvard University.

[3] Banks, James, Diamond, Peter, and James Mirrlees. 2008. The base for direct taxation, Mirrlees Report, mimeo.

[4] Bennett, John. 1987. The second-best lump-sum taxation of observable characteristics. Public Finance/Finances Publiques. vol 42. 227–235.

[5] Boadway, Robin and Laurence Jacquet. 2008. Optimal marginal and average income taxation under maxi-min. Journal of Economic Theory, forthcoming.

[6] Boadway, Robin and Pierre Pestieau. 2006. Tagging and redistributive taxation. Annales d'Economie et de Statistique 83/84, 123–150.

[7] DeNavas-Walt, Carmen, Cleveland, Robert, and Bruce H. Webster Jr. 2003. Income in the United States: 2002. US Census Bureau, Current Population Reports, P60–221, US Government Printing Office, Washington DC.

[8] Dilnot, Andrew W., Kay, John A., and C.N. Morris. 1984. The Reform of Social Security. Oxford: Clarendon Press.

[9] Hamilton, Jonathan and Pierre Pestieau. 2005. Optimal income taxation and the ability distribution: implications for migration equilibria. International Tax and Public Finance vol 12, 29–45.

[10] Immonen, Ritva, Kanbur, Ravi, Keen, Michael, and Matti Tuomala. 1998. Tagging and taxing: the optimal use of categorical and income information in designing tax/transfer schemes. Economica, vol 65, 179–192.

[11] Jorgenson, Dale W., McCall, John J., and Roy Radner. 1967. Replacement of Industrial Equipment—Mathematical Models. Amsterdam: North Holland Publishing Company.

[12] Kremer, Michael. 2001. Should taxes be independent of age? Unpublished manuscript, Harvard University.

[13] Mankiw, N. Gregory and Matthew Weinzierl. 2007. The optimal taxation of height: A case study of utilitarian income redistribution. Unpublished manuscript, Harvard University.

[14] Parsons, Donnald O. 1996. Imperfect 'tagging' in social insurance programs. Journal of Public Economics, vol 62, 183–207.

[15] Saez, Emmanuel 2001. Using elasticities to derive optimal income tax rates. Review of Economic Studies, vol 68, 205–229.

[16] Tuomala,Matti. 1990. Optimal income tax and redistribution. Oxford: Clarendon Press

[17] Salanié, Bernard. 2003. The Economics of taxation. Cambridge, Massachusetts: MIT Press.

[18] Stern, Nicholas. 1982. Optimum taxation with errors in administration. Journal of Public Economics, vol 17, 181–211.

[19] Viard, Alan D. 2001. Some results on the comparative statics of optimal categorical transfer payments. Public Finance Review, vol 29, 148–180.

[20] Viard, Alan D. 2001. Optimal categorical transfer payments: the welfare economics of limited lump-sum redistribution. Journal of Public Economic Theory, vol 3, 483–500.