Tagging and income taxation: theory and an application*

Helmuth Cremer
Toulouse School of Economics
(University of Toulouse and Institut universitaire de France)

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

Jean-Marie Lozachmeur
Toulouse School of Economics
(CNRS)

April 2008
Revised, December 2008

*We thank Sören Blomquist, Ray Rees, two referees of this journal, and the Editor, Alan Auerbach, for helpful comments and advice.

Abstract

We derive a set of analytical results for optimal income taxation with tags using quasilinear preferences and a Rawlsian social welfare function. Secondly, assuming a constant elasticity of labor supply and log-normality of the skills distribution, we analytically identify the winners and losers of tagging. Third, we prove 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. Finally, we calibrate our model to the US workers using gender as tag. Welfare implications are dramatic: Only male high-wage earners lose; everyone else gains and some substantially.

Key words: Tagging, optimal income taxation, hazard rate, gender, race.

JEL Classification: H21, H23.

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, George A. 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.

Akerlof 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, and many studies, the question of how tagging might change the properties of an optimal general income tax schedule remains, to a large extent, an open one.[2] Nor do we know much beyond Akerlof's original findings on who gains or loses as a result of tagging. This paper aims to fill in these gaps, albeit under certain strong assumptions. It considers a model with a continuum of individuals who can

[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.

[2] The papers in this area include Nicholas Stern (1982), Andrew W. Dilnot, John A. Kay, and C.N. Morris (1984), John Bennett (1987), Donald O. Parsons (1996), Ritva Immonen, Ravi Kanbur, Michael Keen, and Matti Tuomala (1998), Michael Kremer (2001), Viard (2001a, b), Jonathan Hamilton and Pierre Pestieau (2004), Robin Boadway and Pestieau (2006), Alberto Alesina, Andrea Ichino, and Loukas Karabarbounis. (2007), N. Gregory Mankiw and Matthew Weinzierl (2007), Sören Blomquist and Luca Micheletto (2008), and Weinzierl (2008).

be divided into two groups with different ability distributions over the same support. Assuming that preferences are quasi-linear and the social welfare function is Rawlsian, it derives a set of analytical results for optimal income taxation with tags. The tagging is based on a publicly and costlessly observable exogenous characteristic. Two additional assumptions, namely log-normality of the skills distribution and a constant elasticity of labor supply, allow us to analytically identify the winners and losers of tagging.

Specifically, we show that if the hazard rates in the two tagged groups do not cross, *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 enjoy 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. If the hazard rates cross, gainers and losers switch sides after the crossing.

Our other results include the finding 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. This result holds for a general utility function as long as the social welfare function is Rawlsian. It also holds for a utilitarian social welfare function with decreasing weights in skills provided that preferences are again quasi-linear.

In terms of analytical results, the closest paper to ours is Boadway and Pestieau (2006). They 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). As with our model, this formulation allows 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 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]

To supplement our analytical results, we study the implications of using gender tags in designing optimal income taxes in the US. According to PSID data, the median earnings in 1993 were approximately \$30,000 for men as compared to \$20,000 for women, with corresponding mean logarithmic deviations of 0.25 and 0.20. There are of course many other differences between these groups beyond those due to average earnings. We ignore them all, assuming that all groups have lognormal distribution of skills that differ in mean incomes only. Additionally, we assume that all workers have a constant wage elasticity of labor supply (which we set equal to 0.5). The reason for these assumptions is consistency with our theoretical model which concentrated on the differences in average earnings alone. These somewhat unrealistic features must be borne in mind when interpreting our quantitative results. In particular, the numbers are not meant to show what will actually happen in the US as a result of tagging. They are simply suggestive.

We find that tagging on the basis of gender has dramatic implications for welfare and optimal tax rates. It increases the welfare of the least well-off male and female workers alike by \$1,174 per year (equivalent to an increase of 6% in terms of utility). At higher wage levels, women gain at all deciles. The gains range from \$1,677 per year for a female worker at the lowest decile to \$13,728 per year at the highest decile. Interestingly, low-earning male workers also gain with those at the first decile gaining by as much as

[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.

[4] Boadway and Pestieau (2004) also state 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).

\$935 per year. The gains decrease for higher deciles, turning negative for the fifth. After that, male workers lose more and more as their earning ability increases, with the loss reaching \$8,397 per year mark at the tenth decile. Concerning tax rates, we derive, for all income brackets between \$10,000 and \$175,000, the values of the optimal marginal and average income tax rates for males and females as well as for when the population is pooled. The marginal tax rates are always decreasing in income. The highest marginal tax rate faced by men is 80% percent versus only 26% for women. Average tax rates increase sharply initially and then level off. The highest average tax rates are 26% for men versus to less than one per cent for women.

I 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", 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 lower average skills as group l, and to the other group as group h. 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 $F_l(w)$ in group l, and $F_h(w)$ in group h. The support of the two distributions are the same and given by $[\underline{w}, \overline{w}]$. The corresponding density functions, $f_j(w) = F_j'(w)$, $j = l, h$, 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

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

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.

A The generic optimal income tax problem

As a background to the optimal tax problem with tagging, we first sketch the generic tax problem with quasi-linear preferences and a Rawlsian objective function. Bernard Salanié (2003) and Boadway and Laurence Jacquet (2008) give details. Let $c(w)$, $I(w) = wL(w)$, and $u(w)$ denote consumption, income, and utility of an individual of type w.

Denote the net taxes collected from the population by R. It follows from the quasi-linear specification (1) that

$$R \equiv \int_{\underline{w}}^{\overline{w}} [I(w) - c(w)] f(w) dw = \int_{\underline{w}}^{\overline{w}} \left[I(w) - u(w) - \varphi\left(\frac{I(w)}{w}\right) \right]. \tag{2}$$

Integrating the local incentive compatibility constraint, $du/dw = \left(I(w)/w^2\right) \varphi'\left(I(w)/w\right)$, we have

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

where $\underline{u} = u(\underline{w})$ is the utility of the poorest individual. One can then easily show that

$$R = -\underline{u} + \Psi(F), \tag{3}$$

where

$$\Psi(F) \equiv \int_{\underline{w}}^{\overline{w}} \left[I(w) - \varphi\left(\frac{I(w)}{w}\right) \right] f(w) dw - \int_{\underline{w}}^{\overline{w}} \frac{I(w)}{w^2} \varphi'\left(\frac{I(w)}{w}\right) [1 - F(w)] dw. \tag{4}$$

Observe that expression (4) measures the maximum tax revenue the government can, in the second best and for a given $I(w)$, 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 $w > \underline{w}$). Assuming that the government's external revenue requirement is $\overline{R}$, and given the skills distribution $F(\cdot)$, the utility of the poorest individual is given by

$$\underline{u}\left(\overline{R}, F\right) = -\overline{R} + \Psi(F), \tag{5}$$

With a Rawlsian social welfare function, the optimal income tax problem reduces to the choice of $I(w)$ to maximize $\Psi(F)$, or tax revenues.

Maximize $\Psi(F)$ with respect to $I(w)$ and simplify the first-order condition to get

$$\frac{1 - \frac{1}{w}\varphi'\left(I(w)/w\right)}{\frac{1}{w}\varphi'\left(I(w)/w\right)} = \left[\frac{1 - F(w)}{wf(w)}\right] \left[1 + \frac{I(w)}{w} \frac{\varphi''\left(I(w)/w\right)}{\varphi'\left(I(w)/w\right)}\right]. \tag{6}$$

Denote the solution to (6) by $I^*(w, F)$ and write $\Psi^*(F)$ for the maximized value of $\Psi(F)$. One can then establish that

$$\begin{aligned}
\underline{u}^*\left(\overline{R}, F\right) &= -\overline{R} + \Psi^*(F), && (7)\\
u^*\left(w, \overline{R}, F\right) &= \underline{u}^*\left(\overline{R}, F\right) + \int_{\underline{w}}^{w} \frac{I^*(s, F)(s)}{s^2}\varphi'\left(\frac{I^*(s, F)}{s}\right) ds, && (8)\\
c^*\left(w, \overline{R}, F\right) &= u^*\left(w, \overline{R}, F\right) + \varphi\left(\frac{I^*(w, F)}{w}\right). && (9)
\end{aligned}$$

Equations (8)–(9) 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 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.

To determine the implication of the solution $\underline{u}^*(\cdot)$, $u^*(\cdot)$, $c^*(\cdot)$ for the shape of the implementing tax function, denoted by $T(I^*(w, F))$, assume that $T(\cdot)$ is differentiable. Now, when an individual of type w maximizes $u = c - \varphi(I/w)$ subject to the budget constraint $c = I - T(I)$, he will have to solve the first-order condition, $1 - T'(I) - \varphi'(I/w)/w = 0$. Thus, the implementing tax function is determined according to $T'(I^*(w, F)) = 1 - \varphi'(I^*(w, F)/w)/w$. Rearranging this equation and introducing,

$$H(w; F) \equiv 1 + \frac{L^*(w, F)\varphi''(L^*(w, F))}{\varphi'(L^*(w, F))}, \tag{10}$$

$$\Omega(w) \equiv \frac{1 - F(w)}{wf(w)}, \tag{11}$$

the implementing tax function is characterized by

$$\frac{T'}{1 - T'} = H(w, F) \times \Omega(w). \tag{12}$$

With $\varphi'(\cdot) > 0$ and $\varphi''(\cdot) > 0$, it follows from (10) that $H(w, F) > 1$. Equation (12) then implies $T'(\cdot) > 0$ as long as $F(w) < 1$; it also implies that $T'(\cdot) = 0$ at $w = \overline{w}$ where $F(w) = 1$. Taxpayers' preferences enter (12) only through the $H(w, F)$ term. Observe also that $L^*(\cdot)\varphi''(\cdot)/\varphi'(\cdot)$ is equal to the inverse of the wage elasticity of labor supply for a w-type individual. Thus equation (12) 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.

If the elasticity of labor supply is constant, the marginal income tax rate is determined entirely by the shape of the hazard rate and the size of the wage rate. The hazard rate for lognormal distribution, has an initially increasing segment followed by a decreasing part; see Dale W. Jorgenson John J. McCall, and Roy Radner (1967). The corresponding marginal income tax rate must then be initially decreasing in w, but it may eventually become increasing.[5] Salanié (2003, p. 95) 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.

II Optimal taxation with and without tagging

We start with the characterization of optimal tax solution when the information from the tag is discarded. Then, we proceed to examine how conditioning the tax on the tag changes the properties of the optimal tax solution.

A Optimal solution without tagging

Denote the distribution function for the entire population by $F_{hl}(w)$ and its corresponding density by $f_{hl}(w)$. They are related to the distribution and the density functions of the two groups that comprise it according to:

$$F_{hl}(w) = \frac{F_l(w) + F_h(w)}{2}, \tag{13}$$

$$f_{hl}(w) = \frac{f_l(w) + f_h(w)}{2}. \tag{14}$$

[5] 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 optimal tax structure in this case is precisely the same as the one derived for the generic optimal tax problem, with $F_{hl}(w)$ and $f_{hl}(w)$ replacing $F(\cdot)$ and $f(\cdot)$. Denote the optimal solution without tagging, and with an external revenue requirement $\overline{R} = 0$, by $\underline{u}^*_{hl} = \underline{u}^*(0, F_{hl})$, $u^*_{hl}(w) = u^*(w, 0, F_{hl})$, and $c^*_{hl}(w) = c^*(w, 0, F_{hl})$. To be consistent, we denote the optimal values of $I(w)$ and $\Psi(F_{hl})$, which are independent of $\overline{R}$, by $I^*_{hl}(w) = I^*(w, F_{hl})$ and $\Psi^*_{hl} = \Psi^*(F_{hl})$. The optimal solutions are thus characterized by

$$\underline{u}^*_{hl} = \Psi^*_{hl}, \tag{15}$$

$$u^*_{hl}(w) = \underline{u}^*_{hl} + \int_{\underline{w}}^{w} \frac{I^*_{hl}(s)}{s^2} \varphi' \left(\frac{I^*_{hl}(s)}{s} \right) ds, \tag{16}$$

$$c^*_{hl}(w) = u^*_{hl}(w) + \varphi \left(\frac{I^*_{hl}(w)}{w} \right). \tag{17}$$

B Optimal solution under tagging

Given our 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 $\min[\underline{u}_l, \underline{u}_h]$. 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 Section I wherein $I_j(w), c_j(w; \overline{R}_j)$ are derived to maximize $\underline{u}_j$ under the constraint that $R_j \geqq \overline{R}_j$, $j = l, h$, 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 $\overline{R}_l, \overline{R}_h$ to maximize $\min \left[\underline{u}^*_l \left(\overline{R}_l \right), \underline{u}^*_h \left(\overline{R}_h \right) \right]$ subject to $\overline{R}_l + \overline{R}_h = 0$. This requires $\underline{u}^*_l \left(\overline{R}_l \right)$ and $\underline{u}^*_h \left(\overline{R}_h \right)$ to be equalized. Using (7), the optimal values of $\overline{R}_l, \overline{R}_h$ are found as the solution to

$$\begin{aligned} -\overline{R}_l + \Psi^*(F_l) &= -\overline{R}_h + \Psi^*(F_h), \\ \overline{R}_l + \overline{R}_h &= 0, \end{aligned}$$

and denoted by $\overline{R}_l^*$ and $\overline{R}_h^*$. We have,

$$\overline{R}_l^* = \frac{\Psi^*(F_l) - \Psi^*(F_h)}{2}, \tag{18}$$

$$\overline{R}_h^* = \frac{\Psi^*(F_h) - \Psi^*(F_l)}{2}. \tag{19}$$

This procedure highlights the fact that the interaction between groups occurs through the transfers only.

Denote $\underline{u}_l^*\left(\overline{R}_l^*\right) = \underline{u}_h^*\left(\overline{R}_h^*\right) \equiv \underline{u}^*$. The value of $\underline{u}^*$ can then be determined from equation (7) using the values of $\overline{R}_l^*$ and $\overline{R}_h^*$ from equations (18)–(19). Subsequently, one can determine the values of $u_j^*(w)$ and $c_j^*(w)$, $j = l, h$, from equations (8)–(9). We have

$$\underline{u}^* = \frac{\Psi^*(F_e) + \Psi^*(F_i)}{2}, \tag{20}$$

$$u_j^*(w) = \underline{u}^* + \int_{\underline{w}}^{w} \frac{I_j^*(s)}{s^2}\varphi'\left(\frac{I_j^*(s)}{s}\right) ds, \tag{21}$$

$$c_j^*(w) = u_j^*(w) + \varphi\left(\frac{I_j^*(w)}{w}\right). \tag{22}$$

The following proposition summarizes the results of this section.

Proposition 1 *Assume preferences are quasi-linear and the social welfare function is Rawlsian. There are two groups of individuals of equal size, l and h, each with a continuum of skills distributed over the same support $[\underline{w}, \overline{w}]$. Let $I_j^*(w)$, $j = l, h, hl$, denote the solution to equation (6) when $F_j(w)$ replaces $F(w)$. The optimal utility and consumption levels for every w-type individual are given by equations (15)–(17) when the two groups are pooled together, and by equations (20)–(22) under tagging.*

C Marginal income tax rates

On the basis of equation (12), and denoting $H(w, F_j)$ by $H_j(w)$, one can write the marginal income tax rates for the tagged and pooled solution as[6]

$$\frac{T_j'}{1 - T_j'} = H_j(w) \times \Omega_j(w), \quad j = l, h, hl. \tag{23}$$

[6]Kremer (2001) derives a similar result.

Observe that under tagging, the marginal tax rate for any given individual type depends only on the characteristics of the group to which he belongs (specifically on the elasticity of the labor supply and the hazard rate).[7] Second, *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. Third, 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). In all cases, the marginal income tax rate is characterized by equation (23).

We end this section by demonstrating a relationship between the hazard rates of the two groups and that of the entire population. This will be found useful later on.

Lemma 1 *Assume there are two groups of individuals, l and h, 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 $F_l(w)$ and $F_h(w)$. 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. It follows from equations (13)–(14) and the definition of $\Omega(w)$ that

$$\begin{aligned}\Omega_{hl}(w) &= \frac{1-F_{hl}(w)}{wf_{hl}(w)} = \frac{1-[F_l(w)+F_h(w)]/2}{w[f_l(w)+f_h(w)]/2} \\ &= \frac{f_l(w)}{f_l(w)+f_h(w)}\Omega_l(w) + \frac{f_h(w)}{f_l(w)+f_h(w)}\Omega_h(w). \qquad (24)\end{aligned}$$

■

[7] 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.

III Welfare and redistribution

There are two questions concerning redistribution. One is how tagging impacts the welfare of low- and high-ability individuals in the two groups; the second is the redistribution between groups. Ultimately, of course, it is the first question that underlies a government's tax policy. Tagging is done not with a view to transfer resources from one group to another; rather, it is undertaken in order to more efficiently transfer resources from high- to low-ability persons. We examine this question here, leaving the question of transfers between groups to the next section.

It is of course a trivial observation that tagging can never lower social welfare. One can always offer the pooled solution to the two groups. In our notation, $I^*_{hl}(w)$ and the corresponding, $\underline{u}^*_{hl}, u^*_{hl}(w), c^*_{hl}(w)$, constitute a feasible allocation to the optimal tax problem with tagging as well. Consequently, if the tagged solution differs from the untagged solution, the former must yield a better allocation (i.e., $\underline{u}^* > \underline{u}^*_{hl}$). Observe also that the first-stage of our two-stage optimal tax problem yields a solution $I^*_j(w) = I^*_{hl}(w)$, $j = l, h$, if $F_l(w) = F_h(w)$. 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.

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 leaving them equally well off.[8] However, one cannot a priori determine what happens to the welfare of the other individuals in either group. To shed light on this, we have to make other simplifying assumptions.

The first assumption we make is that labor supply exhibits a constant wage elasticity. Thus set $\varphi(L) = L^{1+1/\varepsilon}$, where ε is the labor supply elasticity.[9] Observe that the strict convexity of $\varphi(L)$ implies $\varepsilon > 0$. This assumption leads to a closed-form solution for

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

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

optimal incomes which, for $j = l, h, hl$, are given by[10]

$$I_j^*(w) = w^{1+\varepsilon}\left(\frac{\varepsilon}{1+\varepsilon}\right)^{\varepsilon}\left[1+\left(1+\frac{1}{\varepsilon}\right)\Omega_j(w)\right]^{-\varepsilon}. \tag{25}$$

With $\Omega_{hl}(w)$ being a weighted average of $\Omega_l(w)$ and $\Omega_h(w)$, equation (25) tells us that $I_{hl}^*(w)$ always lies between $I_l^*(w)$ and $I_h^*(w)$. Moreover, $I_h^*(w) \gtreqless I_l^*(w) \Leftrightarrow \Omega_h(w) \lesseqgtr \Omega_l(w)$. That is, $I_h^*(w) \gtreqless I_l^*(w)$ if and only if the hazard rate for the w-type in group h is greater than/equal to/smaller than the hazard rate for the w-type in group l. This makes perfect sense. With the hazard rate indicating the "proportion" of w-type in the segment of population with wages w and above, a higher hazard rate implies that we want to assign a higher income level to the w-types.

From the expression for $I_j^*(w)$ in (25), the definition of $\Omega_j(w)$, and the relationship between $F_l(w), F_h(w)$, and $F_{hl}(w)$, it then also follows that

$$\int_{\underline{w}}^{\overline{w}} I_l^*(w)\ f_l(w)dw + \int_{\underline{w}}^{\overline{w}} I_h^*(w)\ f_h(w)dw \geqq 2\int_{\underline{w}}^{\overline{w}} I_{hl}^*(w)\ f_{hl}(w)dw. \tag{26}$$

That is, tagging increases aggregate output.

The closed-form solution for $I_j^*(w)$ allows us to also derive the expressions for $u_{hl}^*(w)$, $c_{hl}^*(w)$, $u_j^*(w)$, and $c_j^*(w)$:

$$u_{hl}^*(w) = \underline{u}_{hl}^* + \left(\frac{\varepsilon}{1+\varepsilon}\right)^{\varepsilon}\int_{\underline{w}}^{w} s^{\varepsilon}\left[1+\left(1+\frac{1}{\varepsilon}\right)\Omega_{hl}(s)\right]^{-1-\varepsilon} ds, \tag{27}$$

$$c_{hl}^*(w) = u_{hl}^*(w) + \left(\frac{\varepsilon}{1+\varepsilon}\right)^{1+\varepsilon} w^{1+\varepsilon}\left[1+\left(1+\frac{1}{\varepsilon}\right)\Omega_{hl}(w)\right]^{-1-\varepsilon}, \tag{28}$$

$$u_j^*(w) = \underline{u}^* + \left(\frac{\varepsilon}{1+\varepsilon}\right)^{\varepsilon}\int_{\underline{w}}^{w} s^{\varepsilon}\left[1+\left(1+\frac{1}{\varepsilon}\right)\Omega_j(s)\right]^{-1-\varepsilon} ds,\quad j = l, h, \tag{29}$$

$$c_j^*(w) = u_j^*(w) + \left(\frac{\varepsilon}{1+\varepsilon}\right)^{1+\varepsilon} w^{1+\varepsilon}\left[1+\left(1+\frac{1}{\varepsilon}\right)\Omega_j(w)\right]^{-1-\varepsilon},\quad j = l, h. \tag{30}$$

It is clear from expressions (27)–(30) that one should know the relationship between $\Omega_h(w)$ and $\Omega_l(w)$ for every $w \in [\underline{w}, \overline{w}]$, to be able to determine who gains and who

[10] Setting $\varphi(L) = L^{1+\frac{1}{\varepsilon}}$ simplifies the expression for $\Psi(F_j)$ in (4). Rearranging the first-order condition for the maximization of $\Psi(F_j)$ with respect to $I_j(w)$ then yields (25).

loses from tagging. This, in turn, requires one to make assumptions on the distribution of skills. We assume that skills within each group have a lognormal distribution, and that group h has a higher mean wage than group l. Otherwise, the two lognormal distributions have identical other moments. Given the shape of lognormal distributions, when w is close to zero, the hazard rate is higher in group l. Thus, initially, $\Omega_l(w) < \Omega_{hl}(w) < \Omega_h(w)$. As w increases, however, $\Omega(w)$ decreases for lognormal distributions (at least initially). Two possibilities arise. First, $\Omega_h(w)$ may remain always above $\Omega_l(w)$.[11] In this case the $\Omega_l(w) < \Omega_{hl}(w) < \Omega_h(w)$ property is always satisfied. Second, it is *possible* that as w increases, at some point, $\Omega_h(w)$ crosses $\Omega_l(w)$. Interestingly, however, if this happens, $\Omega_h(w)$ will never curl back to cross $\Omega_l(w)$ again. That is, after the crossing, $\Omega_h(w)$ will always remain below $\Omega_l(w)$. We have the following lemma.

Lemma 2 *Assume the distribution of skills in the two groups is lognormal with one group, h, having a higher mean wage than the other group, l. Let $\Omega_j(w)$ denote the inverse of the hazard rate divided by w for distribution $F_j(w)$, where j stands for groups l, h, and hl (the two groups pooled together). Then, $\Omega_h(w)$ cannot cross $\Omega_l(w)$ more than once.*

Proof. 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

$$\begin{aligned}\frac{d\Omega}{dw} &= -\frac{1}{w} + \frac{\sqrt{2\pi}}{2\sigma w}(\ln w - \mu)\left[1 - \frac{2}{\sqrt{\pi}}\int_0^{\frac{\ln w - \mu}{\sigma\sqrt{2}}} e^{-t^2}dt\right] e^{\frac{(\ln w-\mu)^2}{2\sigma^2}} \\ &= \frac{1}{w}\left[(\ln w - \mu)\frac{\Omega}{\sigma^2} - 1\right]. \end{aligned} \tag{31}$$

[11] At high values of w, $f_l(w) < f_h(w)$ but $F_l(w) > F_h(w)$ so that the hazard rate in group h can remain always below the hazard rate in group l.

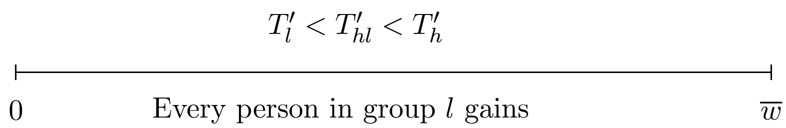

Figure 1: $\Omega_h(w)$ and $\Omega_l(w)$ do not intersect

Now denote the expression for Ω by Ω_l when $\mu = \mu_l$ and by Ω_h when $\mu = \mu_h$, where, by assumption, $\mu_h > \mu_l$. It then follows from equation (31) that

$$\frac{d}{dw}(\Omega_h - \Omega_l) = \frac{1}{w\sigma^2}\left[(\ln w - \mu_h)\,\Omega_h - (\ln w - \mu_l)\,\Omega_l\right]. \tag{32}$$

Equation (32) tells us that whenever $\Omega_h = \Omega_l (= \Omega)$,

$$\frac{d}{dw}(\Omega_h - \Omega_l) = \frac{\Omega}{w\sigma^2}\left[\mu_l - \mu_h\right] < 0.$$

But this can not happen two consecutive times. ■

Lemma 2 tells us that if a point w^D exists at which $\Omega_h(w) = \Omega_l(w)$, then $\Omega_l(w) < \Omega_{hl}(w) < \Omega_h(w)$ for all $w \in [\underline{w}, w^D)$ and $\Omega_l(w) > \Omega_{hl}(w) > \Omega_h(w)$ for all $w \in (w^D, \infty)$. Of course, we also know that if $\Omega_h(w)$ and $\Omega_l(w)$ do not cross, it is always the case that $\Omega_l(w) < \Omega_{hl}(w) < \Omega_h(w)$. Now, as long as $\Omega_l(w) < \Omega_{hl}(w) < \Omega_h(w)$, equations (27)–(29) imply that $u_l^*(w) - u_{hl}^*(w) > 0$ and $u_l^*(w) - u_h^*(w) > 0$. 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 l also become better off. Observe, however, that one does not know how $u_h^*(w)$ and $u_{hl}^*(w)$ compare because while $\underline{u}^* > \underline{u}_{hl}^*$, the second expression in $u_h^*(w)$ is smaller than the second expression in $u_{hl}^*(w)$. This means that it is possible for a person with a low w in group h to become worse off as a result of tagging. Similarly, we have from (28)–(30), $c_l^*(w) - c_{hl}^*(w) > 0$, and $c_l^*(w) - c_h^*(w) > 0$; but that the comparison between $c_h^*(w)$ and $c_{hl}^*(w)$ is ambiguous.

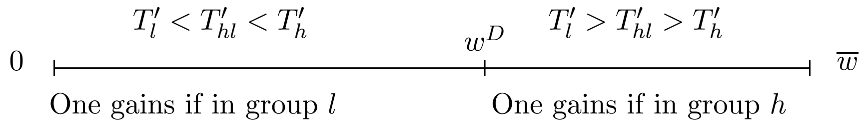

Figure 2: $\Omega_h(w)$ and $\Omega_l(w)$ intersect

Finally, observe that with a constant and identical elasticity of labor supply, the marginal income tax rate formula in (23) is simplified to

$$\frac{T_j'}{1-T_j'} = \left(1+\frac{1}{\varepsilon}\right)\Omega_j(w), \qquad j = l, h, hl. \tag{33}$$

It then immediately follows from (24) 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_l', T_h', and T_{hl}':

$$\frac{T_{hl}'}{1-T_{hl}'} = \frac{f_l(w)}{f_l(w)+f_h(w)}\frac{T_l'}{1-T_l'} + \frac{f_h(w)}{f_l(w)+f_h(w)}\frac{T_h'}{1-T_h'}. \tag{34}$$

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. Equation (34) implies that $T_l'(I(w)) < T_{hl}'(I(w)) < T_h'(I(w))$.

If the hazard rates intersect at w^D, equations (27)–(30) imply that for all $w \in (w^D, \infty)$, $u_h^*(w) - u_{hl}^*(w) > 0$, $u_h^*(w) - u_l^*(w) > 0$, $c_h^*(w) - c_{hl}^*(w) > 0$, and $c_h^*(w) - c_l^*(w) > 0$; but we do not know how $u_l^*(w)$ compares with $u_{hl}^*(w)$ and $c_l^*(w)$ with $c_{hl}^*(w)$. That is, tagging may have the unintended consequence of making even the richer individuals in the rich group h better off. Finally, we also have that $T_l'(I(w)) > T_{hl}'(I(w)) > T_h'(I(w))$.

Figures 1–2 illustrate the two possible cases when $\Omega_h(w)$ and $\Omega_l(w)$ do not intersect and when they do, and Proposition 2 summarizes our results.

Proposition 2 *In addition to the assumptions of Proposition 1, assume that the wage elasticity of labor supply is constant and identical for the groups h and l. Then* *(i) Income of w-type workers in l and h groups under the tagged solution bracket the income that they would earn if the entire population is pooled.*

(ii) Tagging increases aggregate output.

Additionally, assume that the distribution of skills in the two groups is lognormal with group h having a higher mean than group l. Then:

(iii) Either all individuals in group l or those with earning ability $w \in [\underline{w}, w^D)$, assuming there exists a w^D at which the hazard rates for group l and group h intersect, benefit from tagging (in terms of consumption and utility levels and compared to the pooled equilibrium solution). These individuals also earn more, consume more, and have higher utility levels than corresponding w-type persons in group h. Similar individuals in group h may lose, as well as gain.

(iv) Every individual $w \in (w^D, \infty)$ in group h, assuming there exists a w^D at which the hazard rates for group l and group h intersect, benefit from tagging. These individuals also earn more, consume more and have higher utility levels than corresponding w-type persons in group l. Similar individuals in group l may lose, as well as gain.

(v) The pattern of marginal income tax rates are $T_l'(I(w)) < T_{hl}'(I(w)) < T_h'(I(w))$ either for all w, or for all $w \in [\underline{w}, w^D)$. In the latter case, $T_l'(I(w)) > T_{hl}'(I(w)) > T_h'(I(w))$ for all $w \in (w^D, \overline{w}]$.

IV Redistribution between groups

We now turn to the question of redistribution between groups. Interestingly, the result we will derive below holds for a setting more general than what we have considered thus far.

The basic idea of tagging is that if one group contains relatively more disadvantaged people whom the government wants to help, then by “favoring” that group one succeeds in redistribution towards the disadvantaged people in more effective way. In the context of our model, the question is whether tagging entails redistribution from group h to group l. Akerlof’s (1978) original example was based on the assumption that one of the groups consisted of only low-ability individuals. It was then by way of redistributing towards this group that he was able to show that every member of the group received more transfers and became better off (as compared to the pooled equilibrium solution).[12] Note that redistribution from the not-favored to the favored group is essentially a redistribution from the total population to the favored group.

To shed light on this particular question, again one has to know more about the skills distributions beyond the fact that they are different. However, one need not resort to log-normality of skills assumption. The question can be answered for any distribution of skills as long as one stochastically dominates the other. The direction of transfers will be from the dominant group to the dominated one. Interestingly too, quasi-linearity of preferences is not needed for this result as long as the social welfare function is Rawlsian. On the other hand, with quasi-linearity, the result holds with more general social welfare functions. Proposition 3 states and proves these results.

Proposition 3 *There are two groups of individuals of equal size, l and h, each with a continuum of skills distributed according to the distribution functions $F_l(w)$ and $F_h(w)$ over the support $w \in [\underline{w}, \overline{w}]$. Group h first-order stochastically dominates group l so that $F_h(w) \leq F_l(w)$ for all $w \in [\underline{w}, \overline{w}]$ and $F_h(w) < F_l(w)$ for some w. Then tagging implies redistribution from group h to group l under either of these conditions:*

(i) The social welfare function is Rawlsian.

(ii) Preferences are quasi-linear and the social welfare function is weighted utilitarian, $\int_{\underline{w}}^{\overline{w}} \gamma(w) V(w) f(w)\, dw$.

Proof. Part (i): Under a Rawlsian social welfare function, the objective of the government is to maximize tax revenues subject to incentive compatibility and resource constraints. Now observe that incentive compatibility does not depend on the distribution of types so that the allocation $(I_l^*(w), c_l^*(w))$ is incentive compatible in h as well

[12] However, one could not determine the impact that tagging had on the welfare of low-ability persons in the not-favored group. 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.

as in l. But, given that $(I_h^*(w), c_h^*(w))$ is the optimal solution in h, we must have[13]

$$\int_{\underline{w}}^{\overline{w}} [I_h^*(w) - c_h^*(w)]\, dF_h(w) \geqq \int_{\underline{w}}^{\overline{w}} [I_l^*(w) - c_l^*(w)]\, dF_h(w). \tag{35}$$

Now 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

$$\int_{\underline{w}}^{\overline{w}} [I_l^*(w) - c_l^*(w)]\, dF_h(w) > \int_{\underline{w}}^{\overline{w}} [I_l^*(w) - c_l^*(w)]\, dF_l(w). \tag{36}$$

Inequalities (35)–(36) then imply that

$$\int_{\underline{w}}^{\overline{w}} [I_h^*(w) - c_h^*(w)]\, dF_h(w) > \int_{\underline{w}}^{\overline{w}} [I_l^*(w) - c_l^*(w)]\, dF_l(w). \tag{37}$$

This means that the government can extract more tax revenue from group h than from group l so that the direction of transfers is from group h to group l.

Part (ii): The government maximizes $\int_{\underline{w}}^{\overline{w}} \gamma(w) V_j\left(w\right) f_j\left(w\right) dw$ for each group subject to local incentive compatibility constraint and the revenue requirement $\overline{R}_j$ in the first stage, and then equalizes the shadow cost of public funds across the groups in the second stage. Denote the optimized value of $V_j\left(w\right)$ by $V_j^*\left(w\right)$ and the Lagrange multipliers associated with the revenue requirement by μ_j. We have

$$\mu_j = -\frac{d}{d\overline{R}_j}\int_{\underline{w}}^{\overline{w}} \gamma(w) V_j^*\left(w\right) f_j\left(w\right) dw.$$

Thus there will be redistribution from group h to group l if and only if $\mu_l > \mu_h$. Now, with quasi-linear preferences, $dV_j^*\left(w\right)/d\overline{R}_j = -1$. Substituting in the above equation

[13] Remember that with our assumptions the optimal tax schedule maximizes tax revenues.

for μ_j and integrating by parts, one can write the expression for μ_j as

$$\mu_j = \int_{\underline{w}}^{\overline{w}} \gamma(w) f_j(w)\, dw = \gamma(\overline{w}) - \int_{\underline{w}}^{\overline{w}} \gamma'(w) F_j(w)\, dw.$$

Subtracting μ_h from μ_l,

$$\mu_l - \mu_h = \int_{\underline{w}}^{\overline{w}} \gamma'(w) \left[F_h(w) - F_l(w)\right] dw.$$

The assumptions that $F_h(w)$ first-order stochastically dominates $F_l(w)$ and $\gamma'(w) < 0$ imply that $\mu_l - \mu_h > 0$. ■

V Simulations

This section applies our theoretical model to the taxation of US workers using PSID data for 1993 with gender as tag. Assume that the distribution of skills in each of the group of workers is truncated lognormal over the same support. The computed mean logarithmic deviation of income is 0.25 for men and 0.20 for women. Configure the means of the skills distributions in such a way that the median gross income will be $30,000 for men and $20,000 for women. Also assume, in line with our theoretical construct, that the wage elasticity of labor supply is constant and identical for the two tagged groups; setting its value at 0.5 which is a figure used in this literature.

Figures 3 presents the graph of the optimal marginal income tax rates for men versus women. Observe that, the two tagged groups' marginal tax rates bracket the pooled population's tax rate at every income level with men facing the higher rates. At the income level of $10,000, the marginal income tax rate is 80% for men, 26% for women, and 50% for the pooled population. The rates decrease all the way to zero for the top income level of $175,000.

Unlike the marginal income tax rates, average tax rates are increasing in income indicating the progressivity of the tax schedules; see Figure 4. The rates are close across the groups up to an income level of $20,000 after which the tax rate for the

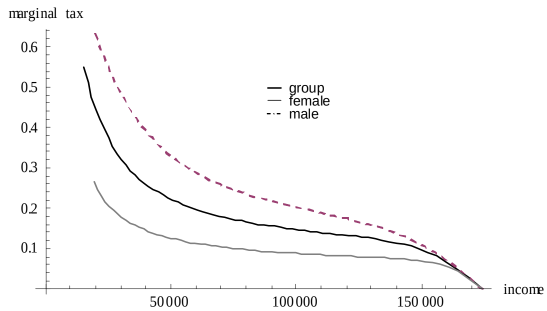

Figure 3: Optimal marginal income tax rates by gender.

pooled population is bracketed by those of men (at a higher rate) and women (at a lower rate).

The higher average tax rates faced by male high-wage earners show that they are the losers in the tagging procedure. To get a clearer picture of these welfare effects, Table 1 reports how men and women are affected by tagging at different decile levels of skills. Observe first that tagging improves the welfare of the least well-off male and female workers alike by \$1,174 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 6%. At higher wage levels, women gain at all deciles. The gains range from \$1,677 per year for a female worker at the lowest decile to \$13,728 per year at the highest decile. Low-wage male workers also gain with those at the first decile gaining by as much as \$935 per year. The gains decrease for higher deciles, turning negative for the fifth. After that, male workers lose more and more as their wage increases, with the loss increasing to \$8,397 per year at the tenth decile.

Given our Rawlsian perspective, the gains and losses to workers beyond least well-

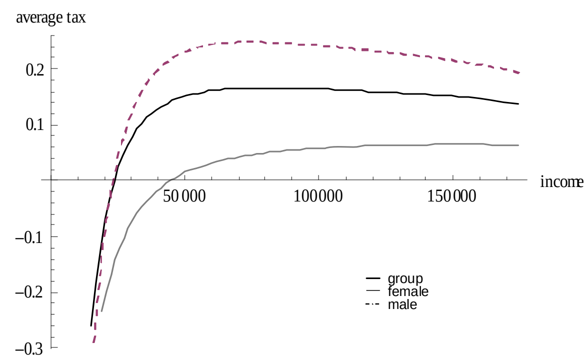

Figure 4: Optimal average income tax rates by gender.

off individuals are not part of the calculus of social welfare. The improvement in social welfare is the $1,174 per year gain enjoyed by the least well-off workers, females and males alike. However, that *all* low-wage workers gain, regardless of their tag, attests to the "efficiency-enhancing" power of tagging.

VI 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, and every person having the same constant wage elasticity of labor supply, we have been able to derive a set of analytical results for optimal income taxation with tags. Secondly, two additional assumptions, namely log-normality of the skills distribution and a constant elasticity of labor supply,

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

	female		male	
	Utility change %	Compensating variation	Utility change %	Compensating variation
Least well-off	6	1,174	6	1,174
Skills deciles				
1	9	1,677	5	935
2	11	2,132	3	671
3	14	2,562	2	384
4	15	2,982	0	71
5	17	3,409	−1	−271
6	19	3,859	−3	−655
7	20	4,359	−5	−1,101
8	21	4,961	−7	−1,652
9	23	5,818	−9	−2,444
10	15	13,728	−9	-8,397

have allowed us to analytically identify the winners and losers of tagging.

Specifically, we have shown that if the hazard rates in the two tagged groups do not cross, *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. If the hazard rates cross, the direction of gainers and losers change after the crossing.

Third, we have demonstrated 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. This result holds for a general utility function as long as the social welfare function is Rawlsian. It also holds for a utilitarian social welfare function with decreasing weights in skills provided that preferences are quasi-linear.

Finally, we have studied the implications of tagging on the basis of gender, assuming that the skills distribution in each group is truncated lognormal. The distributions are calibrated such that average earnings and mean logarithmic deviation of income in each group are those of the US in 1993 according to the PSID data. We have found that optimal marginal income tax rates are about 80% for men and 26% for women at the $10,000 income level, steadily decreasing to zero at an income level of $175,000.

We have also found that tagging has dramatic impacts on welfare. The welfare of the least well-off male and female workers increase by the equivalent of $1,174 per year (6% in utility terms). 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. Most interestingly, *all* low-wage workers gain regardless of their tag. The only people who lose are male high-earners whose losses are more than compensated by the gains to females of equal ability.

In interpreting these numbers, the shortcomings of our calibrated model should be borne in mind. The distributions are calibrated to reflect the differences in average earnings and mean logarithmic deviation of incomes across groups; however, we have not allowed for differential labor supply responses. Yet the finding that all low-wage workers gain suggests that tagging enhance the efficiency of the tax system considerably. This merits further investigation.

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

[4] Blomquist, Sören and Luca Micheletto (2008). Age related optimal income taxation. Scandinavian Journal of Economics, vol. 110, 45–71.

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

[8] 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.

[9] 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.

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

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

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

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

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

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

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

[17] 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.

[18] Weinzierl, Matthew. 2008. The surprising power of age-dependent taxes. Unpublished manuscript, Harvard University.

Appendix

Derivation of the expression for $I_j^*(w)$ in equation (25): Setting $h(L) = L^{1+\frac{1}{\varepsilon}}$ simplifies the expression for $\Psi(F_j)$ in (4), $j = l, h, hl$, to

$$\Psi(F_j) = \left\{ I_j(w) - \left[1 + \left(1 + \frac{1}{\varepsilon}\right) \Omega_j(w)\right] \left(\frac{I_j(w)}{w}\right)^{1+\frac{1}{\varepsilon}} \right\} f_j(w)dw. \tag{A1}$$

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

$$I_j^*(w) = wL_j^*(w) = w^{1+\varepsilon} \left(\frac{\varepsilon}{1+\varepsilon}\right)^{\varepsilon} \left[1 + \left(1 + \frac{1}{\varepsilon}\right) \Omega_j(w)\right]^{-\varepsilon}, \quad j = l, h, hl. \tag{A2}$$

Derivation of the expressions for $\underline{u}_{hl}^*$ in (27) and $\underline{u}^*$ in (29): First, substitute the expression for $I_j^*(w)$ from (A2) into (A1) and simplify to get, for $j = l, h, hl$,

$$\Psi^*(F_j) = \varepsilon^{\varepsilon} (1+\varepsilon)^{-1-\varepsilon} \int_{\underline{w}}^{\overline{w}} w^{1+\varepsilon} \left[1 + \left(1 + \frac{1}{\varepsilon}\right) \Omega_j(w)\right]^{-\varepsilon} f_j(w)dw. \tag{A3}$$

Second, given the expression for $\Psi^*(F_{hl})$ and using (15), we have

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

Finally, given the expression for $\Psi^*(F_j)$, $j = l, h$, and using (20), we have

$$\begin{aligned} \underline{u}^* \;=\; & \frac{1}{2}\varepsilon^{\varepsilon} (1+\varepsilon)^{-1-\varepsilon} \left\{ \int_{\underline{w}}^{\overline{w}} w^{1+\varepsilon} \left[1 + \left(1 + \frac{1}{\varepsilon}\right) \Omega_h(w)\right]^{-\varepsilon} f_h(w)dw \right. \\ & \left. + \int_{\underline{w}}^{\overline{w}} w^{1+\varepsilon} \left[1 + \left(1 + \frac{1}{\varepsilon}\right) \Omega_l(w)\right]^{-\varepsilon} f_l(w)dw \right\}. \end{aligned}$$

Derivation of equations (27)–(30): With $h(L) = L^{1+\frac{1}{\varepsilon}}$, we have, for $j = l, h, hl$,

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

Substitute for $L_j^*(w)$ from (A2) 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 $u_{hl}^*(w)$ and $u_j^*(w)$, $j = l, h$, derived in (16) and (21), to (27) and (29). Similarly, one simplifies the expressions for $c_{hl}^*(w)$ and $c_j^*(w)$, $j = l, h$, derived in (17) and (22), to (28) and (30).

Proof of inequality (26): Let

$$\begin{aligned} \Delta \;&\equiv\; \int_{\underline{w}}^{\overline{w}} I_h^*(w)\ f_h(w)dw + \int_{\underline{w}}^{\overline{w}} I_l^*(w)\ f_l(w)dw - 2\int_{\underline{w}}^{\overline{w}} I_{hl}^*(w)\ f_{hl}(w)dw \\ &=\; \int_{\underline{w}}^{\overline{w}} \left[I_h^*(w) - I_{hl}^*(w)\right] f_h(w)dw + \int_{\underline{w}}^{\overline{w}} \left[I_l^*(w) - I_{hl}^*(w)\right] f_l(w)dw. \end{aligned}$$

Substitute for $I_i^*(w)$, $j = l, h, hl$, from (A2) in above and simplify. We have

$$\begin{aligned} &\Delta\left(\frac{1+\varepsilon}{\varepsilon}\right)^{\varepsilon} = \\ &\int_{\underline{w}}^{\overline{w}} w^{1+\varepsilon}\left\{\left[1+\left(1+\frac{1}{\varepsilon}\right)\Omega_h(w)\right]^{-\varepsilon} - \left[1+\left(1+\frac{1}{\varepsilon}\right)\Omega_{hl}(w)\right]^{-\varepsilon}\right\} f_h(w)dw + \\ &\int_{\underline{w}}^{\overline{w}} w^{1+\varepsilon}\left\{\left[1+\left(1+\frac{1}{\varepsilon}\right)\Omega_l(w)\right]^{-\varepsilon} - \left[1+\left(1+\frac{1}{\varepsilon}\right)\Omega_{hl}(w)\right]^{-\varepsilon}\right\} f_l(w)dw. \end{aligned} \tag{A4}$$

Now subtract the expression for $\underline{u}_{hl}^*$ from the expression for $\underline{u}^*$:

$$\begin{aligned} \underline{u}^* - \underline{u}_{hl}^* \;=\; & \frac{1}{2}\varepsilon^{\varepsilon}(1+\varepsilon)^{-1-\varepsilon}\left\{\int_{\underline{w}}^{\overline{w}} w^{1+\varepsilon}\left[1+\left(1+\frac{1}{\varepsilon}\right)\Omega_h(w)\right]^{-\varepsilon} f_h(w)dw + \right. \\ & \left.\int_{\underline{w}}^{\overline{w}} w^{1+\varepsilon}\left[1+\left(1+\frac{1}{\varepsilon}\right)\Omega_l(w)\right]^{-\varepsilon} f_l(w)dw\right\} - \\ & \varepsilon^{\varepsilon}(1+\varepsilon)^{-1-\varepsilon}\int_{\underline{w}}^{\overline{w}} w^{1+\varepsilon}\left[1+\left(1+\frac{1}{\varepsilon}\right)\Omega_{hl}(w)\right]^{-\varepsilon} f_{hl}(w)dw \end{aligned}$$

$$
\begin{aligned}
= \quad & \frac{1}{2}\varepsilon^{\varepsilon}\left(1+\varepsilon\right)^{-1-\varepsilon}\left\{\int_{\underline{w}}^{\overline{w}} w^{1+\varepsilon}\left[1+\left(1+\frac{1}{\varepsilon}\right)\Omega_h(w)\right]^{-\varepsilon} f_h(w)dw - \right. \\
& \left. \frac{1}{2}\int_{\underline{w}}^{\overline{w}} w^{1+\varepsilon}\left[1+\left(1+\frac{1}{\varepsilon}\right)\Omega_{hl}(w)\right]^{-\varepsilon}\left[f_h(w)+f_l(w)\right]dw\right\} + \\
& \frac{1}{2}\varepsilon^{\varepsilon}\left(1+\varepsilon\right)^{-1-\varepsilon}\left\{\int_{\underline{w}}^{\overline{w}} w^{1+\varepsilon}\left[1+\left(1+\frac{1}{\varepsilon}\right)\Omega_l(w)\right]^{-\varepsilon} f_l(w)dw - \right. \\
& \left. \frac{1}{2}\int_{\underline{w}}^{\overline{w}} w^{1+\varepsilon}\left[1+\left(1+\frac{1}{\varepsilon}\right)\Omega_{hl}(w)\right]^{-\varepsilon}\left[f_h(w)+f_l(w)\right]dw\right\} \\
= \quad & \frac{1}{2}\varepsilon^{\varepsilon}\left(1+\varepsilon\right)^{-1-\varepsilon}\int_{\underline{w}}^{\overline{w}} w^{1+\varepsilon}\left\{\left[1+\left(1+\frac{1}{\varepsilon}\right)\Omega_h(w)\right]^{-\varepsilon} - \right. \\
& \left.\left[1+\left(1+\frac{1}{\varepsilon}\right)\Omega_{hl}(w)\right]^{-\varepsilon}\right\} f_h(w)dw + \frac{1}{2}\varepsilon^{\varepsilon}\left(1+\varepsilon\right)^{-1-\varepsilon}\int_{\underline{w}}^{\overline{w}} w^{1+\varepsilon}\times \\
& \left\{\left[1+\left(1+\frac{1}{\varepsilon}\right)\Omega_l(w)\right]^{-\varepsilon} - \left[1+\left(1+\frac{1}{\varepsilon}\right)\Omega_{hl}(w)\right]^{-\varepsilon}\right\} f_l(w)dw. \qquad \text{(A5)}
\end{aligned}
$$

Substituting from (A5) into (A4), we get

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