

The Normative Analysis of ‘Tagging’ Revisited: Dealing with Stigmatization*

Laurence Jacquet[†] and Bruno Van der Linden[‡]

February 20, 2006

Abstract

Should income transfers be conditional upon personal characteristics of the potential recipients (so-called ‘tagging’) or should they only be tied to reported incomes? A still widespread social norm consists in requiring that individuals (of working age) support themselves and their families. Being a welfare recipient is then socially disapproved because it reveals to others that one is unable to fend for oneself and one’s family. In this context, tagging is always suboptimal under a maxi-min criterion. With a utilitarian criterion, tagging can only be recommended if the distribution of the intensity of stigmatization relative to earnings has small mean and variance and if the mean and variance of the distribution of abilities among the high-ability people are neither too large nor too small.

Key words : tagging, optimal taxation, welfare programs, stigmatization.

JEL Classification : H2, I3

1 Introduction

It is typically argued that a combination of income taxation and testing (‘tagging’) of social security payments, as with a targeted minimum income guarantee, leads to a better trade-off between the incentive costs of distortionary income taxes and the social welfare gains of redistribution. If tagging does not involve any monetary cost and can prevent uneligible people from benefiting from the targeted transfers, society is necessarily better off by its use since it enables a redistribution from the untagged to the tagged group without creating mimicking effects or work disincentives among the former group. This view is challenged in this paper which introduces the demeaning or stigmatizing effect of tagging.

*The editor, Wolfram Richter and the comments of two anonymous referees are gratefully acknowledged. We also deeply thank Maurice Marchand and Linda Andersson, Robin Boadway, Bart Cockx, Kate Cuff, Jean-Yves Duclos, Ginette Herman, Jean Hindriks, Etienne Lehmann and Philippe Van Parijs for very useful discussions and comments. Any remaining error could only be attributed to the authors. The authors acknowledge financial support from the Belgian Federal Government (Grant PAI P5/21, “Equilibrium theory and optimization for public policy and industry regulation”)

[†]Belgian National Scientific Fund (FNRS), Department of economics, Université Catholique de Louvain.

[‡]Department of economics, Université Catholique de Louvain, FNRS and IZA.

The literature on optimal income taxation (see the seminal papers by Mirrlees, 1971; and Stiglitz, 1987) characterizes the optimal shape of marginal tax rates under imperfect information about the earning ability of the tax payers. This approach has been used to recommend a form of non linear negative income tax (Boadway and Keen, 2000). A negative income tax is a refundable tax credit with an explicit tax schedule which can be, but need not by definition be, linear (Cournot, 1838; Friedman, 1962; Tobin, 1966; 1968). Another strand of the literature is concerned with optimal welfare programs (Akerlof, 1978; Diamond and Sheshinski, 1995; Parsons, 1996; Boadway, Marceau and Sato, 1999; Salanié, 2002). In this literature, the welfare agencies are assumed to observe more than only reported income levels and the aggregate distribution of abilities. The agencies condition (tag) transfers on personal characteristics (such as disability or ‘employability’) of potential recipients, that provide some imperfect information about the earning ability. Akerlof (1978) deals with errors of type I (some of those who are entitled to the benefits are rejected) and Parsons (1996) adds errors of type II (individuals not entitled are accepted). Both argue that tagging increases aggregate welfare. Moreover, as long as the disutility of work is not too large, Parsons (1996) shows that the optimal system should let tagged people work. In Akerlof (1978), Diamond and Sheshinski (1995), Parsons (1996), and Salanié (2002), tagging is costless. The accuracy of the tag and therefore the probabilities of errors are taken as given. Boadway, Marceau and Sato (1999) emphasize the role of social workers whose (imperfectly observable) effort affects the magnitude of these errors and induces administrative costs. They show that the choice between transferring income to the poor through tagging or exclusively via a non linear negative income tax system depends on the magnitude of administrative costs relative to the benefits of targeting.

The use of tagging can be questioned due to the evidence that non-take-up is important. Many reasons can be invoked to explain this phenomenon: imperfect information among the eligible population, lack of literacy or numeracy, transaction costs related in particular to the time spent queuing and filling out forms, the loss of privacy of claiming benefits that require an assessment of personal characteristics (social workers evaluate the eligibility of claimants through enquiries and tests of a searching and detailed kind), the demeaning or stigmatizing effect of applying and receiving such benefits. Moffitt (1983) provides an econometric test for stigma in the Aid to Families with Dependent Children program. His results show “definite evidence of a stigma-related disutility of participation” (p. 1024) in welfare programs. Along these lines, an empirical literature has studied the non-take-up of various types of welfare or means-tested benefits. Stigma is mentioned among the non-pecuniary participation costs (see, e.g., Ashenfelter, 1983; Moffitt, 1983; Blundell, Fry and Walker, 1988; Blank and Ruggles, 1996; Duclos, 1995; 1997; Hancock, Pudney, Barker,

Hernandez and Sutherland, 2004; and Currie, 2004).

In this paper, we emphasize stigma as an explanation to the non-take-up phenomenon. This focus is motivated by the growing evidence that stigma is important and by the relative lack of interest for this explanation in the economic literature.

Those who apply and receive assistance benefit suffer from a loss of utility due to stigmatization. “Stigmatized individuals possess (or are believed to possess) some attribute, or characteristic, that conveys a social identity that is devaluated in some particular social context” (Crocker, Major and Steele, 1998). Rules (or norms) about who is a member of the devaluated category are defined by society. A still widespread social norm consists in requiring that individuals (of working age) support themselves and their families. Being a welfare recipient is then socially disapproved because it reveals to others that one is unable to fend for oneself and one’s family. This phenomenon has been stressed by Sen (1995) who writes: “Any system of subsidy that requires people to be identified as poor and that is seen as special benefaction for those who cannot fend for themselves would tend to have some effects on their self-respect as well as on the respect accorded them by others (...) there are (also) direct costs and losses involved in feeling —and being— stigmatized.” In a similar perspective, Yaniv (1997) writes: “The shame, embarrassment, and social disapproval afflicting a claimant whose participation in a welfare program is *observed* by others or becomes known to significant others (i.e., family, friends neighbors, employers, etc) have been strongly stressed by sociologists addressing the issue of welfare stigma (e.g., Goffman, 1963; Waxman, 1983; Spicker, 1984)”. It is important to notice that stigmatization would not occur if redistribution was only based on income reported to the tax authority. For, the handling of income reports by tax authority can be considered as fairly anonymous: Receiving a transfer from the tax authority (e.g., a refundable tax credit) does not require to reveal to other people that one is unable to fend for oneself.

Our theoretical setting is close to the one presented in Akerlof (1978). There are two types of workers, the low-ability workers (whose productivity, at the limit, can be zero - the “disability case”) and the high-ability ones. As in the standard optimal taxation literature, each productive individual has access to a job which remunerates him according to his productivity. As in Akerlof (1978), the targeted transfer, if any, is added to the labor earnings of tagged people. This contrasts with Parsons (1996) and Boadway, Marceau and Sato (1999). In their models, the population is subdivided between (non-working) disable and able people and only the latter are able to work. We deliberately neglect errors of type II and administrative costs linked to imperfect monitoring of social workers. By assumption, the latter do costlessly observe the ability of workers and

can prevent high-ability workers from benefiting from the assistance scheme. Under these rather extreme assumptions, it is expected that the combination of income taxation designed by the tax authority and an assistance scheme targeted on the less able is superior to using only (non linear) negative income tax. This paper shows that the introduction of stigmatization challenges the well-known superiority of costless tagging. As Besley and Coate (1992) notice, there is no reason to believe that claiming welfare benefits will affect all individuals in the same way. Therefore, in our model, low-ability workers are endowed with an individual-specific parameter that measures the impact of stigma on well-being. This parameter is exogenously distributed in the population. Low-ability people are aware of their eligibility, but part of them, possibly, do not claim assistance benefits because they prefer not to be stigmatized.

To the best of our knowledge, stigmatization has never been considered in an optimal income taxation framework. Some economists have modeled the decision to apply for welfare when the application process entails a disutility that can be attributed to one of the above reasons (see Bishop, 1982; or Moffitt, 1983). Besley and Coate (1992) model stigmatization which results from people's disapproval of and resentment against those who choose to go on welfare (and are considered either as responsible for their recipient status or as fraudulent recipients) and whose benefits are financed by mandatory taxation. Yaniv (1997) studies how the decision of taking up welfare benefits is affected by stigmatization (generated through observed participation in welfare programs) and/or by the expected penalty for underreporting income to qualify for welfare benefits. Redistribution mechanisms based on tagging also raise issues of horizontal equity (which requires that those with equal status, whether measured by ability or some other appropriate scale, should be treated the same) and of political feasibility (Sen, 1995). The more targeted the transfers towards the truly needy, the lower the support from the excluded middle income class. This can have detrimental effects on the actual level of redistribution (Gelbach and Pritchett, 1996; De Donder and Hindriks, 1998).

The paper is organized as follows. Section 2 introduces assumptions and notations. Analytical results are presented in Section 3. Since this normative analysis does not lead to clear-cut conclusions in the utilitarian case, a numerical analysis is summarized in Section 5. Section 6 concludes the paper.

2 Assumptions and notations

By assumption, workers supply one unit of labor. Let us consider two types of individuals who differ in terms of their abilities indexed by j . Moreover, there are two types of jobs and pretax earnings are specific to the job. The low-ability workers ($j = l$) are characterized

by exogenous gross or pretax earnings, $w_l \geq 0$.¹ If $w_l = 0$, these individuals are by assumption unable to work (say, because of a disability). A high-ability worker ($j = h$) can either occupy a job designed for his ability (labelled a “skilled job”) or he can have an occupation requiring low abilities (labelled an “unskilled job”). Because of their low ability, type- l workers cannot perform skilled jobs. Gross earnings in skilled jobs are given and denoted by w_h , with $w_h > w_l$. Compared to unskilled jobs, we assume that performing skilled work requires more effort. In this respect, high-ability workers are assumed to be heterogeneous. If they occupy a skilled job, their consumption level is c_h and their level of utility is $u(c_h) - \delta$, where δ is the disutility of working in a skilled job rather than an unskilled one. By assumption, δ is distributed on the interval $[0, +\infty[$ ² according to the cumulative distribution $F(\delta)$ and the density function $f(\delta)$. While δ is an individual attribute that no other agent can observe, $F(\delta)$ is common knowledge. By assumption, $f(\delta) > 0 \forall \delta \in [0, +\infty[$ with $\lim_{\delta \rightarrow +\infty} f(\delta) = 0$. $u(c)$ is a continuous, differentiable, strictly increasing and strictly concave function with $\lim_{c \rightarrow 0} u'(c) = +\infty$. If instead a high-ability worker chooses an unskilled job, his consumption level is c_l and his utility level $u(c_l)$.

As is usually assumed, the tax authority knows the proportion of individuals with low ability γ ($0 \leq \gamma \leq 1$) but it does not observe the ability of a given individual. It perfectly and costlessly observes reported income. As is standard in the optimal taxation literature, the occupational choice of the high-ability workers will limit the extent of redistribution which the tax authority can implement. To relax this self-selection constraint, let us introduce welfare agencies. They occupy social workers who have access to more information than the tax authority. To adopt a simple setting where tagging would be advocated in the absence of stigmatization, let us assume idealized welfare agencies which costlessly and perfectly can assess the ability of claimants. In this idealized setting, welfare agencies can separate low-ability claimants from high-ability individuals with high disutility δ . Welfare agencies only provide a targeted transfer to low-ability individuals. Not all of these individuals will however claim the targeted benefit. This is not because of a lack of information or literacy. They are instead deterred by the losses involved in feeling and being stigmatized. The level of utility of low-ability individuals who choose to

¹As it is standard in the optimal tax literature, employers perfectly and costlessly observe the ability of the worker. This is however not the source of stigmatization. As we explain above, the latter is due to the participation to a welfare program.

²The assumption that δ is non negative could be criticized if δ is interpreted in a broader sense. The occupation requiring low abilities could be interpreted as poor jobs because they are the only type of occupation accessible to low-ability workers. The support of δ would then probably include negative values in order to deal with individuals for whom the negative status of these unskilled jobs outweighs the penibility of effort required to perform a skilled job. This interpretation would however not change the nature of our results as long as the support of δ is not restricted to negative values. So, we stick to the case where δ lies in $[0, +\infty[$.

claim assistance benefits is written as:

$$u(c_l^T) - \sigma \tag{1}$$

where c_l^T equals w_l plus the targeted transfer net of tax liabilities, if any (superscript T denotes the tagged status). Here, σ is an individual parameter that represents the intensity of stigma. There is no reason to believe that claiming welfare benefit will affect all individuals in the same way. So, σ is by assumption distributed on the interval $[0, +\infty[$, according to the cumulative distribution $G(\sigma)$ and its associated density function $g(\sigma)$. It is assumed that $g(\sigma) > 0 \forall \sigma \in [0, +\infty[$ with $\lim_{\sigma \rightarrow +\infty} g(\sigma) = 0$. It is also assumed that the individual value of σ is unknown by all except the person herself, while $G(\sigma)$ is common knowledge. The utility level of the low-ability individuals who do not claim assistance benefit is equal to

$$u(c_l) \tag{2}$$

where c_l denotes their consumption level.

The choice of taxes and transfers is equivalent to the determination of consumption bundles c_l^T , c_l and c_h . Since the unobservable parameters δ and σ are distributed on a wide (formally of infinite size) support, it simply becomes too costly to induce all high-ability individuals to work in skilled jobs³ and to induce all type l individuals to claim the targeted transfer. So, whatever the allocation of consumption levels, there will be some finite cut-off levels $\tilde{\delta}$ and $\tilde{\sigma}$ such that only those high-ability individuals with $\delta < \tilde{\delta}$ are occupied in skilled jobs and only those with $\sigma < \tilde{\sigma}$ and a low ability opt for the targeted transfer. Therefore, in the subpopulation with net income c_l , one necessarily finds some low-ability people and the high-ability ones with a sufficiently large disutility of effort. These cut-off values or threshold levels satisfy the following equalities:

$$u(c_h) - \tilde{\delta} = u(c_l) \tag{3}$$

$$u(c_l^T) - \tilde{\sigma} = u(c_l) \tag{4}$$

Table 1 displays the proportions of individuals in each position and Figure 1 summarizes the model.

TABLE 1 ABOUT HERE

FIGURE 1 ABOUT HERE

³This element distinguishes our model from standard models of redistribution which rely on the revelation principle to induce all individuals to report their true types in an optimum.

Let us now turn to the normative criteria used in this paper. This paper compares outcomes under a utilitarian criterion and a maxi-min criterion. The results are later extended to a more general social welfare function where the degree of aversion to inequality can take values in a broader range. The utilitarian criterion is as usual a sum of the individuals' utility functions weighted by their share in the population:

$$\begin{aligned}
W(c_l^T, c_l, c_h, \tilde{\sigma}, \tilde{\delta}) \equiv & \\
& \gamma \left\{ u(c_l^T)G(\tilde{\sigma}) - \int_0^{\tilde{\sigma}} \sigma g(\sigma) d\sigma + u(c_l)(1 - G(\tilde{\sigma})) \right\} \\
& + (1 - \gamma) \left\{ u(c_h)F(\tilde{\delta}) - \int_0^{\tilde{\delta}} \delta f(\delta) d\delta + u(c_l)(1 - F(\tilde{\delta})) \right\} \quad (5)
\end{aligned}$$

The maxi-min (or ‘‘Rawlsian’’⁴) criterion will place all the weight on the most needy (in terms of utility). In our model, by definition of $\tilde{\sigma}$, those who claim a targeted transfer are at least as well-off as those who do not. The latter benefit from a level of utility $u(c_l)$ which is equal to the level that high-ability workers get if they opt for an unskilled job. So the appropriate criterion which should be maximized under the maxi-min (or ‘‘Rawlsian’’) objective is:

$$u(c_l) \quad (6)$$

The utilitarian and maxi-min criteria are standard in the literature. However, some authors have criticized this approach when individuals differ in several dimensions. Appendix 7.1 deals with an alternative welfare criterion for which no compensation is given for individual characteristics falling within the responsibility of individuals. In this context, Appendix 7.1 considers criterion (5) where σ and δ are ignored.

3 Normative analysis: Analytical results

This section focuses on first-order necessary conditions for an optimum under asymmetric information. As will become clear, these are not in general sufficient conditions for a global maximum. This section is organized as follows. First, using the utilitarian criterion (5), we explain why the case with and without tagging can fulfill the first-order conditions with a utilitarian criterion. Second, analytical properties are derived with the maxi-min objective (6).

⁴Following Okun's (1975) interpretation of Rawls, when all the weight is placed on the most needy, the optimal tax literature uses the term ‘‘Rawlsian objective’’. Obviously, this interpretation does not adequately reflect the richness of the Theory of Justice of Rawls (1971) but it is common practice in economics.

3.1 The utilitarian criterion

The ability characteristics (j), the disutility parameter in skilled job (δ) and the stigma parameter (σ) cannot be observed by the tax authority. The latter perfectly observes reported incomes. When a low-ability worker applies for the targeted transfer, his ability level is observed by social workers. The threshold values $\tilde{\delta}$ and $\tilde{\sigma}$ result from decisions taken by the individuals conditional on the tax and transfer systems (see Equations (3) and (4)). The government chooses the tax-transfer schedule that maximizes the utilitarian social welfare function (5) subject to Equations (3) and (4) and its budget constraint:

$$\phi(c_l^T, c_l, c_h, \tilde{\sigma}, \tilde{\delta}) \equiv \pi_l^T(w_l - c_l^T) + \pi_l(w_l - c_l) + \pi_h(w_h - c_h) = 0 \quad (7)$$

where $\pi_l^T = \gamma G(\tilde{\sigma})$, i.e. the share of the population which is low-able and targeted, $\pi_l = \gamma(1 - G(\tilde{\sigma})) + (1 - \gamma)(1 - F(\tilde{\delta}))$, i.e. the share of the population (high-ability persons as well as low-ability ones) which is occupied in unskilled jobs and $\pi_h = (1 - \gamma)F(\tilde{\delta})$, i.e. the share of the population which is high-able and in a skilled job. Without tagging ($\pi_l^T = 0$), only two consumption levels can be optimally chosen. It is expected that high-ability workers pay taxes ($w_h - c_h > 0$) and low-ability workers receive transfers ($w_l - c_l < 0$). It is also expected that tagged individuals will receive a higher transfer than in the absence of tagging. Whether all low-ability individuals receive a transfer or only the tagged ones is an open question.

Let $\lambda_1, \lambda_2, \lambda_3$ be the multipliers associated respectively with the budget constraint (7), Equations (3) and (4). The Lagrangian expression is:

$$\begin{aligned} \mathcal{L}(c_l^T, c_l, c_h, \tilde{\sigma}, \tilde{\delta}, \lambda_1, \lambda_2, \lambda_3) \equiv & \\ & \gamma \left\{ u(c_l^T)G(\tilde{\sigma}) - \int_0^{\tilde{\sigma}} \sigma g(\sigma) d\sigma + u(c_l)(1 - G(\tilde{\sigma})) \right\} \\ & + (1 - \gamma) \left\{ u(c_h)F(\tilde{\delta}) - \int_0^{\tilde{\delta}} \delta f(\delta) d\delta + u(c_l)(1 - F(\tilde{\delta})) \right\} \\ & + \lambda_1 [\pi_l^T(w_l - c_l^T) + \pi_l(w_l - c_l) + \pi_h(w_h - c_h)] \\ & + \lambda_2 [u(c_h) - \tilde{\delta} - u(c_l)] + \lambda_3 [u(c_l^T) - \tilde{\sigma} - u(c_l)] \end{aligned} \quad (8)$$

From the assumptions on the utility function, the optimal c_l^T, c_l and c_h are necessarily positive. If, at the optimum, $\tilde{\sigma} = 0$, by Equation (4), $c_l^T = c_l$. Therefore, no low-ability person will choose to claim the targeted transfer.⁵ In that sense, tagging is not optimal. In contrast, if at the optimum $\tilde{\sigma} > 0$, by Equation (4), $c_l^T > c_l$, i.e. tagging is optimal. Constraints (3) and (4) can be rewritten respectively as $\phi_1(c_l, c_h, \tilde{\delta}) = 0$ and $\phi_2(c_l^T, c_l, \tilde{\sigma}) = 0$. It can be checked that ϕ_1 and ϕ_2 are quasiconcave. The necessary but not the sufficient

⁵Those characterized by $\sigma = 0$ are actually indifferent between claiming the targeted transfer or not.

conditions for ϕ to be quasiconcave are fulfilled. Trivially the objective $W(c_l^T, c_l, c_h, \tilde{\sigma}, \tilde{\delta})$ is in general not quasiconcave. So, a vector $(c_l^T, c_l, c_h, \tilde{\sigma}, \tilde{\delta}, \lambda_1, \lambda_2, \lambda_3)$ satisfying the following first-order conditions is not necessarily an optimum. These conditions are nevertheless instructive. The first-order conditions can be written as⁶:

$$(\pi_l^T + \lambda_3)u'(c_l^T) = \lambda_1\pi_l^T \quad (9)$$

$$(\pi_l - \lambda_2 - \lambda_3)u'(c_l) = \lambda_1\pi_l \quad (10)$$

$$(\pi_h + \lambda_2)u'(c_h) = \lambda_1\pi_h \quad (11)$$

$$\tilde{\sigma} \frac{\partial \mathcal{L}}{\partial \tilde{\sigma}} = 0 \text{ and } \frac{\partial \mathcal{L}}{\partial \tilde{\sigma}} \leq 0 \quad (12)$$

$$\text{with } \frac{\partial \mathcal{L}}{\partial \tilde{\sigma}} = \lambda_1\gamma g(\tilde{\sigma})[c_l - c_l^T] - \lambda_3 \quad (13)$$

$$\tilde{\delta} \frac{\partial \mathcal{L}}{\partial \tilde{\delta}} = 0 \text{ and } \frac{\partial \mathcal{L}}{\partial \tilde{\delta}} \leq 0 \quad (14)$$

$$\text{with } \frac{\partial \mathcal{L}}{\partial \tilde{\delta}} = \lambda_1(1 - \gamma)f(\tilde{\delta})[w_h - c_h - (w_l - c_l)] - \lambda_2 \quad (15)$$

and Equations (3), (4) and (7).

Proposition 1 *Whatever the value of $\tilde{\sigma}$, the inverse of the marginal cost of public funds is equal to the marginal cost of increasing by a unit the utility of each individual in each group weighted by the share in the population (Diamond and Sheshinski, 1995):*

$$\frac{1}{\lambda_1} = \frac{\pi_l^T}{u'(c_l^T)} + \frac{\pi_l}{u'(c_l)} + \frac{\pi_h}{u'(c_h)} \quad (16)$$

Proof. The proof is straightforward by adding Equations (9), (10) and (11). ■

Having scrutinized the first-order conditions, it turns out that both the case with tagging ($\tilde{\sigma} > 0$) and the one without tagging ($\tilde{\sigma} = 0$) can verify these conditions. The rest of this section is divided into three parts. First, we explain intuitively why $\tilde{\sigma} = 0$ can be optimal, when this property should not hold and why $\tilde{\delta}$ has to be positive. Then, theoretical properties are derived when tagging is optimal and finally when it is not.

3.1.1 Basic mechanisms affecting $\tilde{\sigma}$ and $\tilde{\delta}$

Let us first analytically show that $\tilde{\sigma} = 0$ satisfies the first-order condition $\frac{\partial \mathcal{L}}{\partial \tilde{\sigma}} = 0$. From Equations (9) and (13), we can write:

$$\frac{\partial \mathcal{L}}{\partial \tilde{\sigma}} = \frac{1}{u'(c_l^T)}\pi_l^T[u'(c_l^T) - \lambda_1] + \lambda_1\gamma g(\tilde{\sigma})[c_l - c_l^T] \quad (17)$$

⁶The constraint qualifications have been checked.

From (4), taking c_l constant, a marginal increase in $\tilde{\sigma}$ requires an increase in c_l^T such that $\frac{dc_l^T}{d\tilde{\sigma}} = \frac{1}{u'(c_l^T)}$. The social marginal value of giving dc_l^T to all welfare claimants is $\pi_l^T u'(c_l^T)$ while, $\pi_l^T \lambda_1$ is the social marginal cost of the corresponding increase in public expenditures. Finally, $\lambda_1 \gamma g(\tilde{\sigma})[c_l - c_l^T]$ is the net cost due to marginal individuals shifting from c_l to the higher c_l^T . If $\tilde{\sigma} = 0$, the fraction of the tagged population, π_l^T , is zero. Moreover $c_l = c_l^T$. Hence, from (17), $\frac{\partial \mathcal{L}}{\partial \tilde{\sigma}} = 0$. The intuition behind this property is the following. By (4), the marginal individuals who enter the tagged population are always indifferent because their utility levels are the same whether they are tagged or not. Therefore, there is no direct impact on the utilitarian criterion W . Their earnings, w_l , and hence their contribution to aggregate output are also the same. Moreover, in $\tilde{\sigma} = 0$, $c_l^T = c_l$. A marginal increase in $\tilde{\sigma}$ has therefore no impact on net resources.⁷

Under which circumstances, is it plausible that a positive value of $\tilde{\sigma}$ also solves the first-order conditions? It will turn out that the shape of the distributions of σ and δ is critical. From the discussion above, a necessary condition is that $\frac{\partial \mathcal{L}}{\partial \tilde{\sigma}} > 0$ for some positive values of $\tilde{\sigma}$. From (17), $\tilde{\sigma}$ should increase if:

$$\frac{dc_l^T}{d\tilde{\sigma}} [u'(c_l^T) - \lambda_1] > \frac{g(\tilde{\sigma})}{G(\tilde{\sigma})} [c_l^T - c_l] \lambda_1 \quad (18)$$

As $\tilde{\sigma}$ starts increasing from zero, c_l^T also grows (otherwise there would be less redistribution from the high-ability to the low-ability people). It will be shown that $\lambda_3 < 0$ when $\tilde{\sigma} > 0$. Hence from (9), $u'(c_l^T) > \lambda_1$. The left-hand side of (18) is then positive. As $\tilde{\sigma}$ starts increasing from zero, $c_l^T - c_l$ increases too and it becomes more difficult to fulfill (18). So, a decline of $\frac{g(\tilde{\sigma})}{G(\tilde{\sigma})}$ as $\tilde{\sigma}$ rises helps to fulfill (18). A distribution for which $\frac{g(\sigma)}{G(\sigma)}$ is decreasing is said to have decreasing monotone reversed hazard (or failure) rate. Equivalently, the log of the cumulative distribution function, G , has to be concave. This condition is satisfied by most of the usual distributions.⁸

The distribution of δ also affects the chances of finding a positive value of $\tilde{\sigma}$ that solves the first-order conditions. From Equations (11) and (15), we can rewrite:

$$\frac{\partial \mathcal{L}}{\partial \tilde{\delta}} = \frac{(1 - \gamma)F(\tilde{\delta})}{u'(c_h)} [u'(c_h) - \lambda_1] - \lambda_1 (1 - \gamma) f(\tilde{\delta}) [c_h - w_h - (c_l - w_l)] \quad (19)$$

If $\tilde{\delta} = 0$, $F(\tilde{\delta}) = 0$ and $c_h = c_l$ therefore $\frac{\partial \mathcal{L}}{\partial \tilde{\delta}} \Big|_{\tilde{\delta}=0} = -\lambda_1 (1 - \gamma) f(0) (w_l - w_h) > 0$. So, $\tilde{\delta}$ has

⁷It should be mentioned that $\tilde{\sigma} = 0$ can also solve the first-order conditions when $w_l > 0$ and the tagged population does not work.

⁸A sufficient condition for G to be log-concave is that the density is log-concave. Families of distributions that always have log-concave density functions include the uniform, the normal, the logistic, the extreme-value, the chi-square, the chi, the exponential and the Laplace distributions. However some families of distributions do not have log-concave density for all their parameter values. Still they may have a log-concave distribution function. This is the case with the Gamma, the power function and the Weibull distributions for instances. See Bagnoli and Bergström (1989).

to be positive. As long as $\tilde{\delta}$ should increase, one has

$$\frac{dc_h}{d\tilde{\delta}} [u'(c_h) - \lambda_1] > \lambda_1 \frac{f(\tilde{\delta})}{F(\tilde{\delta})} [c_h - w_h - (c_l - w_l)] \quad (20)$$

The structure of this expression and the one of (18) are similar. If many high-ability workers are characterized by low values of δ , $\tilde{\delta}$ will be such that a large proportion of high-ability individuals will work in skilled jobs. This reduces the marginal cost of public funds, λ_1 , and from (18), increases the probability that a positive value of $\tilde{\sigma}$ is optimal.

Section 5 will be devoted to numerical simulations which will provide additional insights on situations where tagging is or not optimal. Meanwhile, the analytical properties derived in each environment will be presented.

3.1.2 Analytical properties under tagging ($\tilde{\sigma} > 0$)

The relative value of c_l^T and c_h is a major issue here. The next lemma will be useful later on.

Lemma 1 *If $\tilde{\sigma} > 0$: $\frac{1}{u'(c_l)} \leq \frac{\theta}{u'(c_l^T)} \leq \frac{1}{u'(c_h)}$ with strict inequalities if $\tilde{\delta} > 0$, where $\theta = \frac{\pi_l^T}{\pi_l^T + \lambda_3} \left[\frac{1 - \pi_l^T - \lambda_3}{1 - \pi_l^T} \right] \leq 1$.*

Proof. Equation (9) can be rewritten as:

$$\frac{1}{\lambda_1} = \frac{\pi_l^T}{\pi_l^T + \lambda_3} \frac{1}{u'(c_l^T)} \quad (21)$$

Putting this in (16) and dividing by $\pi_l + \pi_h \equiv 1 - \pi_l^T$ yields:

$$\frac{\theta}{u'(c_l^T)} = \frac{x}{u'(c_l)} + \frac{1-x}{u'(c_h)} \quad (22)$$

where $x = \frac{\pi_l}{\pi_l + \pi_h}$ is the probability of occupying an unskilled job conditional on being untagged ($0 < x < 1$) and $\theta = \frac{\pi_l^T}{\pi_l^T + \lambda_3} \left[\frac{1 - \pi_l^T - \lambda_3}{1 - \pi_l^T} \right]$. Hence, θ is nonnegative. According to the sign of λ_3 , θ can be higher or lower than 1. Since $\tilde{\delta} \geq 0$, Equation (3) ensures that $u(c_h) \geq u(c_l)$. Therefore, $\frac{\theta}{u'(c_l^T)}$ lies (strictly) between $\frac{1}{u'(c_l)}$ and $\frac{1}{u'(c_h)}$ (if $\tilde{\delta} > 0$). ■

Proposition 2 *If $\tilde{\sigma} > 0$, then consumption levels are ordered in the following way:*

$$c_h > c_l^T > c_l \quad (23)$$

Proof. First, $\tilde{\sigma} > 0$ and Equation (4) insures that $c_l^T > c_l$. Hence, $u'(c_l^T) < u'(c_l)$. Second, from (16), $\lambda_1 > 0$. So, since $g(\tilde{\sigma}) > 0$, the first-order conditions (12)-(13) imply that $\lambda_3 < 0$. Therefore, $\theta > 1$. This and Lemma 1 yield that $u'(c_h) < u'(c_l^T)$. Hence, putting these two results together, one has : $u'(c_h) < u'(c_l^T) < u'(c_l)$. This is equivalent to (23). ■

Proposition 2 is in line with Parsons (1996).

Proposition 3 *If $\tilde{\sigma} > 0$, the tax levied on those occupied in skilled jobs is strictly higher than the tax paid by workers occupied in unskilled jobs: $w_h - c_h > w_l - c_l$.*

Proof. Section 3.1.1 has shown that $\tilde{\delta} > 0$. Then, Equations (14) and (15) imply that the sign of λ_2 is the one of $w_h - w_l - (c_h - c_l)$. The latter is apparently ambiguous because $w_h - w_l$ and $c_h - c_l$ are both positive (see Proposition 2). By contradiction, it can however be shown that λ_2 is positive. Let us assume that $w_h - c_h \leq w_l - c_l$. Then $\lambda_2 \leq 0$. Recalling that $\lambda_3 < 0$, $u'(c_l) < \lambda_1$ is then a consequence of (11). However, from (9), $u'(c_l^T) > \lambda_1$. Combining these two results leads to $u'(c_l^T) > u'(c_l)$ or $c_l^T < c_l$, which is in contradiction with Proposition 2. Therefore, $\lambda_2 > 0$, which implies that $w_h - c_h > w_l - c_l$. ■

Proposition 4 *If $\tilde{\sigma} > 0$, it cannot be ruled out that untagged individuals with low ability pay taxes.*

Proof. From the budget constraint (7), it can be shown that:

$$w_l - c_l = \pi_l^T(c_l^T - c_l) - \pi_h[w_h - c_h - (w_l - c_l)] \quad (24)$$

From Propositions 2 and 3, both $c_l^T - c_l$ and $w_h - c_h - (w_l - c_l)$ are positive. Hence, the sign of $w_l - c_l$ is ambiguous. In other words, the gross income of untagged low-ability and high-ability individuals can be increased (in case of a transfer: $w_l - c_l < 0$) or decreased (in case of a tax: $w_l - c_l > 0$) by the optimal tax-transfer system. ■

3.1.3 Analytical properties when tagging does not prevail ($\tilde{\sigma} = 0$)

Some quite intuitive propositions can be shown when $\tilde{\sigma} = 0$.

Proposition 5 *If $\tilde{\sigma} = 0$, then consumption levels are ordered in the following way:*

$$c_h > c_l = c_l^T \quad (25)$$

Proof. This property follows immediately from $\tilde{\sigma} = 0$, $\tilde{\delta} > 0$ (see Section 3.1.1) and Equation (4). ■

Proposition 6 *If $\tilde{\sigma} = 0$, $w_h - c_h > w_l - c_l$, with $w_h - c_h > 0$ and $w_l - c_l < 0$.*

Proof. The proof of this proposition follows the same lines as the proof of Proposition 3 above. From Proposition 3, $\tilde{\delta} > 0$ (see Section 3.1.1) and the result $\lambda_1 > 0$, since $\lambda_2 > 0$ and $\lambda_2 = \lambda_1(1 - \gamma)f(\tilde{\delta})[(w_h - c_h) - (w_l - c_l)]$, $w_h - c_h > w_l - c_l$. Moreover, from the budget constraint, it is immediately seen that $w_h - c_h$ and $w_l - c_l$ have opposite signs. ■

Figure 2 summarizes our previous results. As illustrated by the upper panel of Figure 2, when tagging is suboptimal, all low-ability workers (that is a proportion γ of the population) and a proportion $1 - F(\tilde{\delta})$ of the high-ability people are in unskilled jobs and have $u(c_l)$ as utility level. Their consumption level (i.e. their gross labor income increased by a transfer) is (strictly) lower than the one of the high-ability workers in skilled job. The latter (whose proportion is represented by the grey area) have their utility levels included between $u(c_h)$ and $u(c_l)$. The lower panel of Figure 2 represents an economy where tagging is optimal: A proportion $G(\tilde{\sigma})$ of low-ability workers (the dark area) have a consumption level c_l^T (strictly) higher than c_l (which, in this case, can be lower than gross income) and (strictly) lower than c_h .

FIGURE 2 ABOUT HERE

3.2 The maxi-min welfare criterion

If the utilitarian objective is replaced by the ‘‘Rawlsian’’ one, the consumptions levels should maximize (6) subject to the budget constraint and the two conditions (3) and (4) defining the cut-off levels. Intuitively, introducing tagging enables the government to increase the transfer given to the targeted individuals. Yet, to finance this, more high-ability workers have to opt for a skilled job. This requires a decline of the consumption level of the low-ability workers who do not claim the targeted transfer. Since these people are the least well-off, such a change cannot be advocated in a maxi-min perspective. Put another way:

Proposition 7 *The maxi-min optimum requires $\tilde{\sigma} = 0$.*

Proof. Totally differentiating the budget constraint and the two conditions defining the cut-off levels, (3) and (4), and rearranging yields:

$$\begin{aligned} u'(c_l^T) \left[\frac{\pi_l^T}{u'(c_l^T)} + \frac{\pi_l}{u'(c_l)} + \frac{\pi_h}{u'(c_h)} \right] dc_l^T = \\ \left[\frac{\pi_l}{u'(c_l)} + \frac{\pi_h}{u'(c_h)} - \gamma g(\tilde{\sigma}) (c_l^T - c_l) \right] d\tilde{\sigma} \\ + \left[(1 - \gamma) f(\tilde{\delta}) (w_h - c_h - (w_l - c_l)) - \frac{\pi_h}{u'(c_h)} \right] d\tilde{\delta} \end{aligned} \quad (26)$$

and

$$dc_l = \frac{u'(c_l^T) dc_l^T - d\tilde{\sigma}}{u'(c_l)} \quad (27)$$

Substituting (26) into (27), it is easily seen that

$$\frac{\partial c_l}{\partial \tilde{\sigma}} = \frac{1}{u'(c_l)} \left[\frac{\frac{\pi_l}{u'(c_l)} + \frac{\pi_h}{u'(c_h)} - \gamma g(\tilde{\sigma}) (c_l^T - c_l)}{\frac{\pi_l^T}{u'(c_l^T)} + \frac{\pi_l}{u'(c_l)} + \frac{\pi_h}{u'(c_h)}} - 1 \right] \quad (28)$$

This partial derivative is zero at $\tilde{\sigma} = 0$ and is negative for any $\tilde{\sigma} > 0$. Put differently, extending the size of the tagged population from $\tilde{\sigma} > 0$ to $\tilde{\sigma} + d\tilde{\sigma}$ implies that the untagged low-ability group will be worse off. Since the utility which is maximized under the maxi-min criterion (6) is the one of this group, a ‘‘Rawlsian’’ cannot accept such an extension. Conversely, for any $\tilde{\sigma} > 0$, a marginal decrease in $\tilde{\sigma}$ has a positive effect on c_l . This holds true until $\tilde{\sigma} = 0$. The maxi-min optimum definitely requires $\tilde{\sigma} = 0$. ■

The same conclusion obviously also holds if the lowest consumption level, namely c_l , is maximized.

4 Two extensions

First, we consider a more general social welfare function and show how the aversion to inequality affects the decision to use tagging. Second, we show that our results are robust to the introduction of a ‘‘social recognition’’ of the low-ability individuals who fend for themselves. The latter can be seen as the counter-part of stigmatization.

4.1 A general social welfare function

Consider the following standard general social welfare function

$$\gamma \left\{ \int_0^{\tilde{\sigma}} \frac{[u(c_l^T) - \sigma]^{1-v}}{1-v} g(\sigma) d\sigma + \int_{\tilde{\sigma}}^{\infty} \frac{u(c_l)^{1-v}}{1-v} g(\sigma) d\sigma \right\} \\ + (1-\gamma) \left\{ \int_0^{\tilde{\delta}} \frac{[u(c_h) - \delta]^{1-v}}{1-v} f(\delta) d\delta + \int_{\tilde{\delta}}^{\infty} \frac{u(c_l)^{1-v}}{1-v} f(\delta) d\delta \right\} \quad \text{for } v \neq 1$$

where v is aversion to inequality. We have already studied the case in which $v = 0$ (utilitarian) and that in which $v \rightarrow \infty$ (maxi-min). Consider now an intermediate value of v and assume that the conditions under which tagging is optimal are satisfied. The first-order condition with respect to c_l^T and (13) now lead to:

$$\frac{\partial \mathcal{L}^v}{\partial \tilde{\sigma}} = \frac{1}{u'(c_l^T)} \left[\gamma \int_0^{\tilde{\sigma}} \frac{u'(c_l^T)}{[u(c_l^T) - \sigma]^v} g(\sigma) d\sigma - \lambda_1 \pi_l^T \right] + \lambda_1 \gamma g(\tilde{\sigma})(c_l - c_l^T)$$

where \mathcal{L}^v denotes the Lagrangian. As previously, from (4), taking c_l as constant, a marginal increase in $\tilde{\sigma}$ requires an increase in c_l^T such that $\frac{dc_l^T}{d\tilde{\sigma}} = \frac{1}{u'(c_l^T)}$. The social marginal value of giving dc_l^T to all welfare claimants is $\gamma \int_0^{\tilde{\sigma}} \frac{u'(c_l^T)}{[u(c_l^T) - \sigma]^v} g(\sigma) d\sigma$ while, $\pi_l^T \lambda_1$ is the social marginal cost of the corresponding increase in public expenditures. Finally, the last term is the same as in (17). If $\tilde{\sigma} > 0$, the first-order condition is violated if

$$\frac{1}{u'(c_l^T)} \left[\gamma \int_0^{\tilde{\sigma}} \frac{u'(c_l^T)}{[u(c_l^T) - \sigma]^v} g(\sigma) d\sigma - \lambda_1 \pi_l^T \right] < -\lambda_1 \gamma g(\tilde{\sigma})(c_l - c_l^T) \quad (29)$$

From (4), the right-hand side of this last expression is strictly positive when $\tilde{\sigma} > 0$. As the aversion to inequality v rises, less and less weight is given to the marginal utility of the tagged population (through the decreasing multiplier $1/[u(c_l^T) - \sigma]^v$). The following property is therefore to be expected. Starting from a case where tagging is optimal under a utilitarian criterion $v = 0$, there exists a higher level of aversion to inequality v under which tagging is optimal and above which it is not.

4.2 Introducing “social approval”

Imagine people who are not tagged benefit from social approbation. This counter-part of stigmatization increases the utility of people who fend for themselves (i.e. everyone in the economy except tagged people). Let us denote social approbation by a scalar $a > 0$. The threshold level for low-ability workers satisfies now the following equality:

$$u(c_l^T) - \tilde{\sigma} = u(c_l) + a \quad (30)$$

The threshold level $\tilde{\delta}$ is still defined by (3).

Let us first examine if the suboptimality of tagging which prevails with a maxi-min criterion is affected when considering social approval. The appropriate maxi-min criterion becomes:

$$u(c_l) + a$$

It can easily be shown that (28) is still valid in this case. Therefore, the suboptimality of tagging with a maxi-min criterion is maintained when social approbation is considered.

Let us now turn to a utilitarian criterion. The utilitarian criterion without social approbation (5) is slightly modified by considering social approbation: a new term $[1 - \gamma G(\tilde{\sigma})]a$ has to be added. It can easily be checked that the first-order conditions with respect to c_l^T and c_l remain unchanged. Moreover the first-order condition with respect to $\tilde{\sigma}$ is exactly the same because the effect of a marginal change in $\tilde{\sigma}$ on the utilitarian criterion, namely:

$$\gamma[u(c_l^T) - \tilde{\sigma} - u(c_l) - a]g(\tilde{\sigma})$$

is still zero by (30). In sum, the introduction of social approval does not change the conclusions of the theoretical analysis.

5 A numerical analysis under utilitarianism

This section focuses on the utilitarian criterion (5). With a log-concave distribution for δ , there always exists a solution ($\tilde{\sigma} = 0, \tilde{\delta} > 0$) to the first-order conditions (see Appendix 7.2). In addition, there can exist (at least) one other solution such that $\tilde{\sigma} > 0, \tilde{\delta} > 0$.

Then, several local optima are observed. Numerical methods are therefore needed to see when tagging is optimal and when it is not. The following numerical exercise emphasizes the crucial role of the densities of σ and δ . The numerical analysis shows that tagging is only optimal when the distribution of σ is characterized by a density highly concentrated on low values of σ (i.e., simultaneously, small mean and variance), where ‘low’ should be understood relative to $u(w_l)$. However, a distribution sufficiently concentrated on relatively low values of σ is only a necessary condition because the distribution of δ also plays a crucial role. In some cases, where σ is highly concentrated on low values, tagging is not optimal because the distribution of the ability parameters, δ , has a very low or a very large variance (and mean).

This section illustrates and explains these results. The presentation is organized as follows. First, we rewrite and calibrate the model. Second, we illustrate how the parameters’ space can be divided into an area where tagging is optimal and another where a utilitarian would avoid it. It is shown that very restrictive assumptions on the distributions of σ and δ are needed in order to conclude that tagging is optimal. Third, we conduct a sensitivity analysis.

5.1 Reformulation of the model and calibration

Combining constraints (3), (4), (7) and the utilitarian criterion (5), it is convenient to rewrite the problem as:

$$\begin{aligned} W(c_l, \tilde{\sigma}, \tilde{\delta}) \equiv & \gamma \left[\tilde{\sigma} G(\tilde{\sigma}) - \int_0^{\tilde{\sigma}} \sigma g(\sigma) d\sigma \right] + \\ & + (1 - \gamma) \left[\tilde{\delta} F(\tilde{\delta}) - \int_0^{\tilde{\delta}} \delta f(\delta) d\delta \right] + u(c_l) \end{aligned} \quad (31)$$

with

$$c_l = \frac{w_l + (1 - \gamma) F(\tilde{\delta})(w_h - w_l)}{1 + \gamma G(\tilde{\sigma}) \left[\frac{c_l^T}{c_l} - 1 \right] + (1 - \gamma) F(\tilde{\delta}) \left[\frac{c_h}{c_l} - 1 \right]} \quad (32)$$

Equation (32) can be rewritten as $\phi_3(c_l, c_l^T, c_h, \tilde{\sigma}, \tilde{\delta}) = 0$. Using a logarithmic utility function, $u(\cdot) \equiv \log_k(\cdot)$ with $k > 1$ and again Equations (3), (4), it is convenient to rewrite ϕ_3 as:

$$c_l(\tilde{\sigma}, \tilde{\delta}) = \frac{w_l + (1 - \gamma) F(\tilde{\delta})(w_h - w_l)}{1 + \gamma G(\tilde{\sigma})(k^{\tilde{\sigma}} - 1) + (1 - \gamma) F(\tilde{\delta})(k^{\tilde{\delta}} - 1)} \quad (33)$$

Substituting (33) into the objective function (31), the problem becomes a maximization with respect to $(\tilde{\sigma}, \tilde{\delta})$.

FIGURE 3 ABOUT HERE

Let δ and σ be distributed according to Gamma distributions.⁹ Figure 3 represents Gamma densities with parameter r respectively equal to 0.2, 0.42, 1, 1.5 and 3. These values illustrate the variety of shapes of the Gamma density. Moreover, $r = 0.42$ will be a critical value for our results. As r decreases below 1, the density, $f(x)$, becomes quickly negligible when x increases. This phenomenon captures what is here meant by “concentration”. This notion is a relative one. It depends on the values of w_h and w_l and on the shape of $u(\cdot)$.

Let r_δ, r_σ be the parameters characterizing Gamma distributions respectively for δ and σ . To emphasize the role of these parameters, the density and distribution functions can be rewritten as $F(\tilde{\delta} | r_\delta), f(\tilde{\delta} | r_\delta), G(\tilde{\sigma} | r_\sigma)$ and $g(\tilde{\sigma} | r_\sigma)$ in (31) and (33).

With two levels of skills, one can hardly base our assumption about w_h and w_l on actual wage distributions. Hence, we take $w_h = 100, w_l = 20$ and develop, in Subsection 5.3, a sensitivity analysis. We assume a logarithmic utility function with a basis k equal to 2 ($u(w_h) = \log_2(w_h) = 6.64$ and $u(w_l) = \log_2(w_l) = 4.32$) and also develop a sensitivity analysis in Subsection 5.3. The proportion of low-ability individuals, γ , is assumed to be equal to 0.5. The role of γ is also discussed in Subsection 5.3.

5.2 The effect of r_σ and r_δ

Methodology

Let $W(\tilde{\sigma}, \tilde{\delta}, r_\sigma, r_\delta)$ denote (31) after substitution of (33). The optimum $(\tilde{\sigma}, \tilde{\delta})$ verifies

$$\frac{\partial W}{\partial \tilde{\sigma}}(\tilde{\sigma}, \tilde{\delta}, r_\sigma, r_\delta) = 0 \quad (34)$$

and

$$\frac{\partial W}{\partial \tilde{\delta}}(\tilde{\sigma}, \tilde{\delta}, r_\sigma, r_\delta) = 0 \quad (35)$$

This system is highly nonlinear. Therefore, according to the chosen initial values and the numerical method used, a solution to (34)-(35) need not be a global optimum. So, for each pair (r_σ, r_δ) , the objective function (31) (with c_l defined by (33)) is evaluated for a wide range of values of the endogenous variables $(\tilde{\sigma}, \tilde{\delta})$. Through this, we check whether the solution found to (34)-(35) is the global optimum.

The above system defines an implicit relationship between the optimal values of $\tilde{\sigma}, \tilde{\delta}$ and r_σ and r_δ . This system is however too complex to be studied analytically.

⁹A positive random variable follows a Gamma law of parameter r if its density is given by:

$$f(x) = \frac{1}{\Gamma(r)} \exp(-x)x^{r-1}$$

The parameter r of a Gamma distribution is equal to the mean and the variance of the distribution. We have checked that our conclusions are maintained with other continuous distributions defined on the infinite support $[0, +\infty)$.

Since (34) and (35) characterize an optimum, we can totally differentiate them with respect to $\tilde{\sigma}, \tilde{\delta}, r_\sigma$ and r_δ as

$$\frac{\partial^2 W}{\partial \tilde{\sigma}^2} d\tilde{\sigma} + \frac{\partial^2 W}{\partial \tilde{\delta} \partial \tilde{\sigma}} d\tilde{\delta} = -\frac{\partial^2 W}{\partial r_\sigma \partial \tilde{\sigma}} dr_\sigma - \frac{\partial^2 W}{\partial r_\delta \partial \tilde{\sigma}} dr_\delta \quad (36)$$

and

$$\frac{\partial^2 W}{\partial \tilde{\sigma} \partial \tilde{\delta}} d\tilde{\sigma} + \frac{\partial^2 W}{\partial \tilde{\delta}^2} d\tilde{\delta} = -\frac{\partial^2 W}{\partial r_\sigma \partial \tilde{\delta}} dr_\sigma - \frac{\partial^2 W}{\partial r_\delta \partial \tilde{\delta}} dr_\delta \quad (37)$$

This formulation will be helpful in the numerical analysis to study the effect(s) of small changes in r_σ and r_δ . The effects of dr_σ and dr_δ on $d\tilde{\sigma}$ are given by:

$$\frac{d\tilde{\sigma}}{dr_\sigma} = \frac{\frac{\partial^2 W}{\partial \tilde{\delta} \partial \tilde{\sigma}} \frac{\partial^2 W}{\partial r_\sigma \partial \tilde{\delta}} - \frac{\partial^2 W}{\partial r_\sigma \partial \tilde{\sigma}} \frac{\partial^2 W}{\partial \tilde{\delta}^2}}{\frac{\partial^2 W}{\partial \tilde{\sigma}^2} \frac{\partial^2 W}{\partial \tilde{\delta}^2} - \left(\frac{\partial^2 W}{\partial \tilde{\sigma} \partial \tilde{\delta}}\right)^2} \quad (38)$$

$$\frac{d\tilde{\sigma}}{dr_\delta} = \frac{\frac{\partial^2 W}{\partial \tilde{\delta} \partial \tilde{\sigma}} \frac{\partial^2 W}{\partial r_\delta \partial \tilde{\delta}} - \frac{\partial^2 W}{\partial r_\delta \partial \tilde{\sigma}} \frac{\partial^2 W}{\partial \tilde{\delta}^2}}{\frac{\partial^2 W}{\partial \tilde{\sigma}^2} \frac{\partial^2 W}{\partial \tilde{\delta}^2} - \left(\frac{\partial^2 W}{\partial \tilde{\sigma} \partial \tilde{\delta}}\right)^2} \quad (39)$$

The effects of dr_σ and dr_δ on $d\tilde{\delta}$ are given by:

$$\frac{d\tilde{\delta}}{dr_\sigma} = -\frac{\frac{\partial^2 W}{\partial \tilde{\sigma} \partial \tilde{\delta}}}{\frac{\partial^2 W}{\partial \tilde{\delta}^2}} \frac{d\tilde{\sigma}}{dr_\sigma} - \frac{\frac{\partial^2 W}{\partial r_\sigma \partial \tilde{\delta}}}{\frac{\partial^2 W}{\partial \tilde{\delta}^2}} \quad (40)$$

$$\frac{d\tilde{\delta}}{dr_\delta} = -\frac{\frac{\partial^2 W}{\partial \tilde{\sigma} \partial \tilde{\delta}}}{\frac{\partial^2 W}{\partial \tilde{\delta}^2}} \frac{d\tilde{\sigma}}{dr_\delta} - \frac{\frac{\partial^2 W}{\partial r_\delta \partial \tilde{\delta}}}{\frac{\partial^2 W}{\partial \tilde{\delta}^2}} \quad (41)$$

The signs of these expressions are in general ambiguous. In the following numerical exercise, the components of (38)-(41) will be evaluated. This information will be useful to understand how and why $\tilde{\sigma}$ and $\tilde{\delta}$ vary with r_σ and r_δ , respectively.

A necessary condition for the optimality of tagging: a low r_σ

In a (r_δ, r_σ) space, Figure 4 displays the area where tagging is optimal (see the shaded area). It highlights that tagging can only be optimal for low values of r_σ . This result turns out to be true in all reported and unreported simulations. When r_σ decreases, the distribution of σ is more concentrated on low values. In our example, tagging can never be optimal if $r_\sigma > 0.42$.¹⁰ This illustrates the idea of concentration: Very few people have a σ which is nonnegligible compared to $u(w_l)$.

FIGURE 4 ABOUT HERE

¹⁰If $r_\sigma = 0.42$, 45.8% of the low-ability workers have a σ above 0.2 and 9.7% of the low-ability workers have an σ above 1.2. The utility from the gross wage in a unskilled job is $u(w_l) = 4.3$, and therefore, $\sigma = 0.2$ (respectively $\sigma = 1.2$) is only 4.6% (respectively 27.8%) of $u(w_l)$, which is relatively low.

This first result is quite intuitive. If stigmatization is considered as a negligible phenomenon (in the sense that the density of σ is concentrated on very low values of σ relative to $u(w_l)$), then the model confirms the traditional result in the tagging literature (Akerlof, 1978; Parsons, 1996; Salanié, 2002). This literature shows that if a portion of the low-ability people can be costlessly tagged (i.e., without neither administrative costs nor prejudices due to stigma), the total utilitarian welfare is raised by giving an allowance targeted on this sub-population. There is an incentive to do so because the benefits paid to tagged low-ability workers (c_l^T) provide no work disincentive to the high-ability workers so that the former sub-population can be treated more generously. In our model, with a density of σ sufficiently concentrated on very low values of σ relative to $u(w_l)$, the prejudice due to stigma is much lower than the disadvantages linked to a redistributive system purely administered by the tax authority. Then, the traditional result holds: tagging is optimal. However, a distribution of the intensity of stigmatization highly concentrated around relatively low values is only a necessary condition. Simulations results will show that the dispersion of abilities among those who can perform skilled jobs also plays a crucial role.

A look at the area where tagging is optimal

Looking at the area where tagging is optimal, we will describe what happens when r_σ (respectively r_δ) increases. Thanks to the methodology previously defined and the simulations, we will be able to locally study the signs of Expressions (38) to (41). Due to space limitation, we explain general features by considering a few examples.

What happens in the interior of the tagging region when r_σ increases? Table 2 displays the main features of the optima when r_σ increases and r_δ is fixed to 1.3. As the dispersion and the mean of the intensity of stigmatization gets wider, c_l increases (starting from values below w_l), c_l^T somewhat decreases, therefore, π_l^T declines, π_l increases and π_h decreases. In the area where tagging is optimal, $\tilde{\sigma}$ and $\tilde{\delta}$ decrease with r_σ . $\forall r_\sigma \in [0.1; 0.42[$, we compute that $\frac{\partial^2 W}{\partial r_\sigma \partial \tilde{\sigma}} < 0$, $\frac{\partial^2 W}{\partial \tilde{\sigma}^2} < 0$, $\frac{\partial^2 W}{\partial \tilde{\delta} \partial \tilde{\sigma}} > 0$, $\frac{\partial^2 W}{\partial r_\sigma \partial \tilde{\delta}} < 0$ and $\frac{\partial^2 W}{\partial \tilde{\sigma}^2} \frac{\partial^2 W}{\partial \tilde{\delta}^2} > (\frac{\partial^2 W}{\partial \tilde{\sigma} \partial \tilde{\delta}})^2 > 0$. Remembering (38), this explains why $\frac{d\tilde{\sigma}}{dr_\sigma} < 0$. The sign of these effects can then be introduced in (40) to understand why $\frac{d\tilde{\delta}}{dr_\sigma} < 0 \forall r_\sigma \in [0.1; 0.42[$.

TABLE 2 ABOUT HERE

TABLE 3 ABOUT HERE

What happens in the interior of the tagging region when r_δ increases? Let us take the case where $r_\sigma = 0.2$ and consider a few values of r_δ . For sufficiently low values of r_δ (here 0.1), the distribution of δ is highly concentrated around low values. With moderate

differences $c_h - c_l$, nearly all high-ability workers opt for a skilled occupation. Given the available resources, stigmatization can be avoided by giving relatively high levels of transfer to all low-ability workers. For $r_\delta \in [0.2; 0.5]$, Table 3 indicates that c_l^T increases, c_l decreases and π_l decreases. So, as the dispersion and the mean of δ increases, tagging is first used more intensively. From the values taken by the components of (39), $\frac{d\tilde{\sigma}}{dr_\delta}$ is positive but less and less so as r_δ increases in the interval $[0.2; 1.05[$. Around $r_\delta = 1.05$, still exploiting (39), $\frac{d\tilde{\sigma}}{dr_\delta}$ becomes negative. Actually, for high values of r_δ (relative to $u(w_h) - u(w_l)$), the share of high-ability workers with a high δ becomes so large that redistribution has to decrease. The substitution away from skilled jobs accelerates drastically (see the evolution of π_h in Table 3). Looking at the components of $\frac{d\tilde{\delta}}{dr_\delta}$ in (41) clarifies why the optimal value of $\tilde{\delta}$ does not vary monotonically with r_δ . This effect and the change in the distribution function nevertheless lead to a monotonic decline in π_h . The well-being of those earning w_l is therefore more heavily weighted in the welfare function. Therefore, c_l increases and c_l^T decreases. In other words tagging is less and less used ($c_l^T - c_l \rightarrow 0$). This numerical result confirms the intuition provided in Subsection 3.1.1.

To sum up, even if the intensity of stigmatization is highly concentrated around low values in comparison with $u(w_l)$, tagging is not optimal if the distribution of the disutility of skilled jobs is either very concentrated on low values or on the contrary if it has a large variance (and large mean).

5.3 Sensitivity analysis

Subsection 5.2 emphasized the role of the distributions of σ and δ on the optimality of tagging. Let us now look at the effect of the other parameters. We consider in turn the basis of the logarithmic utility function, the wage differential and the proportion of low-ability individuals in the total population.

First, if the basis of the logarithmic utility function ($k > 1$) increases, the area where tagging is optimal in (r_σ, r_δ) space is reduced (see Figure 5). Figure 5 displays how the area where tagging is optimal shrinks when k increases from 2 to 2.6. Intuitively, when tagging is optimal, it is used to reduce the difference in utility levels between $u(c_h)$ and $u(c_l^T)$, that is $\log_k(c_h) - \log_k(c_l^T)$. It can be checked that $\log_k(c_h/c_l^T)$ is decreasing convex with respect to $k > 1$. So, for any given $\frac{c_h}{c_l^T} > 1$, the former difference in utility levels softens by itself when k increases. Tagging is then less needed. So, for any pair (r_σ, r_δ) where tagging is optimal but c_l^T is close to c_l , tagging becomes suboptimal ($c_l^T = c_l$) when k increases.

Second, if the difference between the gross wage rates, $w_h - w_l$, decreases, tagging is again less needed. Actually, as in the case of an increase of k , the inequality in utility

levels is reduced when the difference is reduced. Therefore, the area where tagging is optimal is reduced as illustrated Figure 5, yet with a scope of reduction depending on the parameter's variation.

FIGURE 5 ABOUT HERE

Third, if the proportion γ of low-ability individuals in the population increases, the area where tagging is optimal is also reduced. Let us see why by considering a point in the tagging area. If γ increases, the high-ability population receives a lower weight in the utilitarian criterion. In addition, keeping the allocation of resources unchanged is not feasible because of the growing share of low-ability persons. Simulations show that a utilitarian reacts partly through an increase in taxation, $w_h - c_h$, and partly through a decrease of c_l^T . Moreover, to prevent $\tilde{\delta}$ from decreasing too much, c_l is reduced too but to a lower extent than c_l^T . Therefore, close to the boundary of the tagging area where initially $c_l^T \sim c_l$, an increase in γ leads to $c_l^T - c_l \rightarrow 0$ (and $\tilde{\sigma} \rightarrow 0$).

6 Conclusion

This paper has questioned the optimality of conditioning income transfers on personal characteristics of the potential welfare recipients instead of basing them only on reported incomes. In our analysis, wage formation and the demand for labor are assumed to be exogenous. We have developed a simple framework with two categories of (unobservable) abilities, in which individuals with low abilities decide whether or not to claim the targeted transfer and welfare authorities assess eligibility perfectly and costlessly. This assumption about welfare agencies is deliberately in favor of tagging. Being a welfare recipient is however socially disapproved because a social norm is violated. This still widespread norm states that anyone should fend for oneself and one's family. It has been assumed that the disutility of being stigmatized is exogenously distributed in the population.

The main conclusions are as follows. First, with a utilitarian criterion, tax-transfer systems with and without tagging can solve the first-order optimality conditions. These are not necessarily sufficient however for determining the (global) optimum. Therefore, we have developed a numerical analysis that suggests that tagging can only be recommended if the distribution of the intensity of stigmatization is highly concentrated around low values, where low is a relative notion. The intensity of stigmatization is measured in comparison with the level of earnings in unskilled jobs. However, this is only a necessary condition. The dispersion of the disutility of effort among those who can perform a skilled job ('the high-ability people') also plays a crucial role. Even when the intensity of stigmatization can be considered as a minor factor, tagging is not optimal if the dispersion of abilities

among the ‘high-ability people’ is too large or too narrow. After a sensitivity analysis, we conclude that the above properties are robust. Second, tagging is always suboptimal under a maxi-min social criterion (for which the aversion to inequality parameter tends to infinity).

As this paper focuses on stigmatization, utility costs such as transaction costs or the loss of privacy that applying for and receiving assistance benefits entail have not been considered. However, the model could easily be interpreted as including such utility cost.

To sum up, stigmatization which has often been neglected in the economics literature questions the robustness of previous normative conclusions about the advantages of tagging.

7 Appendix

7.1. A welfare criterion only based on income levels

The utilitarian and maxi-min criteria are standard in the literature. They take preferences as given and consider that utility functions are appropriate for a normative analysis. However, some authors have criticized this approach when individuals differ in several dimensions. For example, Fleurbaey and Maniquet (2002) have argued in favor of a distinction between “relevant” and “irrelevant” characteristics: Whereas the former call for compensation, the latter do not, because they are considered as falling within the responsibility of the individuals. In the same vein, Arneson (1990) defends a conception of social justice as equal opportunity for welfare. He also makes a distinction between the part of one’s utility for which one is responsible and the part for which is not. Following these authors, this appendix excludes from the normative criteria the parts of the utilities under control of people. Being a parameter of disutility of effort, δ should then be left aside. One could also argue that individuals are responsible for the intensity of the impact of stigmatization on their well-being (σ). Then, income should not be transferred in order to compensate for high values of σ and δ . This principle could be translated into normative criteria where σ and δ are simply ignored. This appendix is then devoted to the analysis of a welfare criterion which is only based on income levels. The social welfare functions could then use any strictly increasing and concave function of consumption. For simplicity, we here use the function $u(\cdot)$. Our alternative normative criterion is then a sum of such corrected utility functions weighted by the share in the population:

$$V \equiv \gamma \left\{ u(c_l^T)G(\tilde{\sigma}) + u(c_l)(1 - G(\tilde{\sigma})) \right\} + (1 - \gamma) \left\{ u(c_h)F(\tilde{\delta}) + u(c_l)(1 - F(\tilde{\delta})) \right\} \quad (42)$$

Totally differentiating the social welfare objective (42), the budget constraint and the two conditions defining the cut-off levels, (3) and (4), and rearranging, the change in V becomes:

$$dV = u'(c_l^T)dc_l^T + [\gamma g(\tilde{\sigma})(u(c_l^T) - u(c_l)) - (\pi_l + \pi_h)] d\tilde{\sigma} + [(1 - \gamma)f(\tilde{\delta})(u(c_h) - u(c_l)) + \pi_h] d\tilde{\delta}, \quad (43)$$

subject to the balanced budget constraint (26). The change in the utilitarian criterion (5), dW , is different and can be written as:

$$dW = u'(c_l^T)dc_l^T - (\pi_l + \pi_h)d\tilde{\sigma} + \pi_h d\tilde{\delta} \quad (44)$$

subject to the balanced budget constraint (26). Equations (43) and (44) are now employed for comparison. As in the utilitarian case (see Subsection 3.1), it is still true that $\frac{\partial V}{\partial \tilde{\sigma}} = 0$ in $\tilde{\sigma} = 0$. However, it is obvious that an additional positive term pushes $\tilde{\sigma}$ upwards as soon as $\tilde{\sigma} > 0$, namely $\gamma g(\tilde{\sigma})(u(c_l^T) - u(c_l))$. Expression $(1 - \gamma)f(\tilde{\delta})(u(c_h) - u(c_l))$ plays a similar role for $\tilde{\delta}$. Both terms express that an increase in $\tilde{\sigma}$ (respectively, $\tilde{\delta}$) has a favorable effect on Criterion V via the impact on marginal individuals who are actually indifferent. So, it is not surprising that unreported numerical simulations indicate that the global optimum is characterized by tagging as soon as stigmatization is neglected in the normative criterion.

7.2. No tagging optimal with an utilitarian criterion and under asymmetric information, a sufficient condition

This appendix shows the existence of a solution to the first-order conditions characterized by $\tilde{\sigma} = 0$ and $\tilde{\delta} > 0$. This proof is produced for logarithmic utility functions, $u(\cdot) \equiv \log_k(\cdot)$ with $k > 1$. The cut-off definitions (3) and (4) can then respectively be written as:

$$c_h = c_l k^{\tilde{\delta}} \quad (45)$$

$$c_l^T = c_l k^{\tilde{\sigma}}, \quad (46)$$

and the budget constraint becomes:

$$c_l = \frac{w_l + \pi_h(w_h - w_l)}{\pi_l^T k^{\tilde{\sigma}} + \pi_l + \pi_h k^{\tilde{\delta}}} \quad (47)$$

Substituting (45), (46) and (47) into (5), one gets an objective function of $\tilde{\sigma}$ and $\tilde{\delta}$. Maximizing this objective, the first-order conditions are:

$$\phi_{\tilde{\sigma}}(\tilde{\sigma}, \tilde{\delta}) \equiv \pi_l^T u'(c_l^T) \frac{dc_l^T}{d\tilde{\sigma}} + \pi_l u'(c_l) \frac{dc_l}{d\tilde{\sigma}} + \pi_h u'(c_h) \frac{dc_h}{d\tilde{\sigma}} = 0 \quad (48)$$

$$\phi_{\tilde{\delta}}(\tilde{\sigma}, \tilde{\delta}) \equiv \pi_l^T u'(c_l^T) \frac{dc_l^T}{d\tilde{\delta}} + \pi_l u'(c_l) \frac{dc_l}{d\tilde{\delta}} + \pi_h u'(c_h) \frac{dc_h}{d\tilde{\delta}} = 0 \quad (49)$$

where the various derivatives can be computed from (45) to (47).

$\forall \tilde{\delta}$, we know $\phi_{\tilde{\sigma}}(\tilde{\sigma} = 0, \tilde{\delta}) = 0$ (see Subsection 3.1.1). Therefore, what follows is devoted to finding a sufficient condition for $\phi_{\tilde{\delta}}(\tilde{\sigma} = 0, \tilde{\delta}) = 0$. If $\tilde{\sigma} = 0$ and $\tilde{\delta} = 0$, $\phi_{\tilde{\delta}}(\tilde{\sigma} = 0, \tilde{\delta} = 0) = u'(c_l) \frac{dc_l}{d\tilde{\delta}} > 0$. Hence, it suffices to show that $\exists \tilde{\delta}_1 > 0$ such that $\phi_{\tilde{\delta}}(\tilde{\sigma} = 0, \tilde{\delta}_1) < 0$ where $\phi_{\tilde{\delta}}(\tilde{\sigma} = 0, \tilde{\delta}_1) = (1 - \pi_h(\tilde{\delta}_1))u'(c_l) \frac{\partial c_l}{\partial \tilde{\delta}} + \pi_h(\tilde{\delta}_1)u'(c_h) \frac{\partial c_h}{\partial \tilde{\delta}}$, with $\frac{\partial c_l}{\partial \tilde{\delta}} \geq 0$ and $\frac{\partial c_h}{\partial \tilde{\delta}} = \frac{\partial c_l}{\partial \tilde{\delta}} k^{\tilde{\delta}} + c_l k^{\tilde{\delta}} \ln(k)$.

From (45) either $\frac{\partial c_h}{\partial \tilde{\delta}} > 0$ and $\frac{\partial c_l}{\partial \tilde{\delta}} < 0$ or $\frac{\partial c_h}{\partial \tilde{\delta}} < 0$, $\frac{\partial c_l}{\partial \tilde{\delta}} < 0$. In $\tilde{\sigma} = 0$, the latter case cannot occur when $\tilde{\delta}$ grows up because it would violate the budget constraint. Hence two necessary conditions for the existence of $\tilde{\delta}_1 > 0$ such that $\phi_{\tilde{\delta}}(\tilde{\sigma} = 0, \tilde{\delta}_1) < 0$ are:

$$\frac{\partial c_l}{\partial \tilde{\delta}} < 0 \quad (50)$$

$$\frac{\partial c_h}{\partial \tilde{\delta}} > 0, \text{ or } \frac{\partial c_l}{\partial \tilde{\delta}} > -c_l \ln(k) \quad (51)$$

With (47), the last inequality can be rewritten as

$$(1 - \gamma)f(\tilde{\delta}_1)(w_h - k^{\tilde{\delta}_1} w_l) > -\pi_l(w_l + \pi_h(w_h - w_l)) \ln(k) \quad (52)$$

The right-hand side is negative. Since $w_h > c_h$ and $w_l < c_l$, Equality (45) implies $w_h > k^{\tilde{\delta}_1} w_l$. So, the left-hand side of (52) is positive.

Differentiating (47) and rearranging, Inequality (50) becomes

$$\frac{f(\tilde{\delta}_1)}{F(\tilde{\delta}_1)} < \frac{w_l + \pi_h(w_h - w_l)}{w_h - k^{\tilde{\delta}_1} w_l} k^{\tilde{\delta}_1} \ln(k) \quad (53)$$

The right-hand side is positive. So, log-concavity of $F(\cdot)$ is sufficient to guarantee (53).

Finally, $\phi_{\tilde{\delta}}(\tilde{\sigma} = 0, \tilde{\delta}_1) < 0$ and the derivative of (47) with respect to $\tilde{\delta}$ can be combined to yield:

$$\frac{f(\tilde{\delta}_1)}{F(\tilde{\delta}_1)} < \frac{(u'(c_l) - u'(c_h))k^{\tilde{\delta}_1} \ln(k) \pi_l(w_l + \pi_h(w_h - w_l))}{\pi_h u'(c_h) k^{\tilde{\delta}_1} + \pi_l u'(c_l) w_h - k^{\tilde{\delta}_1} w_l}$$

The right-hand side of this inequality being positive, log-concavity of $F(\cdot)$ is a sufficient condition to guarantee the existence of an optimum characterized by $\tilde{\sigma} = 0$ and $\tilde{\delta} > 0$.

References

- [1] Akerlof, G.A. (1978), The Economics of ‘Tagging’ as Applied to the Optimal Income Tax, Welfare Programs, and Manpower Training, *American Economic Review* 68(1), 8-19.
- [2] Arneson, R.J. (1990), Liberalism, Distributive Subjectivism, and Equal Opportunity for Welfare, *Philosophy and Public Affairs* 19(2), 158-194.
- [3] Ashenfelter, O. (1983), Determining participation in income-tested social programs, *Journal of The American Statistical Association* 78(383), 517-525.

- [4] Bagnoli, M. and Bergström, T., (1989), Log-concave probability and its applications, CREST Working Paper, University of Michigan, 8923.
- [5] Besley, T. and Coate, S. (1992), Understanding welfare stigma: Taxpayer resentment and statistical discrimination, *Journal of Public Economics* 48(2), 165-83.
- [6] Bishop, J. (1982), Modelling the decision to apply for welfare. in: Garfinkel, I. (ed.), *Income-tested transfer programs: The case for and against*. Academic Press, London, 59-65.
- [7] Blank, R.M. and Ruggles, P. (1996), When Do Women Use Aid to Families with Dependent Children and Food Stamps? The Dynamics of Eligibility versus Participation, *Journal of Human Resources* 31(1), 57-89.
- [8] Blundell, R., Fry, V. and Walker, I. (1988), Modelling the take-up of means-tested benefits: the case of housing benefits in the United Kingdom, *The Economic Journal* 98 (supplement), 58-74.
- [9] Boadway, R. and Keen, M. (2000), Redistribution, in: Atkinson, A.B. and Bourguignon, F. (eds.), *Handbook of Income Distribution*, Amsterdam: Elsevier, vol.1, chap.12, 677-789.
- [10] Boadway, R., Marceau, N. and Sato, M. (1999), Agency and the design of welfare systems, *Journal of Public Economics* 73(1), 1-30.
- [11] Cournot, A. (1838), *Recherches sur les principes mathématiques de la théorie des richesses* Librairie Hachette, Paris. (new edition Paris: Marcel Riviere, 1938).
- [12] Crocker, J., Major, B. and Steele, C. (1998), Social stigma, in: Gilbert, D.T., Fiske, S.T. and Lindzey, G. (eds.), *Handbook of social psychology*, 4th ed., vol.2, Boston: McGraw-Hill, 504-553.
- [13] Currie, J. (2004), *The Take up of Social Benefits*, NBER Working Paper 10488.
- [14] De Donder, P. and Hindriks, J. (1998), The Political Economy of Targeting, *Public choice* 95(1), 177-200.
- [15] Diamond, P. and Sheshinski, E. (1995), Economic Aspects of Optimal Disability Benefits, *Journal of Public Economics* 57(1), 1-23.
- [16] Duclos, J.-Y. (1995), Modelling the Take-up of State Support, *Journal of Public Economics* 58(3), 391-415.
- [17] Duclos, J.-Y. (1997), Estimating and testing a model of welfare participation: the case of supplementary benefits in Britain, *Economica* 64(253), 81-100.
- [18] Fleurbaey, M. and Maniquet, F. (2002), Compensation and responsibility, in Arrow, K. Sen, A. and Suzumura, A. (eds.), *Handbook of Social Choice and Welfare*, Amsterdam: North-Holland.
- [19] Friedman, M. (1962), *Capitalism and freedom*, Chicago: University of Chicago Press.
- [20] Gelbach, J.B. and Pritchett, L.H. (2002), Is more for the poor less for the poor? The politics of means-tested targeting, *Topics in Economic Analysis and Policy* 2(1), 1027-1027.
- [21] Goffman, E. (1963), *Stigma: Notes on the management of Spoiled Identities*, Englewood Cliffs: Prentice-Hall.
- [22] Hancock, R., Pudney, S., Barker, G., Hernandez, M. and Sutherland, H. (2004), The take-up of means-tested benefits by British pensioners: Evidence from the Family Resources Surveys, *Fiscal Studies* 25(3), 279-303.
- [23] Mirrlees, J.A. (1971), An Exploration in the Theory of Optimum Income Taxation, *Review of Economic Studies* 38(2), 175-208.

- [24] Moffitt, R. (1983), An Economic Model of welfare Stigma, *American Economic Review* 73(5), 1023-35.
- [25] Okun, A.M. (1975), *Equality and Efficiency: The Big Tradeoff*, Washington D.C.: The Brookings Institution.
- [26] Parsons, D.O. (1996), Imperfect 'Tagging' in Social Insurance Programs, *Journal of Public Economics* 62(1), 187-207.
- [27] Rawls, J. (1971), *A theory of justice*, Oxford: Oxford University Press.
- [28] Salanié, B. (2002), Optimal Demogrants with Imperfect Tagging, *Economics Letters* 75(3), 319-324.
- [29] Sen, A.K. (1995), The political economy of targeting, in: van De Walle, D. and Nead, K. (eds.), *Public Spending and the Poor*, Baltimore: J. Hopkins University Press, 11-24.
- [30] Spicker, P. (1984), *Stigma and social welfare*, St Martin's Press, New York.
- [31] Stiglitz, J.E. (1987), Pareto Efficient and Optimal Taxation and the New New Welfare Economics, in: Auerbach, A.J. and Feldstein, M.S. (eds.), *Handbook of Public Economics*, Amsterdam: Elsevier, vol.2, chap.15, 991-1006.
- [32] Tobin, J. (1966), The Case for an Income Guarantee, *The Public Interest* 4, 31-41.
- [33] Tobin, J. (1968), Raising the Incomes of the Poor, in: Kermit, G. (ed.), *Agenda for the Nation*, Washington D.C.: The Brookings Institution, 77-116.
- [34] Waxman, C. (1983), *The stigma of poverty*, Pergamon Press, New York.
- [35] Yaniv, G. (1997), Welfare fraud and welfare stigma, *Journal of Economic Psychology* 18(4), 435-451.

Full address of the corresponding author:

Laurence Jacquet
 Department of economics
 Université Catholique de Louvain
 Place Montesquieu 3
 B-1348 Louvain-la-Neuve
 Belgium
 laurence.jacquet@uclouvain.be

TABLES AND FIGURES

Ability level (indexed by j)	claimants working in unskilled jobs ^(*)	non-claimants working in unskilled jobs ^(*)
low, $j = l$	$\gamma G(\tilde{\sigma})$	$\gamma(1 - G(\tilde{\sigma}))$
	working in skilled jobs	working in unskilled jobs ^(*)
high, $j = h$	$(1 - \gamma)F(\tilde{\delta})$	$(1 - \gamma)(1 - F(\tilde{\delta}))$

Table 1 *Distribution of individuals in the population*

(*) if $w_l > 0$, otherwise they are inactive.

($r_\delta = 1.3$)	Welfare level	Cutoff levels		Consumption levels			Proportions of popu- lation		
	W	$\tilde{\sigma}$	$\tilde{\delta}$	c_h	c_l	c_l^T	π_h	π_l	π_l^T
$r_\sigma = 0.1$	5.23	1.82	2.09	62.3	14.6	51.8	0.40	0.10	0.50
$r_\sigma = 0.2$	5.19	1.44	1.84	64.1	17.9	48.6	0.38	0.13	0.49
$r_\sigma = 0.3$	5.17	0.77	1.49	68.7	24.5	41.7	0.34	0.22	0.44
$r_\sigma = 0.4$	5.16	0.04	1.28	73.8	30.8	31.6	0.30	0.54	0.16
$r_\sigma = 0.42$	5.16	0	1.25	74.2	31.1	31.1	0.30	0.70	0

Table 2 *Simulations' results when the variance and the mean of the intensity of stigmatization (r_σ) increases and $r_\delta = 1.3$*

($r_\sigma = 0.2$)	Welfare level	Cutoff levels		Consumption levels			Proportions of popu- lation		
	W	$\tilde{\sigma}$	$\tilde{\delta}$	c_h	c_l	c_l^T	π_h	π_l	π_l^T
$r_\delta = 0.1$	5.81	0	0.51	68.4	48.2	48.2	0.47	0.53	0
$r_\delta = 0.2$	5.74	0.30	0.74	67.7	40.4	49.9	0.46	0.13	0.41
$r_\delta = 0.5$	5.57	1.10	1.37	64.1	24.9	53.2	0.45	0.07	0.48
$r_\delta = 1$	5.33	1.48	1.78	63.5	18.5	51.5	0.41	0.1	0.49
$r_\delta = 1.5$	5.11	1.36	1.84	64.7	18	46.3	0.36	0.16	0.48
$r_\delta = 2$	4.91	1.06	1.82	66.8	18.9	39.5	0.27	0.26	0.47
$r_\delta = 3$	4.61	0.28	1.75	73.2	21.8	26.4	0.13	0.47	0.40
$r_\delta = 3.5$	4.52	0.003	1.76	75.6	22.2	22.3	0.08	0.75	0.17
$r_\delta = 4$	4.45	0	1.84	76.7	21.4	21.4	0.06	0.94	0

Table 3 *Simulations' results when r_δ increases and $r_\sigma = 0.2$*

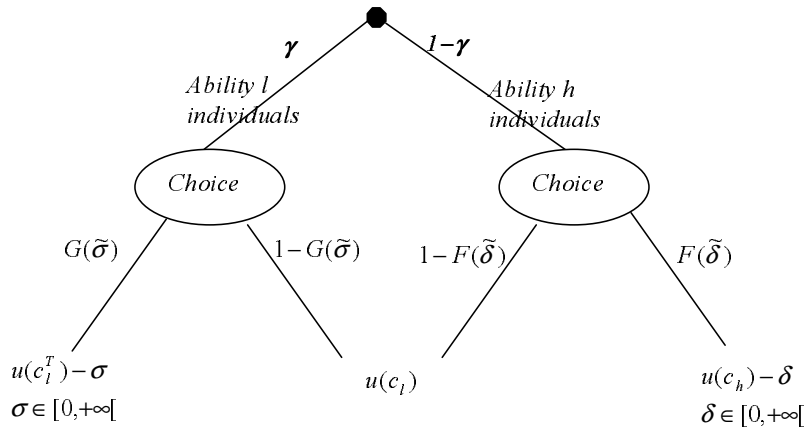


Figure 1 *Summary of the model*

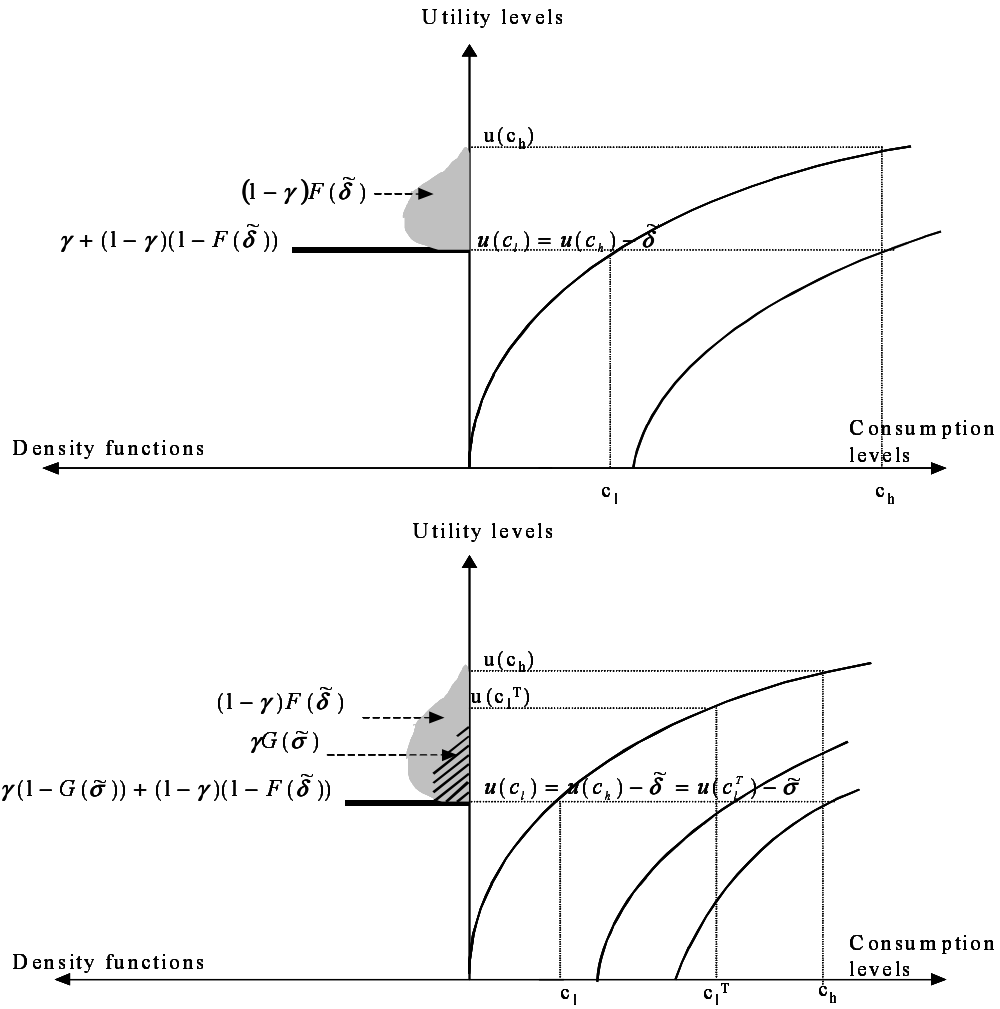


Figure 2 Graphical presentation of the optimal solutions: Without tagging (the upper-part) and with tagging

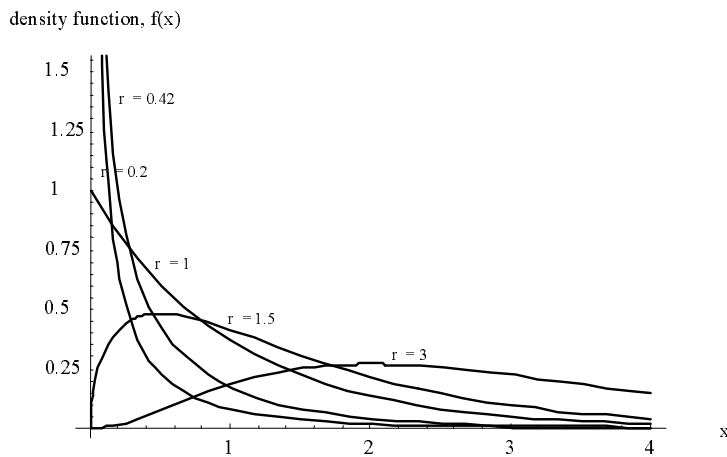


Figure 3 Gamma density functions

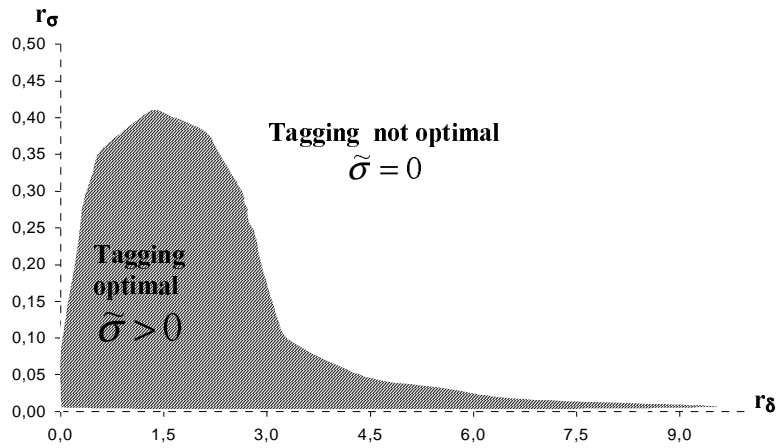


Figure 4 Simulations' results. Values of the means (and the variance) of σ and δ for which tagging is optimal

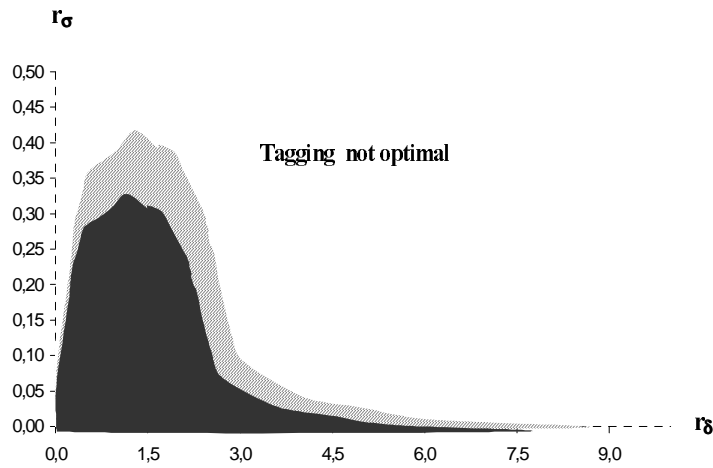


Figure 5 The reduction in the tagging area due to an increase in the basis of the logarithmic utility function (from 2 to 2.6)