School of Economics and Finance Discussion Papers

Optimal Universal and Categorical Benefits with Classification Errors and Imperfect Enforcement

Sean Slack and David Ulph

University of
St Andrews

Optimal Universal and Categorical Benefits with Classification Errors and Imperfect Enforcement

Sean Slack[*],

David Ulph[†]

Abstract

We determine the optimal combination of a universal benefit, B, and categorical benefit, C, for an economy in which individuals differ in both their ability to work – modelled as an exogenous zero quantity constraint on labour supply – and, conditional on being able to work, their productivity at work. C is targeted at those unable to work, and is conditioned in two dimensions: *ex-ante* an individual must be unable to work to be awarded the benefit , whilst *ex-post* a recipient must not subsequently work. However, the *ex-ante* conditionality may be imperfectly enforced due to Type I(false rejection) and Type II (false award) classification errors, whilst, in addition, the *ex post conditionality* may be imperfectly enforced. If there are no classification errors – and thus no enforcement issues – it is always optimal to set $C > 0$, whilst $B = 0$ only if the benefit budget is sufficiently small. However, when classification errors occur, $B = 0$ only if there are no Type I errors and the benefit budget is sufficiently small, while the conditions under which $C > 0$ depend on the enforcement of the *ex-post conditionality*. We consider two discrete alternatives. Under *No Enforcement* $C > 0$ only if the test administering C has some discriminatory power. In addition, social welfare is decreasing in the propensity to make each type of error. However, under *Full Enforcement* $C > 0$ for all levels of discriminatory power, including that of no discriminatory power. Furthermore, whilst social welfare is decreasing in the propensity to make Type I errors, there are certain conditions under which it is *increasing* in the propensity to make Type II errors. This implies that there may be conditions under which it would be welfare enhancing to lower the chosen eligibility threshold – supporting the suggestion by Goodin (1985) to "err on the side of kindness".

JEL: D6, H53, I3

Keywords: Categorical Benefit; Classification errors; Enforcement, Targeting ; Universal Benefit

1. Introduction

Partial universal welfare programmes are defined as those which provide all individuals in society with an unconditional universal benefit, but also allow for additional, targeted, assistance to those considered by the policymaker to be most in need. Many welfare systems can be interpreted as being of this type rather than the polar extremes of purely universal or purely targeted. Partial welfare programmes have been extensively analysed in the economics literature (Atkinson & Sutherland, 1989; Atkinson, 1996; Callan, O'Donoghue, Sutherland, & Wilson, 1999; DO Parsons, 1996; Salanié, 2002). Further, the merits of such partial programs, particularly in terms of administrative simplicity and political support[3], are widely discussed in the social and political science literature (Mkandawire, 2005; Skocpol, 1991; Van Parijs, 2004). The provision of a universal benefit also ensures that needy individuals have some form of financial support should they be incorrectly denied a targeted benefit by Type I (false rejection) error.

One crucial feature of any benefit system that involves targeting is that it imposes what we call *double conditionality*: *ex-ante* the award of the benefit is conditioned on an individual meeting some eligibility conditions; however there are typically also some *ex-post* requirements imposed on a recipient – e.g. in the case of disability benefits, that they do not work; in the case of unemployment benefits that they undertake sufficient job-search. Moreover these two types of conditionality are linked under the presumption that an individual who is genuinely eligible will satisfy ex-post conditionality.[4]

This leads us to a second crucial feature a benefit system that involves targeting which is that each of these two types of conditionality can be subject to enforcement errors.

- In the case of *ex-ante conditionality* there are classification errors of both Type I(false rejection) and Type II (false award). These are to some extent inevitable in targeted systems and can arise for two reasons: (i) the test or monitoring technology administering the benefit may make errors in determining where an applicant lies relative to a chosen eligibility threshold; (ii) the eligibility threshold itself may be incorrect[5].
- In the case of *ex-post conditionality* benefit authorities may fail to fully enforce the requirements placed on recipients. Again there can be two factors at work: (i) authorities may not detect all individuals who break the requirements; and,

---

[3] Skocpol (1991,p.414) refers to such programs as "targeting within universalism" . Drawing on the social policy history of the U.S., she notes that whilst targeted programs in isolation have been politically unsustainable, those that are more universal and spread benefits across groups have received political support and been effective at targeting benefits to the needy.

[4] In the case of disability benefit this would be true because an individual who is truly unable to work could not violate the ex-post conditionality by working. In the case of unemployment benefits targeted at those who are temporarily unemployed but really want to work, the sufficient job-search requirement is likely to be satisfied.

[5] Disability benefits provide a good example where both these sources of error can arise since (i) certain medical conditions are difficult to diagnose; and (ii) tests focus, in part, on the subjective assessment of ability to function in the workplace. Indeed, there has been much work exploring Type II errors and the work capability of recipients in the U.S. Social Security Disability Insurance program (Autor & Duggan, 2006; Bound, 1989, 1991; Chen & van der Klaauw, 2008; Parsons, 1991, 1980; Wachter, 2011).

further, (ii) the sanctions imposed may be insufficient to deter individuals from this behaviour.

As indicated above, these two enforcement issues are linked in the sense that if there were no Type II errors then there would be no potential for violating ex-post conditionality.

In the context of partial universal programs three central questions therefore arise:

(i)   How does the propensity of the benefit system to make Type I and Type II errors affect are the optimal levels of a universal and targeted benefit respectively? In particular, are there conditions under which it is optimal to resort to either a purely universal or purely targeted system of welfare provision, as opposed to a partial system?

(ii)   How do these error propensities affect the resulting level of social welfare?

(iii)   How do the answers to both of these questions depend on how well the *ex-post* conditionality is enforced?

The existing literature on the targeting of benefits has tended to focus on the first question.   In doing so a particular enforcement structure in relation to *ex-post conditionality* has been assumed, but there has been no comparison across enforcement regimes. Moreover, while different papers have made different assumptions about the way *ex-post conditionality* is enforced other differences in the modelling frameworks have not facilitated direct comparison.   This is illustrated in the two related contributions of Parsons (1996) and Salanié (2002) , where both papers explore the optimal provision of targeted benefits with classification errors. Parsons uses a framework where individuals are *ex-ante* identical[6] but face an exogenous probability of becoming unable to work, and allows able individuals who are tagged by Type II error to work[7]. However, in a more general framework where able indivdiuals differ over a productivity continuum, Salanié (2002) imposes the restriction that tagged able individuals do not work, and shows that it is optimal to award a higher benefit to tagged individuals than those who are not tagged. Notably, Salanié does not model the decision of individuals to apply for the targeted benefit – instead, a fixed proportion, corresponding to the Type II error probability, of able individuals are tagged and do not work. Yet, if individuals have to apply for tagged benefits then, given the enforced no work restriction, those with a sufficiently high level of productivity would choose not to apply, instead preferring to receive the lower unconditional benefit and be allowed to work.

The contribution of this paper is to address the three questions raised above within a framework that allows for a systematic comparison of how the answer to the first two questions depends on how *ex-post conditionality* is enforced. We consider a framework where individuals differ in two dimensions. There is a categorical dimension – they are either able to work or unable to work (modelled as a zero-hours quantity constraint on

---

[6] In particular, Parsons (1996) assumes that all individuals have the same productivity.

[7] Parsons (1996) refers to this as a 'dual negative income tax' scheme, whereby consumption levels are designed which incentivise tagged able individuals to work. Indeed, Parsons demonstrates that a system where tagged able indivdiuals are incentivsed to work is welfare superior to one where they are not.

labour supply) – and, for those who are able to work, their productivity differs along a continuum. The government operates a tax/benefit system that comprises: (i) a constant marginal tax rate on income; (ii) a tax-free universal benefit, B, which is received automatically by all individuals in society, (iii) a categorical benefit, C, which is targeted at those who are unable to work; and, for simplicity, (iv) a fixed budget for benefits[8].

Individuals have to apply for the categorical benefit and the benefit authority then applies a test to determine whether or not each applicant is eligible to receive the benefit. This test is subject to Type I and Type II errors, and, as in Salanié (2002), we assume that the propensity to make each Type of error is fixed and, in particular that the propensity Type II errors is independent of individual productivity. The test will be said to have:

- *perfect discriminatory* power if the propensities to make both Type I and Type II errors are both zero;
- *zero discriminatory* power if the benefit is equally likely to be awarded to both able and unable individuals;
- *some discriminatory* power in intermediate cases.

It is assumed that applications are costless in terms of time, stigma and money[9], and that no penalty attaches to an able individual applying for the categorical benefit and subsequently being awarded it through a Type II error.

Given that the categorical benefit is intended for those who are unable to work, there is an *ex-post conditionality* requirement that recipients do not work. As indicated, a number of different enforcement assumptions are made in the literature on categorical transfers administered with classification errors[10]. In this paper we consider two discrete alternative enforcement regimes: *No Enforcement* and *Full Enforcement*.

---

[8] The UK government has recently introduced an overall cap on benefits. However, the conditions under which it is optimal to adopt a targeted system, as opposed to a pure universal system, generalise to the case where there is no such cap(Slack & Ulph 2014).

[9] Given this assumption, our previous assumption that the propensity to make Type II errors is independent of individual productivity is innocuous. Specifically, because applications for the categorical benefit are costless, the utility of an applicant in the rejected state corresponds with that from having not applied. Accordingly, individuals decisions to apply for the categorical benefit will not depend on the size of Type II errors, but solely on whether the utility from receiving the categorical benefit (where this utility will be conditional on the enforcement regime in place) exceeds that from not receiving the categorical benefit.

[10] The literature on optimal benefits/transfers can be partitioned into three related strands. The first strand focuses on the design of welfare benefits when the benefit authority has no formal discriminatory test to determine eligibility. It instead chooses consumption bundles/transfer levels such that the non-needy opt against masquerading as the needy (see Besley & Coate, 1992; Blackorby & Donaldson, 1988; Cuff, 2000; P. A. Diamond & Mirrlees, 1978; Kreiner & Tranaes, 2005; Nichols & Zeckhauser, 1982). An important second strand analyses transfers within the optimal income taxation framework (Mirrlees, 1971; Sheshinski, 1972). In the standard model where individuals differ only through unobservable ability, Atkinson (1996) models a universal benefit financed by a linear income tax and allows for an unable subpopulation than cannot work. Following Akerlof (1978), a number of papers have modelled taxes and transfers when categorical information can be perfectly observed (Immonen, Kanbur, Keen, & Tuomala, 1998; Mankiw & Weinzierl, 2010; Viard, 2001a, 2001b). The third important strand – and most related to this paper – models categorical transfers administered with classification errors (see Diamond

- Under a *No Enforcement* regime there are no effective mechanisms in place to deter able recipients of C from subsequently working. In this case all such individuals will apply for the categorical benefit.
- Under the *Full-Enforcement* regime it is assumed that the probability of detection and the penalty regime are sufficiently tough that no able individual who is wrongly awarded the categorical benefit will choose to subsequently work. The categorical benefit policy is therefore two-dimensional: individuals receive a monetary amount, C, and also a fully enforced zero quantity constraint on labour supply. Since able individuals can choose whether or not to apply, it is clear that if an individual's productivity is sufficiently high the cost of the enforced labour-supply constraint will outweigh the monetary benefit and they will choose not to apply. Moreover some of those who do apply would have chosen to work under the *No Enforcement* set-up and so are constrained by the *full enforcement* of the *ex-post conditionality* requirement.

The government's objective function is taken to be strictly utilitarian (i.e. the linear sum of individual utilities). A desire for redistribution is thus captured solely through the degree of concavity that individual utility exhibits – which is an empirical proposition – as opposed to additional value judgements on society's concern for inequality (Kaplow, 2010)[11]. The government chooses the level of categorical benefit C and universal benefit, B, to maximise social welfare given (a) the benefit budget; (b) the propensity to make Type I and Type II errors and (c) the *ex-post conditionality* enforcement regime that is in place.

The major conclusions of this paper are:

- Under a *No Enforcement* regime:

(i)   It is optimal to set $C > 0$ only if the test administering C has some discriminatory power, no matter how small. If Type I errors occur with

& Sheshinski, 1995; Jacquet, 2006, 2013; Kleven & Kopczuk, 2011; DO Parsons, 1996; Salanié, 2002; Stern, 1982)

[11] In the absence of any welfare provision, there are two types of inequality in this model.
  (i)   For able individuals there is *within-group inequality* – those with higher productivity have higher absolute utility but lower social marginal utilities of income.
  (ii)   There is *between group* inequality: those who are able to work will have a higher average level of utility but lower average social marginal utility of income than those who are unable to work.

As is well known from the literature on optimal categorical benefits in the absence of classification errors (see Beath, Lewis, & Ulph, 1988; Immonen, Kanbur, Keen, & Tuomala, 1998; Viard, 2001a,b), the prime purpose of the categorical benefit is to eliminate between group inequality, as reflected in differences in the average social marginal utility of income. Whether the categorical benefit can achieve this depends on the size of the budget, and while it is always optimal to have C > 0. A point that has received somewhat less attention in the literature is that it will only be optimal to have B > 0 if between group inequality is eliminated. If C is administered with classification errors then Type I errors introduce horizontal inequity within those unable to work while Type II errors introduce horizontal inequity amongst those able to work. Whilst the Utilitarian framework does not account for these inequities beyond differences in SMVI across individuals, it is useful to note throughout the analysis the inequities that do emerge from classification errors (for a useful discussion, see Stern 1982)

positive propensity then it is always optimal to set $B > 0$ to avoid some unable individuals having no source of income to consume. However, if Type I errors do not occur and, in addition, the benefit budget falls at or below a minimum critical level, it is optimal to set $B = 0$. So if the test administering C has no discriminatory power it is optimal to adopt a *pure universal system*, if it has some discriminatory power it is optimal to adopt a targeted system. This will be a *pure targeted system* if no Type I errors are made and the budget is sufficiently small, but will be a *partial targeted system* in all other cases.

(ii)   The associated maximum social welfare is decreasing in both Type I and Type II errors respectively. This implies that while society will benefit from improvements in the accuracy of the test, changes that alter the liability threshold – i.e. make it tougher or weaker - will involve a trade-off between Type I and Type II errors.

- Under a *Full Enforcement* regime:

(i)   It is optimal to set $C > 0$ for all levels of discriminatory power. If Type I errors occurs with positive propensity then it is always optimal to set $B > 0$. If Type I errors do not occur and, in addition, the benefit budget falls at or below a minimum critical level, it is optimal to set $B = 0$. So under Full Enforcement it is *always* optimal to adopt a targeted system – a pure universal system is *never* chosen. Once more, this will be a *pure targeted system* if no Type I errors are made and the benefit budget is sufficiently small, but will be a *partial targeted system* otherwise.

(ii)   Whilst maximum welfare is unambiguously decreasing in Type I errors, there are conditions under which it can be *increasing* in Type II errors. This implies that changes that increase the liability threshold – make the test tougher – will increase Type I errors but reduce Type II errors and so are unambiguously welfare decreasing. It follows that changes which lower the liability threshold – i.e. make the test less tough – are welfare increasing. This supports the suggestion by Goodin (1985) to "err on the side of kindness".

The remainder of this paper is structured as follows. In Section 2 we set out the model and discuss how the different enforcement assumptions affect application behaviour. Section 3 then presents the main analysis: we derive the optimum benefit levels under both enforcement assumptions and explore how classification errors affect social welfare.

2. The Model

2.1. Background – Individuals

Individuals in the economy have identical preferences given by the utility function $u(x, l)$, where $x \geq 0$ denotes consumption and l; $0 \leq l \leq 1$ denotes leisure respectively. The standard assumptions apply: $u(\cdot)$ is continuous, differentiable, strictly increasing in both arguments, strictly concave and both goods are normal[12]. We also assume:

$$\lim_{x \to 0} u_x(x, l) = +\infty \, , \lim_{x \to \infty} u_x(x, l) = 0. \tag{1}$$

For an individual with net wage $\omega \geq 0$ and unearned income $M > 0$, labour supply, $H^*(\omega, M)$; $0 \leq H \leq 1$, and the associated indirect utility function, $v(\omega, M)$, are given by:

$$H^*(\omega, M) \equiv \underset{0 \leq H \leq 1}{ARGMAX} \, u(\omega H + M, 1 - H)$$

$$v(\omega, M) = \underset{0 \leq H \leq 1}{MAX} \, u(\omega H + M, 1 - H) = u(\omega H^* + M, 1 - H^*).$$

For all $\omega \geq 0$ we have $v_M(\omega, M) = u_x(\omega H^* + M, 1 - H^*) > 0$, and, from Roy's identity, $v_\omega(\omega, M) = v_M(\omega, M) H^*(\omega, M) > 0$. It is also straightforward to show that concavity of $u(x, l)$ implies diminishing marginal indirect utility of unearned income, i.e. $v_{MM}(\omega, M) < 0$.

Let $\overline{\omega}(M)$ denote the reservation wage at or below which an individual with lump sum unearned income M will choose not to work. Formally, $\overline{\omega}(M)$ is defined by $H^*[\overline{\omega}(M), M] = 0$ and satisfies $\overline{\omega}(M) = u_l(M, 1)/u_x(M, 1)$. Furthermore, given that leisure is normal $\overline{\omega}'(M) > 0$.

For $\omega \leq \overline{\omega}(M)$ we have $H^*(\omega, M) \equiv 0$ which implies $v(\omega, M) \equiv u(M, 1)$, and so $v_M(\omega, M) \equiv u_x(M, 1)$. So, over this range of net wage rates the marginal indirect utility of unearned income is constant and independent of the wage rate. On the other hand, for $\omega > \overline{\omega}(M)$, Roy's identity and the assumption that leisure is normal together imply that $v_{\omega M} = v_{M\omega} = v_{MM} H^* + v_M H_M^* < 0$. [13] So over this range of net wages the marginal indirect utility of unearned income is a strictly decreasing function of the net wage. Formally:

$$\forall \omega > \overline{\omega}(M), v_{\omega M} < 0 \Rightarrow v_M(\omega, M) < v_M(\overline{\omega}(M), M) = u_x(M, 1) \tag{2}$$

2.2. The Population and Tax-Benefit System

In a population of size 1, there are two distributional issues. First, the fraction θ; $0 < \theta < 1$, of individuals face a zero quantity constraint on labour supply and are thus unable to work. Absent any provision of state financial support, these individuals

---

[12] Formally, where subscripts denote partial derivatives, we have: $u_x > 0, u_{xx} < 0, u_l > 0, u_{ll} < 0, u_{xx}u_{ll} - (u_{xl})^2 \geq 0$ (by concavity), $2u_{xl} - (u_x/u_l)u_{ll} - (u_l/u_x)u_{xx} \geq 0$ (by implied quasiconcavity) and $u_l u_{xx} - u_x u_{xl} < 0$ by normality of leisure.

[13] For a useful discussion, see (Christiansen, 1983;p. 367).

would have no source of income. Second, the remaining fraction $(1 - \theta)$ of individuals are able to work, but differ in their productivity, thereby giving rise to earned income inequality.

There is a tax-benefit system comprising four elements:

(i)   A constant marginal tax rate, t; $0 < t < 1$, on all earned income;
(ii) A tax-free universal benefit, $B \geq 0$, paid automatically to everyone[14];
(iii) A tax-free categorical benefit, $C \geq 0$, that is targeted – potentially imperfectly – at those unable to work and is received in addition to B.
(iv) A fixed benefit budget, $\beta > 0$, to be spent on the universal and/or the categorical benefits[15].

We take t as fixed such that productivity differences amongst the able are reflected through the net wage, ω, which is distributed with density $f(\omega)$, where $f(\omega) > 0 \; \forall \omega \geq 0$ and $\int_0^\infty f(\omega)d\omega = 1$. The associated distribution function is $F(\omega) = \int_0^\omega f(z)dz$, where $0 \leq F(\omega) \leq 1$. Note from (2) that:

$$\int_0^\infty v(\omega, M) f(\omega) d\omega > u(M, 1); \quad \int_0^\infty v_M(\omega, M) f(\omega) d\omega < u_x(M, 1) \qquad (3)$$

So conditional on unable and able individuals receiving the same unearned income, the average marginal utility of income for the unable must exceed that of the able. This inequality can be reduced through targeting additional resources at the unable subpopulation.

Given a fixed benefit budget of size β, the government may choose to provide a universal benefit B for everyone and/or target a categorical benefit C at the unable. Whilst B is automatically received, C must be applied for and is subject to what we term *double conditionality:*

- *Ex-ante conditionality:* An individual who has applied for C must be unable to work.
- *Ex-post conditionality*: A recipient of C must not subsequently work.[16]

An individual who has applied for C is subject to a test to determine whether they are indeed unable to work – and thus determine whether *ex-ante conditionality* is satisfied. However, this test may be imperfect and, in statistical parlance, subject to Type I and/or Type II classification errors. A Type I (false negative) error occurs when an individual who is unable to work is incorrectly classified as being able to work. We denote the probability of a Type I error occurring by p_I; $0 \leq p_I \leq 1$. Contrastingly, a Type II (false positive) error arises when an individual who is able to work is incorrectly awarded C. For simplicity, we assume that the probability of Type II error is independent of

---

[14] With just these first two elements, we would have effectively a simple negative income tax system.

[15] Our results for when it is optimal to adopt a pure universal system or a targeted system continue to hold when the assumption of a fixed budget is replaced by an endogenous benefit budget financed by tax revenue (Slack & Ulph 2014)

[16] Indeed, since C is targeted at the unable, it is *de facto* conditional on not subsequently working.

productivity and denote this by p_{II}; $0 \leq p_{II} \leq 1$[17]. We confine our attention to error probabilities satisfying:

$$p_I + p_{II} \leq 1 \tag{4}$$

If $p_I = p_{II} = 0$ then we say that the test awarding C has perfect discriminatory power[18]. Contrastingly, if $p_I + p_{II} = 1$ then we say that the test has no discriminatory power because it awards C to an able applicant with the same probability that it awards C to an unable applicant. For the intermediate cases where $0 < p_I + p_{II} < 1$, we say that the test has some, but not perfect, discriminatory power.

2.3. Benefit Applications and the Enforcement of ex-post conditionality

Given that applications for C are costless, an unable individual will always choose to apply. Turning to able individuals, we assume that no checks or penalties are in place for an able individual who applies for C and, if awarded it, does not work. Further, there are no penalties for being denied C. However, for individual's with productivities sufficiently high that they would consider working when receiving C, their application decision will depend on how the *ex-post conditionality* is enforced. In this regard, we analyse two extreme enforcement assumptions where the requirement that recipients do not work is either (i) not enforced at all or is (ii) fully enforced. We refer to these a *No Enforcement* and *Full Enforcement* regimes and detail these below.

2.3.1. No Enforcement

In this first case we assume that there are no effective mechanisms in place to deter an able recipient of C from working should they wish to. In this sense, benefit policy is one-dimensional because recipients receive only the monetary amount C, with no subsequent restrictions on labour supply. Accordingly, all able individuals along the net wage continuum will choose to apply for C. Of the proportion $(1 - p_{II})$ who are correctly denied the benefit, all those with $\omega \leq \overline{\omega}(B)$ will not work, whilst those with $\overline{\omega}(B) < \omega$ will work. Turning to the proportion p_{II} who are awarded C, all those with $\omega \leq \overline{\omega}(B + C)$ will not work, whilst those with $\overline{\omega}(B + C) < \omega$ will work (see Figure 1 and Figure 2(i) in the Appendix). Note that:

$$\overline{\omega}(B) < \overline{\omega}(B + C) \ \forall \ C > 0 \tag{5}$$

2.3.2. Full Enforcement

In this second alternative case, we assume that there are totally effective mechanisms in place that fully deter any able individual from working whilst receiving C. So benefit policy now has two dimensions: Individuals who are awarded the benefit receive both the monetary amount C *and* an enforced zero quantity constraint on labour supply. As under No Enforcement, those with $\omega \leq \overline{\omega}(B)$ will apply for C because they would never

---

[17] For other applications of this assumption, see, for example, Salanié (2002).

[18] If no Type II errors are made, then *ex-post* conditionality is automatically satisfied because the only individuals who receive the categorical benefit are those who are unable to work.

choose to work. However, those with $\omega > \overline{\omega}(B)$ would choose to work when receiving just B and so will only apply for C if $\omega \leq \overline{\overline{\omega}}(B, C)$, where:

$$v[\overline{\overline{\omega}}(B, C), B] \equiv u(B + C, 1) \tag{6}$$

So $\overline{\overline{\omega}}(B, C)$ is the net wage at which an able individual is just indifferent between (i) not working and receiving total benefit income $B + C$, and (ii) working as much as desired and receiving only B in benefit income (see Figure 1 and Figure 2(ii) in the appendix). It is straightforward to show that:

$$\overline{\overline{\omega}}_C > \overline{\overline{\omega}}_B > 0 \tag{7}$$

Finally, given that $v[\overline{\omega}(B + C), B] = v[\overline{\overline{\omega}}(B, C), B] = u(B + C, 1)$, we have:

$$\overline{\omega}(B + C) < \overline{\overline{\omega}}(B, C) \ \forall \ C > 0 \tag{8}$$

Note that $\overline{\overline{\omega}}(B, C) \to \overline{\omega}(B)$ as $C \to 0$ and $\overline{\overline{\omega}}(B, 0) = \overline{\omega}(B)$.

2.4.   Policy Constrained Individuals

Under Full Enforcement, the only recipients of C for whom the enforced zero quantity constraint on labour actually binds are those with $\overline{\omega}(B + C) < \omega \leq \overline{\overline{\omega}}(B, C)$. Under the No Enforcement regime, we know from (5) that these individuals would optimally choose to work when receiving benefit income $B + C$. They are therefore constrained by the fully enforced benefit policy. A comparison between Figure 2(i) and Figure 2(ii) in the Appendix graphically illustrates this.

3.   Analysis

We now determine the conditions under which (i) $B > 0, C > 0$, (ii) $B = 0, C > 0$ or (iii) $B > 0, C = 0$ [19] for, first, the benchmark case where categorical status is perfectly observed, and second, the more interesting cases where classification errors are made and the conditionality of C may or may not be enforced. For both enforcement assumptions we will determine how classification errors affect the value function of social welfare (i.e. maximum social welfare). In all cases, the social welfare function that the government seeks to maximize is strictly utilitarian, and thus given by the sum of individual (expected) utilities.

3.1.   Perfect Discrimination

When the test administering C can perfectly discern the unable from the able, we can write social welfare as:

$$W^P(B, C; \theta) = \theta u(B + C, 1) + (1 - \theta) \int_0^\infty v(\omega, B) f(\omega) d\omega \tag{9}$$

---

[19] That is, we determine the conditions – if any – under which it is optimal to adopt (i) a partial universal benefit system, (ii) a purely targeted benefit system, or (iii) a pure universal benefit system.

Given the fixed budget size β available for expenditure on B and C, the problem is to choose B and C so as to maximize social welfare. Formally, denote maximum social welfare by:

$$V^P(\beta, \theta) = \underset{B,C}{MAX}\, W^P(B, C; \beta, \theta) \quad \text{s.t.}\ \ B + \theta C = \beta\,, B \geq 0, C \geq 0 \tag{10}$$

Notice that the budget constraint must hold with equality because, if underspent, welfare could always be raised through paying a higher B to all individuals. If we denote the optimum benefit levels by $\hat{B}^P$ and $\hat{C}^P$ respectively we have:

Proposition 1:

$$\text{(i)}\ \ \hat{C}^P > 0\,,$$

$$\text{(ii)}\ \ \hat{B}^P \begin{Bmatrix} > \\ = \end{Bmatrix} \text{ if } \beta \begin{Bmatrix} > \\ \leq \end{Bmatrix} \bar{\beta}^P \tag{11}$$

where $\bar{\beta}^P$ is defined by:

$$\int_0^\infty v_M(\omega, 0)f(\omega)d\omega \equiv u_x\left(\frac{\bar{\beta}^P}{\theta}, 1\right) \tag{12}$$

Proof (See Appendix)

Proposition 1 states that, whilst it is always optimal to set $C > 0$, it is only optimal to set $B > 0$ if inequality in the average social marginal utility of income between the unable and able subpopulations can be eliminated without exhausting the benefit budget[20]. The perfect discrimination optimum is therefore characterised by:

$$u_x\left(\hat{B}^P + \hat{C}^P, 1\right)\begin{Bmatrix} > \\ = \end{Bmatrix} \int_0^\infty v_M\left(\omega, \hat{B}^P\right)f(\omega)d\omega \text{ if } \beta \begin{Bmatrix} < \\ \geq \end{Bmatrix} \bar{\beta}^P\,, \tag{13}$$

It must therefore hold that $B^P + \hat{C}^P > \beta > \hat{B}^P$.

3.2. Imperfect Discrimination

3.2.1. No Enforcement

With no enforcement mechanisms in place to restrict the work behaviour of able individuals who receive C, we know that all able individuals will apply for C. Social welfare is therefore now given by:

---

[20] There is therefore an ordering or priorities. The first priority is to eliminate, if possible, between-group inequality in the average social marginal utility of income – i.e. support unable individuals because they are the most needy in society from the perspective of marginal utility of income. Conditional on this being achieved and the benefit budget not being exhausted, the second priority is to spend the remainder of the benefit budget on the universal benefit.

$$W^N(B, C; p_I, p_{II}, \theta)$$

$$= \theta\{(1 - p_I)u(B + C, 1) + p_I u(B, 1)\}$$

$$+(1 - \theta)\begin{cases} p_{II}\left[F[\bar{\omega}(B + C)]u(B + C, 1) + \displaystyle\int_{\bar{\omega}(B+C)}^{\infty} v(\omega, B + C)f(\omega)d\omega\right] \\ \quad + (1 - p_{II})\left[F[\bar{\omega}(B)]u(B, 1) + \displaystyle\int_{\bar{\omega}(B)}^{\infty} v(\omega, B)f(\omega)d\omega\right] \end{cases} \tag{14}$$

Notice that $W^N(B, C; 0, 0, \theta) = W^P(B, C; \theta)$.[21] The first line on the right side of (14) illustrates that horizontal inequity is introduced into the unable subpopulation. These individuals derive consumption solely from unearned income and so any disparities in their utility arise due to unequal treatment by the benefit system. The proportion p_I are incorrectly denied C, whilst the proportion $(1 - p_I)$ are correctly awarded it. The second line demonstrates that the proportion p_{II} of able individuals are incorrectly awarded C. There is also an element of horizontal inequity introduced for those $\omega \le \bar{\omega}(B)$ who are voluntarily unemployed when receiving only B. Whilst these individuals differ in their productivity, they share the same utility level when receiving B. Some will be awarded the benefit C by Type II error, whilst others will not, thus giving rise to the horizontal inequity[22].

Let the proportion of the population who are awarded C be given by $k^N = \theta(1 - p_I) + (1 - \theta)p_{II}$. Maximum social welfare is therefore given by:

$$V^N(p_I, p_{II}, \beta, \theta) = \underset{B,C}{MAX}\ W^N(B, C; p_I, p_{II}, \theta)$$

$$\text{s.t. } B + \theta k^N = \beta, \qquad B \ge 0, C \ge 0 \tag{15}$$

If we now denote the optimal benefit levels by $\hat{B}^N$ and $\hat{C}^N$ respectively we have:

Proposition 2a:

(i) $\hat{C}^N \left\{\begin{matrix}>\\=\end{matrix}\right\} 0 \ \forall \ p_I + p_{II} \left\{\begin{matrix}<\\=\end{matrix}\right\} 1$

(ii) $\hat{B}^N > 0 \ \forall \ p_I > 0. \ \ \text{If } p_I = 0, \ \hat{B}^N \left\{\begin{matrix}=\\>\end{matrix}\right\} 0 \ as \ \beta \left\{\begin{matrix}\le\\>\end{matrix}\right\} \bar{\beta}^N \ \forall \ 0 \le p_{II} < 1$
                                                                                 (16)

where $\bar{\beta}^N$ is the critical budget level satisfying:

$$k^N u_x\left(\frac{\bar{\beta}^P}{\theta}, 1\right) \equiv \theta u_x\left(\frac{\bar{\beta}^N}{k^N}, 1\right) + (1 - \theta)p_{II}\int_0^{\infty} v_M\left(\omega, \frac{\bar{\beta}^N}{k^N}\right)f(\omega)d\omega. \tag{17}$$

Proof: See Appendix.

The principal messages from Proposition 2 are that (i) a necessary and sufficient condition to have a positive categorical benefit is that the test administering the categorical benefit has some discriminatory power, no matter how small, and (ii) whenever Type I errors are made it is always optimal to have a universal benefit - to avoid some unable individuals having zero consumption. However even if no Type I errors are made then it will only be optimal to have a universal benefit if the total benefit budget is above a minimum critical level. Put differently, if the test administering the categorical benefit has no discriminatory power, it is optimal to adopt a pure universal system (i.e. $B = \beta, C = 0$). Otherwise a targeted system should be used. This will be a pure targeted system (i.e. $B = 0, C = \beta/k^N$) if there are no Type I errors and the benefit budget is sufficiently small. In all other cases it is optimal to have a partial targeted system (i.e. $B > 0, C > 0$).

The implications for maximum social welfare are that:

$$V^N(0,0,\beta,\theta) = V^P(\beta,\theta), \qquad V^N(p_I, 1 - p_I, \beta, \theta) = W^U(\beta,\theta), \tag{18}$$

where W^U is the level of social welfare under a pure universal system, given budget a budget size β:

$$W^U(\beta,\theta) = \theta u(\beta, 1) + (1 - \theta) \int_0^\infty v(\omega, \beta) f(\omega) d\omega \tag{19}$$

To provide some intuition for Proposition 2a(i), let $\bar{M}^D$ and $\bar{M}^{AN}$ denote the benefits received, on average, by the unable and able respectively, under an arbitrary, budget feasible, targeted system (i.e. $B < \beta, C = (\beta - B)/k^N$). Formally:

$$\bar{M}^D = B + (1 - p_I) \cdot \frac{\beta - B}{k^N} , \qquad \bar{M}^{AN} = B + p_{II} \cdot \frac{\beta - B}{k^N} \tag{20}$$

It is straightforward to see that:

$$\bar{M}^{AN} < \beta < \bar{M}^D \ \forall \ p_I + p_{II} < 1$$

$$\bar{M}^{AN} = \beta = \bar{M}^D \ \forall \ p_I + p_{II} = 1 \tag{21}$$

So setting $C > 0$ only provides the unable with more, on average, than the unable, if the test administering C has some discriminatory power. If the test has no discriminatory power then the benefits received by the two subpopulations are, on average, the same.

To save on notation, we denote the expected utility of unable and able individuals by $\mathrm{E}[u(M, 1)]$ and $\mathrm{E}[v(\omega, M)]$ respectively[23]. The relevant considerations in comparing a targeted system with a pure universal system are therefore captured by:

---

[23] Where $\mathrm{E}[u(M, 1)] = p_I u(B, 1) + (1 - p_I) u(B + C, 1)$, and $\mathrm{E}[v(\omega, M)] = (1 - p_{II}) v(\omega, B) + p_{II} v(\omega, B + C)]$.

$$W\left[B < \beta, \left(\frac{\beta - B}{k^N}\right); p_I, p_{II}, \beta, \theta\right] - W^U(\beta, \theta) =$$

$$\theta\{E[u(M, 1)] - u(\bar{M}^D, 1)\} + (1 - \theta)\int_0^\infty \{E[v(\omega, M)] - v(\omega, \bar{M}^{AN})\}f(\omega)d\omega \qquad (22)$$

$$+ \theta\{u(\bar{M}^D, 1) - u(\beta, 1)\} + (1 - \theta)\int_0^\infty \{v(\omega, \bar{M}^{AN}) - v(\omega, \beta)\}f(\omega)d\omega$$

The first line on the right side captures how a targeted system performs relative to a pure universal system in terms *within-group* inequality in utility levels. By the concavity of utility it is (i) negative when $p_I > 0$ and/or $p_{II} > 0$ and (ii) zero when $p_I = p_{II} = 0$. The second line, meanwhile, captures how a targeted system performs relative to a pure universal system in terms of *between-group* inequality in average utility levels. This line is (i) positive when $p_I + p_{II} < 1$ because $\bar{M}^D > \beta > \bar{M}^{AN}$ [24]and (ii) zero when $p_I + p_{II} = 1$ because $\bar{M}^D = \beta = \bar{M}^{AN}$. We thus have:

$$W^N\left(B < \beta, \frac{\beta - B}{A}; p_I, 1 - p_I\right) < W^U(\beta, 0) \qquad (23)$$

So when the test administering C has no discriminatory power, a targeted system does worse than a universal system in terms of within-group inequality (in utility levels), and no better in terms of between-group inequality (in terms of average utility levels), which provides the intuition as to why $\hat{C}^N = 0 \ \forall \ p_I + p_{II} = 1$.

The intuition for Proposition 2a(ii) is firstly that if there are Type I errors and no universal benefit then unable individuals who are wrongly denied the categorical benefit will have no source of income to provide consumption. Secondly even if there are no Type I errors, the first priority of a benefit system is to support the most needy – those unable to work – and a universal benefit should be used only if the budget is above a critical level.

We now turn our attention to how classification errors affect maximum social welfare. From the FOCs characterising the optimum benefit levels (see Appendix) it is straightforward to show that:

$$u_x(\hat{B}^N + \hat{C}^N, 1) \left\{\substack{> \\ =}\right\} \hat{\lambda}^N > \int_0^\infty v_M(\omega, \hat{B}^N)f(\omega)d\omega \ \forall \ p_I > 0, p_{II} \left\{\substack{> \\ =}\right\} 0,$$

$$u_x(\hat{B}^N + \hat{C}^N, 1) > \hat{\lambda}^N \left\{\substack{> \\ \geq}\right\} \int_0^\infty v_M(\omega, \hat{B}^N)f(\omega)d\omega \ \forall \ p_I \left\{\substack{> \\ =}\right\} 0, p_{II} > 0 \qquad (24)$$

This implies that:

---

[24] A first-order Taylor approximation around β on the final line of **(22)** gives:

$$\theta(\bar{M}^U - \beta)\left[u_x(\beta, 1) - \int_0^\infty v_M(\omega, \beta)f(\omega)d\omega\right] \geq 0$$

So if $p_I + p_{II} < 1 \Rightarrow \bar{M}^D > \beta$ this is positive.

$$\text{E}[u_\chi(M,1)] > \hat{\lambda}^N > \text{E}[v_M(\omega,M)] \ \forall \ 0 < p_I + p_{II} < 1. \tag{25}$$

So whenever there are classification errors of either Type I or Type II or both, the optimal targeted system fails to eliminate *between-group* inequality in the average social marginal utility of income. Given (24) we can establish the following result:

Proposition 2b: *Maximum social welfare is decreasing in the propensity to make classification errors of either Type I or Type II. Formally:*

$$\frac{\partial V^N}{\partial p_I} < 0 \ , \frac{\partial V^N}{\partial p_{II}} < 0 \ ; \ \forall \ 0 \le p_I + p_{II} < 1 \tag{26}$$

Proof: See Appendix.

This concludes our discussion of the No Enforcement regime. We now proceed to analyse the Full Enforcement regime.

3.2.2. Full Enforcement

When the condition that recipients of C do not work is fully enforced, we know from (6) that the only able individuals who will apply for C will be those with $\omega \le \bar{\bar{\omega}}$. Given that the unable always apply for C, we denote the fraction of the total population who receive C by $k^F \equiv \theta(1 - p_I) + (1 - \theta)p_{II}F(\bar{\bar{\omega}}); 0 < k^F < 1$. The proportion of the population that does not receive C is therefore given by $(1 - k^F) = \theta p_I + (1 - \theta)\{(1 - p_{II})F(\bar{\bar{\omega}}) + [1 - F(\bar{\bar{\omega}})]\}$. It is composed of those who apply for C but are rejected it, and also those who do not apply for C. Let:

$$\alpha = \frac{\theta p_I}{1 - k^F} \ , \qquad \chi = \frac{(1 - \theta)(1 - p_{II})F(\bar{\bar{\omega}})}{1 - k^F}; \ \ \alpha + \chi < 1 \tag{27}$$

First, α denotes the proportion of these not receiving C who are unable, and thus incorrectly denied the benefit by Type I error. Second, χ denotes the proportion of those not receiving C who are able and applied for C, but were correctly denied it. Notice that $\alpha + \chi < 1$ because only those with $\omega < \bar{\bar{\omega}}$ from the able subpopulation apply for C.

Putting this all together we can write social welfare[25] as:

$$W^F(B,C;p_I,p_{II},\theta)$$

$$= k^F u(B+C,1) + (1 - k^F) \begin{bmatrix} \alpha u(B,1) + \chi \int_0^{\bar{\bar{\omega}}} v(\omega,B)g(\omega)d\omega \\ +(1 - \alpha - \chi)\int_{\bar{\bar{\omega}}}^\infty v(\omega,B)h(\omega)d\omega \end{bmatrix} \tag{28}$$

---

[25] In the Appendix we give a more explicit expression in terms of the underlying parameters.

Where $g(\omega) \equiv f(\omega)/F(\bar{\bar{\omega}})$, $h(\omega) \equiv f(\omega)/[1 - F(\bar{\bar{\omega}})]$ and $\int_0^{\bar{\bar{\omega}}} g(\omega)d\omega = \int_{\bar{\bar{\omega}}}^{\infty} h(\omega)d\omega = 1$.

Maximum welfare under the Full Enforcement is therefore:

$$V^F(p_I, p_{II}, \beta, \theta) = \underset{B,C}{MAX}\ W^F(B, C; p_I, p_{II}, \theta)$$

$$\text{s.t.}\ \ B + k^F C = \beta\ , B \geq 0, C \geq 0$$

(29)

If we denote the optimum benefit levels by $\hat{B}^F$ and $\hat{C}^F$ we have:

Proposition 3a:

(i) $\hat{C}^F > 0\ \forall\ p_I + p_{II} \leq 1$

(ii) $\hat{B}^F > 0\ \forall\ p_I > 0.$ If $p_I = 0$, $\hat{B}^F \begin{Bmatrix} = \\ > \end{Bmatrix} 0$ as $\beta \begin{Bmatrix} \leq \\ > \end{Bmatrix} \bar{\beta}^F\ \forall\ 0 \leq p_{II} < 1$

(30)

The principal messages from Proposition 3 are that (i) it is optimal to provide a categorical benefit for *all* levels of discriminatory power – therefore including the case of no discriminatory power – and (ii) whenever Type I errors are made it is always optimal to have a universal benefit to avoid some unable individuals having zero consumption. Further, even if no Type I errors are made it will only be optimal to provide a universal benefit if the total benefit budget is above a minimum critical level, $\bar{\beta}^F$. Put differently, there are *no* conditions under Full Enforcement where a pure universal system (i.e. $B = \beta, C = 0$) would be adopted. Instead, a targeted system should always be used. This will be a pure targeted system (i.e. $B = 0, C = \beta/k^F$) if there are no Type I errors and the benefit budget is sufficiently small. In all other cases, it is optimal to have a partial

In comparing Proposition 3a(i) and Proposition 2a(i), the principal difference between the two systems – in term of benefit design – is that under No Enforcement it is optimal to adopt a pure universal system whenever there is no discriminatory power, whilst under Full Enforcement it is never optimal to adopt a pure universal system. Put differently, the Full Enforcement benefit system will always involve targeting, whereas the No Enforcement system will only involve targeting if there is some discriminatory power.

To provide some intuition for Proposition 3a(i), recall that able recipients of C with $\bar{\omega}(B + C) < \omega < \bar{\bar{\omega}}$ are quantity constrained by the Full Enforcement of the *ex-post* conditionality. Under the No Enforcement regime, these individuals would have chosen to work when receiving C. Given that the unable face an exogenous zero quantity constraint on labour, we can therefore write the proportion of the population who are in fact quantity constrained as:

$$\theta^Q = \theta + (1 - \theta)p_{II}\{F[\bar{\bar{\omega}}(B, C)] - F[\bar{\omega}(B + C)]\}\ ;\ \frac{\partial \theta^Q}{\partial p_{II}} > 0 \qquad (31)$$

Unsurprisingly, θ^Q is increasing in p_{II} because this increases the number of individuals with $\overline{\omega}(B + C) < \omega \leq \overline{\overline{\omega}}$ who are awarded C and become quantity constrained. So of all the individuals in the economy who are quantity constrained, the proportion who do not receive C must therefore be :

$$p_I^\infty = \frac{\theta p_I}{\theta^Q} = \frac{p_I}{1 + \left(\frac{1 - \theta}{\theta}\right) p_{II}\{F(\overline{\overline{\omega}}) - F[\overline{\omega}(B + C)]\}} \leq p_I \tag{32}$$

Note that $p_I^\infty < p_I \; \forall \; p_I > 0$ and $p_{II} > 0$, whilst $p_I^\infty = p_I$ whenever $p_{II} = 0$ and/or $p_I = 0$. Further, because an increase in p_I increases the number of unable individuals who are denied C, whilst an increase in p_{II} increases the number of quantity constrained individuals, we have:

$$\frac{\partial p_I^\infty}{\partial p_I} > 0, \qquad \frac{\partial p_I^\infty}{\partial p_{II}} < 0 \tag{33}$$

Next, from (31) the proportion of the population who are not quantity constrained in their labour supply is given by $(1 - \theta^Q) = (1 - \theta)\{[1 - F(\overline{\overline{\omega}})] + F(\overline{\overline{\omega}})(1 - p_{II}) + p_{II}F[\overline{\omega}(B + C)]\}$. The proportion of unconstrained individuals who receive C can only be composed of those with $\omega \leq \overline{\omega}(B + C)$, and is thus given by:

$$p_{II}^\infty \equiv \frac{p_{II}F[\overline{\omega}(B + C)]}{[1 - p_{II}F(\overline{\overline{\omega}})] + p_{II}F[\overline{\omega}(B + C)]} = \frac{1}{1 + \left[1 - \frac{p_{II}F(\overline{\overline{\omega}})}{p_{II}F[\overline{\omega}(B + C)]}\right]} \tag{34}$$

If $p_{II} = 0$ then $p_{II}^\infty = p_{II} = 0$. However, if $0 < p_{II} \leq 1$ then $p_{II}^\infty < p_{II} \; \forall \; 0 < p_{II} \leq 1$. It is also straightforward to see that (see Appendix):.

$$\frac{\partial p_{II}^\infty}{\partial p_{II}} > 0 \tag{35}$$

An increase in p_{II} can therefore be seen to generate three effects. First, it increases the proportion of the population who face a zero quantity constraint on labour. Second, it improves the targeting of C towards individuals who are quantity constrained through lowering p_I^∞. Finally however, it worsens targeting through awarding C to those who are not constrained.

Let us define the benefit authority's *effective level* of discriminatory power by $p_I^\infty + p_{II}^\infty$, where Type I and Type II errors are appropriately redefined to account for the fact that benefit policy constrains the labour supply of those with $\overline{\omega}(B + C) < \omega \leq \overline{\overline{\omega}}(B, C)$. From (32) and (36) it must hold that, when $p_{II} > 0$:

$$p_I^\infty + p_{II}^\infty < p_I + p_{II} \leq 1 \tag{36}$$

The test administering C therefore always has some effective level of discriminatory power, and this helps explain why $\hat{C}^F > 0 \; \forall \; p_I + p_{II} \leq 1$.

We now turn our attention to how classification errors affect maximum social welfare. From the FOCs characterising the optimum benefit levels (see Appendix) we can show that:

$$\int_0^{\overline{\overline{\omega}}} v_M(\omega, \hat{B}^F)g(\omega)d\omega > u_x(\hat{B}^F + \hat{C}^F, 1) \geq \hat{\lambda}^F > \int_{\overline{\overline{\omega}}}^{\infty} v_M(\omega, \hat{B}^F)h(\omega)d\omega \qquad (37)$$

This states that that average social marginal utility of rejected able applicants exceeds that of recipient of the categorical benefit, which in turn exceeds the average social marginal utility of those who do not apply for the categorical benefit. So classification errors prevent the categorical benefit from eliminating between-group inequality in the average social marginal utility of income between the unable and able subpopulations.

Given (37) we can establish the following result:

Proposition 3b: *Maximum social welfare is decreasing in the propensity to make Type I classification errors, but there are conditions under which it is increasing in the propensity to make Type II classification errors. Formally:*

$$\frac{\partial V^F}{\partial p_I} < 0 \quad , \frac{\partial V^F}{\partial p_{II}} \left\{ \begin{matrix} \leq \\ > \end{matrix} \right\} 0 \qquad (38)$$

where:

$$\underbrace{\frac{F(\overline{\omega})}{\frac{\theta(1-p_I)}{(1-\theta)p_{II}} + F(\overline{\overline{\omega}})}} \cdot \underbrace{\frac{\overline{\overline{\omega}}f(\overline{\overline{\omega}})}{F(\overline{\overline{\omega}})}}_{\substack{elasticity\ of \\ F\ wrt\ \overline{\overline{\omega}}}} \cdot \underbrace{\frac{\hat{C}^F \overline{\overline{\omega}}_C}{\overline{\overline{\omega}}}}_{\substack{elasticity\ of \\ \overline{\overline{\omega}}\ wrt\ C}} - \left[1 - \frac{F(\overline{\omega})}{F(\overline{\overline{\omega}})} \right] > 0 \Rightarrow \frac{\partial V^F}{\partial p_{II}} > 0$$

Proof: First, it is straightforward to demonstrate that social welfare is falling in Type I errors:

$$\frac{\partial V^F}{\partial p_I} = \theta\{[u(\hat{B}^F, 1) - u(\hat{B}^F + \hat{C}^F 1)] + \hat{\lambda}^F \hat{C}^F\}$$

$$< \theta \hat{C}^F \{\hat{\lambda}^F - u_x(\hat{B}^F + \hat{C}^F, 1)\} \leq 0 \qquad (39)$$

Turning to the effect of Type II errors on social welfare, we have:

$$\frac{\partial V^F}{\partial p_{II}} = (1-\theta)F(\overline{\overline{\omega}})E \ ; \ E = u(\hat{B}^F + \hat{C}^F, 1) - \int_0^{\overline{\overline{\omega}}} v(\omega, \hat{B}^F)g(\omega)d\omega - \hat{\lambda}^F \hat{C}^F \qquad (40)$$

Adding and subtracting $u(\hat{B}^F, 1)$ gives:

$$E = [u(\hat{B}^F + \hat{C}^F, 1) - u(\hat{B}^F, 1) - \hat{\lambda}^F \hat{C}^F] - \int_0^{\overline{\overline{\omega}}} [v(\omega, \hat{B}^F) - u(\hat{B}^F, 1)]g(\omega)d\omega \qquad (41)$$

The first term in square braces is the benefit to an individual – net of exchequer costs – of receiving the categorical benefit conditional on facing a zero quantity constraint on labour supply. The second term, meanwhile, measures the cost of facing a zero quantity constraint on labour conditional on receiving only the universal benefit. Now, given that $v(\omega, M) = u(M, 1) \, \forall \omega \le \overline{\omega}(M)$, the cost of facing the zero quantity constraint and only receiving B must be zero for all those with $\omega \le \overline{\omega}(B)$. Accordingly, we can rewrite (41) as:

$$E = \left[u\left(\hat{B}^F + \hat{C}^F, 1\right) - u\left(\hat{B}^F, 1\right) - \hat{\lambda}^F \hat{C}^F\right]$$

$$- \frac{F(\overline{\overline{\omega}}) - F[\overline{\omega}(B)]}{F(\overline{\overline{\omega}})} \int_{\overline{\omega}(B)}^{\overline{\overline{\omega}}(B,C)} \left[v\left(\omega, \hat{B}^F\right) - u\left(\hat{B}^F, 1\right)\right] g^{\infty}(\omega) d\omega \qquad (42)$$

Where $g^{\infty}(\omega) \equiv f(\omega) / \left[F(\overline{\overline{\omega}}) - F[\overline{\omega}(B)]\right]$ and $\int_{\overline{\omega}(B)}^{\overline{\overline{\omega}}} g^{\infty}(\omega) d\omega = 1$. But since $v(\omega, B) < v(\overline{\overline{\omega}}, B) = u(B + C, 1) \, \forall \, \overline{\omega}(B) \le \omega < \overline{\overline{\omega}}$ it must hold that:

$$E > \left(1 - \frac{F(\overline{\overline{\omega}}) - F[\overline{\omega}(B)]}{F(\overline{\overline{\omega}})}\right) \left[u\left(\hat{B}^F + \hat{C}^F, 1\right) - u\left(\hat{B}^F, 1\right)\right] - \hat{\lambda}^F \hat{C}^F$$

$$= \frac{F[\overline{\omega}(B)]}{F(\overline{\overline{\omega}})} \left[u\left(\hat{B}^F + \hat{C}^F, 1\right) - u\left(\hat{B}^F, 1\right)\right] - \hat{\lambda}^F \hat{C}^F \qquad (43)$$

And thus:

$$E > E' = \frac{F(\overline{\omega})}{F(\overline{\overline{\omega}})} \left[u\left(\hat{B}^F + \hat{C}^F, 1\right) - u\left(\hat{B}^F, 1\right)\right] - \hat{\lambda}^F \hat{C}^F \qquad (44)$$

By a standard property of concave functions[26] we can write:

$$E' > \hat{C}^F \left\{\frac{F[\overline{\omega}(B)]}{F(\overline{\overline{\omega}})} u_x\left(\hat{B}^F + \hat{C}^F, 1\right) - \hat{\lambda}^F\right\}$$

$$= \hat{C}^F \left\{\frac{F[\overline{\omega}(B)]}{F(\overline{\overline{\omega}})} \hat{\lambda}^F \left[1 + \frac{(1 - \theta) p_{II} f(\overline{\overline{\omega}}) \overline{\overline{\omega}}_C \hat{C}^F}{k^F}\right] - \hat{\lambda}^F\right\} \qquad (45)$$

$$= \hat{\lambda}^F \hat{C}^F \left\{\frac{F[\overline{\omega}(B)]}{F(\overline{\overline{\omega}})} \frac{(1 - \theta) p_{II} f(\overline{\overline{\omega}}) \overline{\overline{\omega}}_C \hat{C}^F}{k^F} - \left[1 - \frac{F[\overline{\omega}(B)]}{F(\overline{\overline{\omega}})}\right]\right\}$$

We thus have:

$$E > E' > \hat{\lambda}^F \hat{C}^F E'' \qquad (46)$$

Where:

---

[26] $u_x(B + C, 1)C < u(B + C, 1) - u(B, 1) < u_x(B, 1)$ (see (A5) in the Appendix)

$$E'' = \frac{F[\overline{\omega}(B)]}{F(\overline{\overline{\omega}})} \frac{(1-\theta)p_{II}f(\overline{\overline{\omega}})\overline{\overline{\omega}}_C \hat{C}^F}{k^F} - \left[1 - \frac{F[\overline{\omega}(B)]}{F(\overline{\overline{\omega}})}\right] \tag{47}$$

Substituting in the definition of k^F gives:

$$E'' = \frac{F[\overline{\omega}(B)]}{F(\overline{\overline{\omega}})} \frac{(1-\theta)p_{II}f(\overline{\overline{\omega}})\overline{\overline{\omega}}_C \hat{C}^F}{\theta(1-p_I) + (1-\theta)p_{II}F(\overline{\overline{\omega}})} - \left[1 - \frac{F[\overline{\omega}(B)]}{F(\overline{\overline{\omega}})}\right]$$

$$= \frac{F[\overline{\omega}(B)]}{F(\overline{\overline{\omega}})} \cdot \frac{f(\overline{\overline{\omega}})\overline{\overline{\omega}}_C \hat{C}^F}{\dfrac{\theta(1-p_I)}{(1-\theta)p_{II}} + F(\overline{\overline{\omega}})} - \left[1 - \frac{F[\overline{\omega}(B)]}{F(\overline{\overline{\omega}})}\right] \tag{48}$$

Finally, writing this in terms of elasticities we arrive at:

$$E'' = \frac{F[\overline{\omega}(B)]}{\dfrac{\theta(1-p_I)}{(1-\theta)p_{II}} + F(\overline{\overline{\omega}})} \cdot \underbrace{\frac{\overline{\overline{\omega}}f(\overline{\overline{\omega}})}{F(\overline{\overline{\omega}})}}_{\substack{elasticity\ of \\ F\ wrt\ \overline{\overline{\omega}}}} \cdot \underbrace{\frac{\hat{C}^F \overline{\overline{\omega}}_C}{\overline{\overline{\omega}}}}_{\substack{elasticity\ of \\ \overline{\overline{\omega}}\ wrt\ C}} - \left[1 - \frac{F[\overline{\omega}(B)]}{F(\overline{\overline{\omega}})}\right] \tag{49}$$

If $(1-\theta)p_{II} \approx 0$ then the first term on the right side of (49) is approximately zero such that $E'' < 0$. Yet, given the inequalities in (44) and (46), this is not sufficient to sign $\partial V^F / \partial p_{II}$. However, if $\theta(1-p_I) \approx 0$ then:

$$E'' \approx \frac{F(\overline{\omega})}{F(\overline{\overline{\omega}})} \cdot \frac{\overline{\overline{\omega}}f(\overline{\overline{\omega}})}{F(\overline{\overline{\omega}})} \cdot \frac{\hat{C}^F \overline{\overline{\omega}}_C}{\overline{\overline{\omega}}} - \left[1 - \frac{F(\overline{\omega})}{F(\overline{\overline{\omega}})}\right]$$

$$= \frac{F(\overline{\omega})}{F(\overline{\overline{\omega}})} \cdot \left\{\frac{\overline{\overline{\omega}}f(\overline{\overline{\omega}})}{F(\overline{\overline{\omega}})} \cdot \frac{\hat{C}^F \overline{\overline{\omega}}_C}{\overline{\overline{\omega}}} - \left[\frac{F(\overline{\overline{\omega}})}{F(\overline{\omega})} - 1\right]\right\}$$

If the product of (i) the elasticity of F with respect to ω - evaluated at $\overline{\overline{\omega}}$ - and (ii) the elasticity of $\overline{\overline{\omega}}$ with respect to C - evaluated at $\hat{C}^F$ – are sufficiently high, then we can have $E'' > 0$ and thus $\partial V / \partial p_{II} > 0$. Since these elasticities depend on as yet unspecified properties of the distribution and utility functions respectively, we have enough degrees of freedom to choose parameters so that $E'' > 0$.

3.2.2.1. *A Numerical Example where Welfare is Increasing in Type II Errors.*

We now provide an example under which social welfare is increasing in Type II errors. Let preferences take the Cobb-Douglas form:

$$u(x,l) = x^\alpha l^{1-\alpha} \; ; 0 < \alpha < 1$$

We suppose that $\alpha = 0.5$. Taking net wages as exponentially distributed with shape parameter $\lambda = 2$, the average net wage in the economy is 0.5. We take the fixed benefit budget to be $\beta = 0.05$, thus corresponding to 10 percent of the average wage. Under these parameters, note that $\bar{\beta}^P = 0.008$ such that the assumed budget size would, under perfect discrimination, guarantee that (i) $\hat{C}^P > 0$ eliminates between-group inequality

in the average social marginal utility and (ii) $\hat{B}^P > 0$. Table 1 displays No Enforcement and Full Enforcement Optima for different p_I and p_{II}. The columns $V^N(p_I, p_{II})/V^N(p_I, 0)$ and $V^F(p_I, p_{II})/V^F(p_I, 0)$ capture, for a given level of p_I, whether social welfare is increasing or decreasing in p_{II}. In both cases the welfare changes are small, but, importantly, are of different sign[27]. Under No Enforcement social welfare is always decreasing in p_{II}, whilst under Full Enforcement is can be increasing in p_{II}. The size of the welfare changes may be attributed to the fixed budget assumption – such that errors have no effect on the budget size – and also the fact that there is no express concern for inequality in the social welfare function beyond the concavity of individual utility (see Stern, 1976, p.153).

---

[27] Extensive simulations for different values of θ and λ show the same results. These are available from the authors upon request.

Table 1: Example where Welfare is Increasing in Type II Errors under Full Enforcement

p_{II}	No Enforcement			Full Enforcement		
	$\dfrac{V^N(p_I,p_{II})}{V^N(0,p_{II})}$	$\widehat{B}^N$	$\widehat{C}^N$	$\dfrac{V^F(p_I,p_{II})}{V^N(0,p_{II})}$	$\widehat{B}^F$	$\widehat{C}^F$
	$p_I = 0$			$p_I = 0$		
0	1.000	0.041	0.184	1.000	0.041	0.184
0.2	0.995	0.033	0.069	0.999	0.037	0.085
0.4	0.993	0.030	0.046	1.002	0.031	0.078
0.6	0.992	0.028	0.035	1.005	0.022	0.081
0.8	0.991	0.027	0.028	1.009	0.012	0.085
	$p_I = 0.2$			$p_I = 0.2$		
0	1.000	0.043	0.178	1.000	0.043	0.178
0.2	0.995	0.038	0.053	1.000	0.039	0.076
0.4	0.994	0.038	0.030	1.002	0.034	0.070
0.6	0.993	0.040	0.016	1.005	0.028	0.070
0.8	0.993	0.050	0.000	1.008	0.018	0.077
	$p_I = 0.4$			$p_I = 0.4$		
0	1.000	0.045	0.173	1.000	0.045	0.173
0.2	0.996	0.042	0.038	1.000	0.042	0.068
0.4	0.995	0.044	0.015	1.002	0.037	0.063
0.6	0.995	0.050	0.000	1.005	0.031	0.065
0.8	0.995	0.050	0.000	1.008	0.023	0.069

4. Concluding Remarks

Real world targeted benefits typically feature what we term *double conditionality*. *Ex-ante* the award of the benefit is conditioned on an individual meeting some eligibility conditions, whilst *ex-post* requirements are placed on benefit recipients. The literature on the targeting of benefits with classification errors has focused on the optimal provision of benefits under particular enforcement structures in relation to *ex-post* conditionality, but there has been no comparison within a single framework across enforcement regimes. Further, the affect of classification errors on social welfare under differing enforcement structures has been given little attention.

In relation to these gaps in the literature, we posed three questions at the outset. First, how does the propensity to make classification errors of Type I and Type II affect the choice between (a) a pure universal benefit system, or (b) a targeted system which may take the form of a purely targeted system or a partially targeted system with a universal benefit? Second, how the propensities to make classification errors affect maximum social welfare. Third, how do the answers both of these questions depend on how well the *ex-post* conditionality is enforced?

In a framework where (i) individuals differ in both their ability to work – modelled as an exogenous zero quantity constraint on labour supply – and, conditional on being able to work, their productivity at work, and (ii) the government chooses the optimal levels of (a) a universal benefit and (b) a categorical benefit that is targeted at the unable but administered with classification errors and a no work requirement that may be either fully enforced or not enforced at all, we have analysed each of the above questions.

The principal messages are as follows:

- Under a No Enforcement regime:

(i)   It is optimal to adopt a pure universal system only if the test administering the categorical benefit has no discriminatory power. In all other cases it is optimal to adopt a targeted system. If Type I errors are made this will always be a partial universal system to avoid some unable individuals having no source of income to consume. However, if no Type I errors are made and the benefit budget is sufficiently small, it is optimal to adopt a purely targeted system.

(ii)   Maximum social welfare is decreasing in the propensity to make both Type I and Type II classification errors. This implies that while society will benefit from improvements in the accuracy of the test, changes that alter the liability threshold – i.e. make it tougher or weaker - will involve a trade-off between Type I and Type II errors.

- Under a Full Enforcement regime:

(i)   It is optimal to adopt a targeted system for all levels of discriminatory power. A pure universal system is *never* chosen. Once more, whether a purely targeted or partially targeted system is chosen will depend on whether or not Type I errors are made and on the size of the benefit budget

(ii)   Whilst maximum welfare is unambiguously decreasing in Type I errors, there are conditions under which it can be *increasing* in Type II errors. This implies that changes that increase the liability threshold – make the test tougher – will increase Type I errors but reduce Type II errors and so are unambiguously welfare decreasing. It follows that changes which lower the liability threshold – i.e. make the test less tough – are welfare increasing. This supports the suggestion by Goodin (1985) to "err on the side of kindness".

There are a number of important research directions that emerge from this framework, but go beyond the scope of this paper. The first concerns how the optimal universal and categorical benefit levels change with Type I and Type II classification error propensities. Insights into this rely on numerical simulations and this is an on going research project[28]. Second, how does the optimal tax rate (i) change with classification errors and (ii) depend on the enforcement structure in place? We note that whilst the paper has assumed a fixed benefit budget, the conditions under which it is optimal to adopt a purely universal or targeted system continue hold when the benefit budget is determined by the tax revenue collected (Slack & Ulph 2014). Finally, a third interesting direction concerns how the analysis changes when there are differing degrees of disability, or quantity constraint.

---

[28] Numerical simulation results are available at request from the authors.

Figure 1: The Functions $\bar{\omega}(M)$ and $\bar{\bar{\omega}}(B,C)$
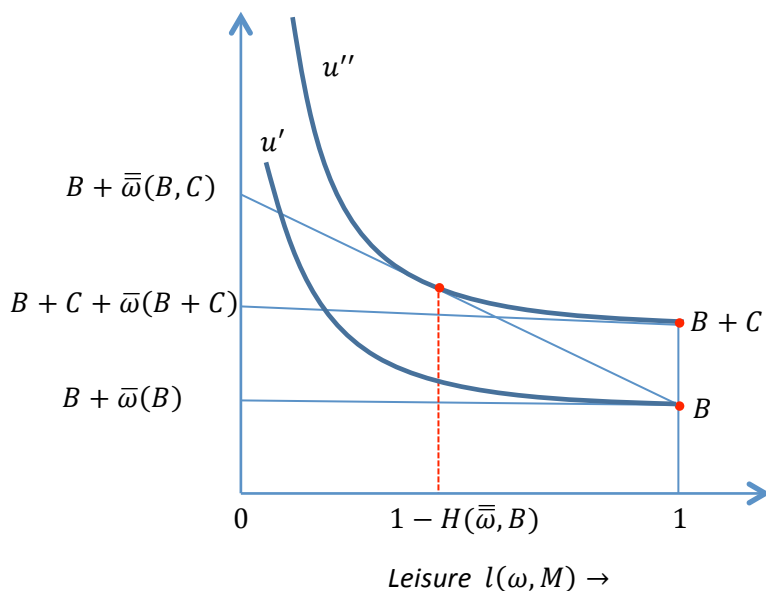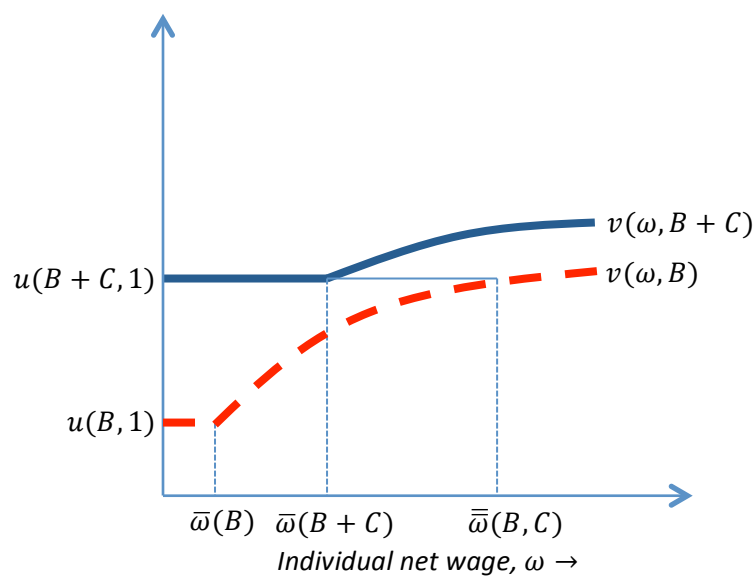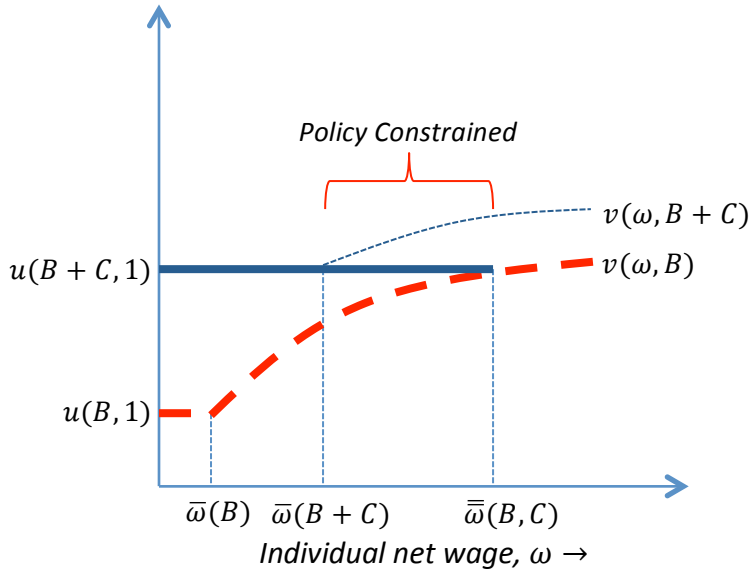

Figure 2: Individual Utility under Alternative Enforcement Structures

(i)   No Enforcement

NB. The bold unbroken curve depicts utility with benefit income $B + C$, whilst the bold broken curve depicts utility with benefit income B.

(ii)   Full Enforcement

Policy Constrained

$v(\omega, B + C)$

$v(\omega, B)$

$u(B + C, 1)$

$u(B, 1)$

$\bar{\omega}(B)$ $\bar{\omega}(B + C)$ $\bar{\bar{\omega}}(B, C)$

Individual net wage, $\omega \rightarrow$

NB. The bold unbroken curve depicts utility with benefit income $B + C$, whilst the bold broken curve depicts utility with benefit income B.

Appendix B: Proofs

Proof of Proposition 1

The FOCs for characterising $\hat{B}^P$ and $\hat{C}^P$ are:

$$
\theta u_x\big(\hat{B}^P + \hat{C}^P, 1\big) + (1 - \theta) \int_0^\infty v_M\big(\omega, \hat{B}^P\big) f(\omega) d\omega \le \hat{\lambda}^P, \hat{B}^P \ge 0
$$
$$
u_x\big(\hat{B}^P + \hat{C}^P, 1\big) \le \hat{\lambda}^P, \hat{C}^P \ge 0 \tag{A1}
$$
$$
\hat{B}^P + \theta \hat{C}^P = \beta
$$

Here the inequalities hold with complementary slackness and λ is the shadow price of public expenditure.

Given that the budget constraint must be exhausted (i.e. $\beta + \theta C = \beta$), we now test two hypotheses. First, suppose that $\hat{B}^P = \beta$ and $\hat{C}^P = 0$. In this case, the FOCs in (A1) imply that $\hat{C}^P = 0$ only if $\int_0^\infty v_M(\omega, \beta) f(\omega) d\omega > u_x(\beta, 1)$, which contradicts (3). It must therefore hold that $\hat{C}^P > 0$. Second, suppose that $\hat{B}^P = 0$ and thus $\hat{C}^P = \beta/\theta$. The FOCs in (A1) now imply that $\hat{B}^P = 0$ only if $\int_0^\infty v_M(\omega, 0) f(\omega) d\omega \le u_x(\beta/\theta, 1)$. *Ceteris paribus*, the right side is decreasing in β and so there is a critical budget level $\bar{\beta}^P$ satisfying (12). Q.E.D.

Proof of Proposition 2a

The FOCs characterising $\hat{B}^N$ and $\hat{C}^N$ are:

$$\theta\{(1 - p_I)u_x(\hat{B}^N + \hat{C}^N, 1) + p_I u_x(\hat{B}^N, 1)\} +$$

$$(1 - \theta)\left\{\begin{array}{l} p_{II}\displaystyle\int_0^\infty [v_M(\omega, \hat{B}^N + \hat{C}^N) - v_M(\omega, \hat{B}^N)]f(\omega)d\omega \\[2mm] + \displaystyle\int_0^\infty v_M(\omega, \hat{B}^N)f(\omega)d\omega \end{array}\right\} \leq \hat{\lambda}^N, \hat{B}^N \geq 0; \tag{A2}$$

$$\frac{1}{k^N}\left\{\begin{array}{l} \theta(1 - p_I)u_x(\hat{B}^N + \hat{C}^N, 1) \\[2mm] +(1 - \theta)p_{II}\displaystyle\int_0^\infty v_M(\omega, \hat{B}^N + \hat{C}^N)f(\omega)d\omega \end{array}\right\} \leq \hat{\lambda}^N, \hat{C}^N \geq 0 \tag{A3}$$

$$\hat{B}^N + k\hat{C}^N = \beta \tag{A4}$$

where the inequalities hold with complementary slackness.

With regard to the above FOCs, we test the following two hypotheses:

(i) $\quad \hat{B}^N = \beta, \hat{C}^N = 0$

The FOCs in (A2) and (A3) reduce to:

$$\theta u_x(\beta, 1) + (1 - \theta)\int_0^\infty v_M(\omega, \beta)f(\omega)d\omega = \lambda$$

$$\left[\theta(1 - p_I)u_x(\beta, 1) + (1 - \theta)p_{II}\int_0^\infty v_M(\omega, \beta)f(\omega)d\omega\right]/k^N \leq \lambda$$

Combining both equations gives:

$$(1 - p_I)\left[\lambda - (1 - \theta)\int_0^\infty v_M(\omega, \beta)f(\omega)d\omega\right] + (1 - \theta)p_{II}\int_0^\infty v_M(\omega, \beta)f(\omega)d\omega \leq \lambda k^N$$

$$\Leftrightarrow (1 - p_I)(1 - \theta)\left[\lambda - \int_0^\infty v_M(\omega, \beta)f(\omega)\,d\omega\right] \leq p_{II}(1 - \theta)\left[\lambda - \int_0^\infty v_M(\omega, \beta)f(\omega)d\omega\right]$$

$$\Leftrightarrow 1 - p_I \leq p_{II}$$

Given our discriminatory power assumption (i.e. $p_I + p_{II} \leq 1$), this condition can only hold with equality, and thus when $p_I + p_{II} = 1$. So $\hat{C}^P = 0$ only if there is no discriminatory power.

(ii) $\hat{B}^N = 0, \hat{C}^N = \beta/k^N$

First off, by (1) we have $\lim_{x \to 0} u_x(0,1) = +\infty$ such that the FOCs above cannot hold and $\hat{B}^N > 0$ whenever $p_I > 0$. However, for $p_I = 0$, the FOCs in (A2) and (A3) now reduce to:

$$\theta u_x\left(\frac{\beta}{k^N}, 1\right) + (1 - \theta)\left\{\int_0^\infty \left[p_{II}v_M\left(\omega, \frac{\beta}{k^N}\right) + (1 - p_{II})v_M(\omega, 0)\right]f(\omega)d\omega\right\} \leq \lambda,$$

$$\left[\theta u_x\left(\frac{\beta}{k^N},1\right)+(1-\theta)p_{II}\int_0^\infty v_M\left(\omega,\frac{\beta}{k^N}\right)f(\omega)d\omega\right]/k^N=\lambda$$

Combining these two equations gives:

$$(1-\theta)(1-p_{II})\cdot[\theta+(1-\theta)p_{II}]\int_0^\infty v_M(\omega,0)\,f(\omega)d\omega$$

$$\leq[1-\theta-(1-\theta)p_{II}]\left\{\theta u_x\left(\frac{\beta}{k^N},1\right)+(1-\theta)p_{II}\int_0^\infty v_M\left(\omega,\frac{\beta}{k^N}\right)f(\omega)d\omega\right\}$$

Rearranging gives:

$$[\theta(1-\theta)(1-p_{II})+(1-\theta)^2p_{II}(1-p_{II})]\int_0^\infty v_M(\omega,0)\,f(\omega)d\omega$$

$$\leq\theta(1-\theta)(1-p_{II})u_x\left(\frac{\beta}{k^N},1\right)+(1-\theta)^2p_{II}(1-p_{II})\int_0^\infty v_M\left(\omega,\frac{\beta}{k^N}\right)f(\omega)d\omega$$

The left side is independent of β, whilst the right side is unambiguously decreasing in β. Suppose that $\beta\to0$, then $\lim_{\beta\to0}u_x\left(\frac{\beta}{k^N},1\right)=+\infty$ such that the right side approaches $+\infty$ and the condition must hold with strict inequality. There must therefore be a critical budget level $\bar\beta^N$ for which the condition holds with equality. Using the definition of $\bar\beta^P$ from (12), we can define $\bar\beta^N$ by:

$$[\theta(1-\theta)(1-p_{II})+(1-\theta)^2p_{II}(1-p_{II})]u_x\left(\frac{\bar\beta^P}{\theta},1\right)$$

$$\equiv\theta(1-\theta)(1-p_{II})u_x\left(\frac{\bar\beta^N}{k^N},1\right)+(1-\theta)^2p_{II}(1-p_{II})\int_0^\infty v_M\left(\omega,\frac{\bar\beta^N}{k^N}\right)f(\omega)d\omega$$

Dividing by $(1-\theta)(1-p_{II})$ gives:

$$k^N u_x\left(\frac{\bar\beta^P}{\theta},1\right)\equiv\theta u_x\left(\frac{\bar\beta^N}{k^N},1\right)+(1-\theta)p_{II}\int_0^\infty v_M\left(\omega,\frac{\bar\beta^N}{k^N}\right)f(\omega)d\omega$$

It is straightforward to see that this implies $u_x(\bar\beta^N/k^N,1)>u_x(\bar\beta^P/\theta,1)\Rightarrow\bar\beta^N<\bar\beta^P\cdot k^N/\theta$. Note that $p_{II}=0\Rightarrow k^N=\theta\Rightarrow\bar\beta^P=\bar\beta^N$. To see how $\bar\beta^N$ behaves with p_{II} we differentiate the above expression with respect to p_{II} to obtain:

$$(1-\theta)\left\{u_x\left(\frac{\bar\beta^P}{\theta},1\right)-\int_0^\infty v_M\left(\omega,\frac{\bar\beta^N}{k^N}\right)f(\omega)d\omega\right\}$$

$$=\left[\frac{k^N\frac{\partial\bar\beta^N}{\partial p_{II}}-(1-\theta)\bar\beta^N}{(k^N)^2}\right]\cdot\left[\theta u_{xx}\left(\frac{\bar\beta^N}{k^N},1\right)+(1-\theta)p_{II}\int_0^\infty v_{MM}\left(\omega,\frac{\bar\beta^N}{k^N}\right)\right]f(\omega)d\omega$$

Given that $\int_0^\infty v_M(\omega,0)f(\omega)d\omega=u_x(\bar\beta^P/\theta,1)$, it must hold that $u_x(\bar\beta^P/\theta,1)>\int_0^\infty v_M(\omega,\bar\beta^N/k^N)f(\omega)d\omega$. Solving for $\partial\bar\beta^N/\partial p_{II}$ therefore gives:

$$\frac{\partial\bar\beta^N}{\partial p_{II}}=(1-\theta)k^N\left[\frac{\left(u_x\left(\frac{\bar\beta^P}{\theta},1\right)-\int_0^\infty v_M\left(\omega,\frac{\bar\beta^N}{k^N}\right)f(\omega)d\omega\right)}{\theta u_{xx}\left(\frac{\bar\beta^N}{k^N},1\right)+(1-\theta)(1-p_{II})\int_0^\infty v_{MM}\left(\omega,\frac{\bar\beta^N}{k^N}\right)}\right]+(1-\theta)\frac{\bar\beta^N}{k^N}$$

Q.E.D.
Proof of Proposition 2b

For $p_I + p_{II} < 1$, the FOCs in (A2) and (A3) can be wrote as:

$$\theta p_I u_x(\hat{B}^N, 1) + (1 - \theta)(1 - p_{II}) \int_0^\infty v_M(\omega, \hat{B}^N) f(\omega) d\omega \leq \hat{\lambda}^N (1 - k^N); \; \hat{B}^N \geq 0,$$

$$\theta(1 - p_I) u_x(\hat{B}^N + \hat{C}^N, 1) + (1 - \theta) p_{II} \int_0^\infty v_M(\omega, \hat{B}^N + \hat{C}^N) f(\omega) d\omega = \hat{\lambda}^N k^N$$

From these conditions it is straightforward to establish (24). To establish the effect of classification errors on social welfare, we make use of the following standard property of concave functions:

$$u_x(B, 1) \cdot C > u(B + C, 1) - u(B, 1) > u_x(B + C, 1) \cdot C \tag{A5}$$

From (24) and the above property, it must therefore hold that:

$$\frac{\partial V^N}{\partial p_I} = \theta \{[u(\hat{B}^N, 1) - u(\hat{B}^N + \hat{C}^N, 1)] + \hat{\lambda}^N \hat{C}^N\} < \theta \hat{C}^N [\hat{\lambda}^N - u_x(\hat{B}^N + \hat{C}^N, 1)] \leq 0$$

$$\frac{\partial V^N}{\partial p_{II}} = (1 - \theta) \left\{ \int_0^\infty [v(\omega, \hat{B}^N + \hat{C}^N) - v(\omega, \hat{B}^N)] f(\omega) d\omega - \hat{\lambda}^N \hat{C}^N \right\}$$

$$< \theta \hat{C}^N \left[\int_0^\infty v_M(\omega, \hat{B}^N) f(\omega) d\omega - \hat{\lambda}^N \right] \leq 0$$

Q.E.D.

Proof of Proposition 3a

The welfare function in (28) can be written more explicitly as:

$$W^F(B, C; p_I, p_{II}, \theta) = [\theta(1 - p_I) + (1 - \theta) p_{II} F(\overline{\overline{\omega}})] u(B + C, 1)$$

$$+ [\theta p_I + (1 - \theta)(1 - p_{II}) F(\overline{\omega})] u(B, 1)$$

$$+ (1 - \theta) \left[(1 - p_{II}) \int_{\underline{\omega}}^{\overline{\overline{\omega}}} v(\omega, B) f(\omega) d\omega + \int_{\overline{\overline{\omega}}}^\infty v(\omega, B) f(\omega) d\omega \right]$$

Given the benefit budget constraint $B + [\theta(1 - p_I) + (1 - \theta) p_{II} F(\overline{\overline{\omega}})] C = \beta$, the FOCs are:

FOC(B):

$$[\theta(1 - p_I) + (1 - \theta) p_{II} F(\overline{\overline{\omega}})] u_x(\hat{B}^F + \hat{C}^F, 1) + (1 - \theta) p_{II} f(\overline{\overline{\omega}}) \overline{\overline{\omega}}_B u(B + C, 1)$$
$$+ [\theta p_I + (1 - \theta)(1 - p_{II}) F(\overline{\omega}) u_x(\hat{B}^F, 1) + (1 - \theta)(1 - p_{II}) f(\overline{\omega}) \overline{\omega}_B u(B, 1)$$
$$+ (1 - \theta) \left[(1 - p_{II}) \int_{\underline{\omega}}^{\overline{\overline{\omega}}} v_M(\omega, \hat{B}^F) f(\omega) d\omega + \int_{\overline{\overline{\omega}}}^\infty v_M(\omega, \hat{B}^F) f(\omega) d\omega \right]$$
$$+ (1 - \theta) \{(1 - p_{II})[\overline{\overline{\omega}}_B v(\overline{\overline{\omega}}, B) f(\overline{\overline{\omega}}) - \overline{\omega}_B v(\overline{\omega}, B) f(\overline{\omega})] - \overline{\overline{\omega}}_B v(\overline{\overline{\omega}}, B) f(\overline{\overline{\omega}})\}$$
$$\leq \hat{\lambda}^F [1 + (1 - \theta) p_{II} f(\overline{\overline{\omega}}) \overline{\overline{\omega}}_B \hat{C}^F], \hat{B}^F \geq 0$$

Using the identities $v(\overline{\omega}, B) \equiv u(B, 1)$ and $v(\overline{\overline{\omega}}, B) \equiv u(B + C, 1)$, this simplifies to:

$\Leftrightarrow$

$$[\theta(1 - p_I) + (1 - \theta)p_{II}F(\overline{\overline{\omega}})]u_x(\hat{B}^F + \hat{C}^F, 1)$$
$$+[\theta p_I + (1 - \theta)(1 - p_{II})F(\overline{\omega})]u_x(\hat{B}^F, 1)$$
$$+(1 - \theta)\left[(1 - p_{II})\int_{\overline{\omega}}^{\overline{\overline{\omega}}} v_M(\omega, \hat{B}^F)f(\omega)d\omega + \int_{\overline{\overline{\omega}}}^{\infty} v_M(\omega, \hat{B}^F)f(\omega)d\omega\right] \quad \text{(A6)}$$
$$\leq \hat{\lambda}^F[1 + (1 - \theta)p_{II}f(\overline{\overline{\omega}})\overline{\overline{\omega}}_B\hat{C}^F], \hat{B}^F \geq 0$$

FOC(C):

$$[\theta(1 - p_I) + (1 - \theta)p_{II}F(\overline{\overline{\omega}})]u_x(\hat{B}^F + \hat{C}^F, 1) + (1 - \theta)p_{II}f(\overline{\overline{\omega}})\overline{\overline{\omega}}_C u(\hat{B}^F + \hat{C}^F, 1)$$
$$+(1 - \theta)\{(1 - p_{II})\overline{\overline{\omega}}_C v(\overline{\overline{\omega}}, B)f(\overline{\overline{\omega}}) - \overline{\overline{\omega}}_C v(\overline{\overline{\omega}}, B)f(\overline{\overline{\omega}})\}$$
$$\leq \hat{\lambda}^F\{[\theta(1 - p_I) + (1 - \theta)p_{II}F(\overline{\overline{\omega}})] + (1 - \theta)p_{II}f(\overline{\overline{\omega}})\overline{\overline{\omega}}_C\hat{C}^F\}, \hat{C}^F \geq 0$$

$\Leftrightarrow$

$$[\theta(1 - p_I) + (1 - \theta)p_{II}F(\overline{\overline{\omega}})]u_x(\hat{B}^F + \hat{C}^F, 1)$$
$$\leq \hat{\lambda}^F\{\theta(1 - p_I) + (1 - \theta)p_{II}[F(\overline{\overline{\omega}}) + f(\overline{\overline{\omega}})\overline{\overline{\omega}}_C\hat{C}^F]\}, \hat{C}^F \geq 0 \quad \text{(A7)}$$

From the definitions of k^F, α, χ, g and h, (A6) and (A7) can therefore be written as:

$$k^F u_x(\hat{B}^F + \hat{C}^F, 1) + (1 - k^F)\left[\begin{array}{l} \alpha u_x(\hat{B}^F, 1) + \chi \int_0^{\overline{\overline{\omega}}} v_M(\omega, \hat{B}^F)g(\omega)d\omega + \\ \\ (1 - \alpha - \chi)\int_{\overline{\overline{\omega}}}^{\infty} v_M(\omega, \hat{B}^F)h(\omega)d\omega \end{array}\right] \quad \text{(A8)}$$
$$\leq \hat{\lambda}^F[1 + (1 - \theta)p_{II}f(\overline{\overline{\omega}})\overline{\overline{\omega}}_B\hat{C}^F]; \quad \hat{B}^F \geq 0$$

$$k^F u_x(\hat{B}^F + \hat{C}^F, 1) \leq \hat{\lambda}^F\{\theta(1 - p_I) + (1 - \theta)p_{II}[F(\overline{\overline{\omega}}) + f(\overline{\overline{\omega}})\overline{\overline{\omega}}_C\hat{C}^F]\}; \quad \hat{C}^F \geq 0 \quad \text{(A9)}$$

where the pairs of inequalities hold with complementary slackness.

We test the following two hypotheses:

(i) $\quad \hat{B}^F = \beta, \hat{C}^F = 0$

Suppose that $\hat{C}^F = 0$ - and thus $\hat{B}^F = \beta$ – such that $\overline{\overline{\omega}}(\beta, 0) = \overline{\omega}(\beta)$. It follows $\forall \omega \leq \overline{\overline{\omega}}(\beta, 0) = \overline{\omega}(\beta): H(\omega, \beta) = 0 \Rightarrow u(\beta, 1) = v(\omega, \beta) \Rightarrow v_M(\omega, \beta) = u_x(\beta, 1)$. The FOCs (A6) and (A7) (or, equivalently (A8) and (A9)) therefore become:

$$[\theta + (1 - \theta)F(\overline{\omega})]u_x(\beta, 1) + (1 - \theta)\int_{\overline{\omega}}^{\infty} v_M(\omega, \beta)f(\omega)d\omega = \hat{\lambda}^F,$$
$$u_x(\beta, 1) \leq \hat{\lambda}^F.$$

Combining these equations implies the contradictory statement:

$$\frac{1}{[1 - F(\overline{\omega})]}\int_{\overline{\omega}}^{\infty} v_M(\omega, \beta)f(\omega)d\omega \geq u_x(\beta, 1)$$

The assertion that $\hat{C}^F > 0$ is therefore false and we instead have $\hat{C}^F > 0 \; \forall \; p_I + p_{II} \leq 1$.

(ii) $\quad \hat{B}^F = 0, \hat{C}^F = \beta/k^F$

If $p_I > 0$ then this assertion must be false by (3). However, if $p_I = 0$ then the FOCs become:

$$[\theta + (1-\theta)p_{II}F(\overline{\overline{\omega}})]u_x\left(\frac{\beta}{k^F}, 1\right)$$

$$+(1-\theta)\left[(1-p_{II})\int_0^{\overline{\overline{\omega}}} v_M(\omega, 0)f(\omega)d\omega + \int_{\overline{\overline{\omega}}}^\infty v_M(\omega, 0)f(\omega)d\omega\right]$$

$$\leq \hat{\lambda}^F\left[1 + (1-\theta)p_{II}f(\overline{\overline{\omega}})\overline{\overline{\omega}}_B\left(\frac{\beta}{k^F}\right)\right]$$

and:

$$[\theta + (1-\theta)p_{II}F(\overline{\overline{\omega}})]u_x\left(\frac{\beta}{k^F}, 1\right) = \hat{\lambda}^F\left\{\theta + (1-\theta)p_{II}\left[F(\overline{\overline{\omega}}) + f(\overline{\overline{\omega}})\overline{\overline{\omega}}_C\left(\frac{\beta}{k^F}\right)\right]\right\}$$

Let $A \equiv (1-\theta)\left[(1-p_{II})\int_0^{\overline{\overline{\omega}}} v_M(\omega, 0)f(\omega)d\omega + \int_{\overline{\overline{\omega}}}^\infty v_M(\omega, 0)f(\omega)d\omega\right]$. This term is independent of β and can be treated as a constant. Combining the two FOCs, we thus obtain:

$$[\theta + (1-\theta)p_{II}F(\overline{\overline{\omega}})]u_x\left(\frac{\beta}{k^F}, 1\right) + A$$

$$\leq \frac{[\theta + (1-\theta)p_{II}F(\overline{\overline{\omega}})]u_x\left(\frac{\beta}{k^F}, 1\right)}{\left\{\theta + (1-\theta)p_{II}\left[F(\overline{\overline{\omega}}) + f(\overline{\overline{\omega}})\overline{\overline{\omega}}_C\left(\frac{\beta}{k^F}\right)\right]\right\}} \cdot \left[1 + (1-\theta)p_{II}f(\overline{\overline{\omega}})\overline{\overline{\omega}}_B\left(\frac{\beta}{k^F}\right)\right]$$

Dividing through by $[\theta + (1-\theta)p_{II}F(\overline{\overline{\omega}})]u_x\left(\frac{\beta}{k^F}, 1\right)$ thus gives:

$$1 + \frac{A}{[\theta + (1-\theta)p_{II}F(\overline{\overline{\omega}})]u_x\left(\frac{\beta}{k^F}, 1\right)} \leq \frac{\left[1 + (1-\theta)p_{II}f(\overline{\overline{\omega}})\overline{\overline{\omega}}_B\left(\frac{\beta}{k^F}\right)\right]}{\left\{\theta + (1-\theta)p_{II}\left[F(\overline{\overline{\omega}}) + f(\overline{\overline{\omega}})\overline{\overline{\omega}}_C\left(\frac{\beta}{k^F}\right)\right]\right\}}$$

Suppose $\beta \to 0$. In this case $\lim_{\beta \to 0} u_x\left(\frac{\beta}{k^F}, 1\right) = +\infty$ such that the left side of the above expression approaches 1. Contrastingly, because the $\lim_{\beta \to 0}\left(\frac{\beta}{k^F}\right) = 0$, the right side of the above expression approaches $1/k^F > 1$. So for a benefit budget sufficiently low, this condition holds with strict inequality. There must therefore be a critical benefit budget, which we denote by $\bar{\beta}^F$ at which this condition holds with equality. Q.E.D.

Proof that $p_{II}^\infty < p_{II}$

From (34), $p_{II}^\infty < p_{II} \Leftrightarrow 1 < p_{II} + \left[1 - p_{II}F[\overline{\overline{\omega}}(B, C)]\right]/F[\overline{\omega}(B + C)] \Leftrightarrow p_{II}\left[F[\overline{\overline{\omega}}(B, C)] - F[\overline{\omega}(B)]\right] < 1 - F[\overline{\omega}(B + C)]$. Q.E.D.

Proof of (37)

Given that $\hat{C}^F > 0 \ \forall \ 0 \le p_I + p_{II} \le 1$, the FOC for characterising $\hat{C}^F$ must hold with equality to give:

$$u_x\big(\hat{B}^F + \hat{C}^F, 1\big) = \hat{\lambda}^F \left\{ 1 + \frac{(1-\theta)p_{II}f(\bar{\bar{\omega}})\bar{\bar{\omega}}_C \hat{C}^F}{k^F} \right\} > \hat{\lambda}^F$$

Substituting this into the FOC characterising $\hat{B}^F$ then gives[29]:

$$\begin{bmatrix} \alpha u_x\big(\hat{B}^F, 1\big) + \chi \displaystyle\int_0^{\bar{\bar{\omega}}} v_M\big(\omega, \hat{B}^F\big)g(\omega)d\omega + \\[2mm] (1 - \alpha - \chi) \displaystyle\int_{\bar{\bar{\omega}}}^{\infty} v_M\big(\omega, \hat{B}^F\big)h(\omega)d\omega \end{bmatrix} \le \hat{\lambda}^F \left[1 + \frac{(1-\theta)p_{II}f(\bar{\bar{\omega}})(\bar{\bar{\omega}}_B - \bar{\bar{\omega}}_C)\hat{C}^F}{1 - k^F} \right] < \hat{\lambda}^F$$

Given that $u_x\big(\hat{B}^F, 1\big) > \int_0^{\bar{\bar{\omega}}} v_M\big(\omega, \hat{B}^F\big)g(\omega)d\omega > u_x(\hat{B}^F + \hat{C}^F, 1)$, the only way in which the above condition can hold is if $\int_{\bar{\bar{\omega}}}^{\infty} v_M\big(\omega, \hat{B}^F\big)h(\omega)d\omega$. Result (37) follows. Q.E.D.

References

Akerlof, G. A. (1978). The Economics of "Tagging" as Applied to the Optimal Income Tax, Welfare Programs, and Manpower Planning. *American Economic Review*, *68*, 8–19.

Atkinson, A. B. (1996). *Public Economics in Action*. Oxford University Press.

Atkinson, A. B., & Sutherland, H. (1989). Analysis of a partial basic income scheme. In A. B. Atkinson (Ed.), *Poverty and Social Security*. Hemel Hempstead: Harvester Press.

Autor, D. H., & Duggan, M. G. (2006). The growth in the Social Security Disability rolls: a fiscal crisis unfolding. *The Journal of Economic Perspectives : A Journal of the American Economic Association*, *20*(3), 71–96.

Beath, J. A., Lewis, G. W., & Ulph, D. T. (1988). Policy targeting in a new welfare framework with poverty. In P. G. Hare (Ed.), *Surveys in Public Sector Economics* (pp. 161–185). Oxford, UK: Basil Blackwell.

Besley, T., & Coate, S. (1992). Workfare versus Welfare: Incentive Arguments for Work Requirements in Poverty-Alleviation Programs. *American Economic Review*, *82*(1), 249–261.

---

[29] This condition states that the government spends on C up until the point where the social marginal utility of income of a recipient of C equates with the shadow price of public expenditure multiplied by the marginal cost of increasing C. This marginal cost is composed of two effects. First, there is the direct cost of increasing the welfare payment to existing recipients of C. Second, a marginal increase in C induces an additional individual at the margin to apply for C. With probability p_{II} this additional applicant is awarded the benefit. So the marginal cost of an increase in C is composed of increased welfare payments to existing and new categorical benefit recipients.

Blackorby, C., & Donaldson, D. (1988). Cash Versus Kind, Self-selection, and Efficient Transfers. *The American Economic Review*, *78*, 691–700.

Bound, J. (1989). The health and earnings of rejected disability insurance applicants. *The American Economic Review*, *79*(3), 482–503.

Bound, J. (1991). The health and earnings of rejected disability insurance applicants: reply. *The American Economic Review*, *81*(5), 1427–1434.

Callan, T., O'Donoghue, C., Sutherland, H., & Wilson, M. (1999). *Comparative analysis of Basic Income proposals: UK and Ireland*. Cambridge: The Microsimulation Unit, Department of Applied Economics, University of Cambridge.

Chen, S., & van der Klaauw, W. (2008). The work disincentive effects of the disability insurance program in the 1990s. *Journal of Econometrics*, *142*(2), 757–784.

Christiansen, V. (1983). Some Important Properties of the Social Marginal Utility of Income. *Scandinavian Journal of Economics*, *85*(3), 359–371.

Cuff, K. (2000). Optimality of workfare with heterogeneous preferences. *Canadian Journal of Economics/Revue Canadienne d'Economique*, *33*(1), 149–174.

Diamond, P. A., & Mirrlees, J. A. (1978). A Model of Social Insurance with Variable Retirement. *Journal of Public Economics*, *10*, 295–336.

Diamond, P., & Sheshinski, E. (1995). Economic aspects of optimal disability benefits. *Journal of Public Economics*.

Feldstein, M. (1976). On the Theory of Tax Reform. *Journal of Public Economics*, *6*(January 1976), 77–104.

Goodin, R. (1985). Erring on the Side of Kindness in Social Welfare Policy. *Policy Sciences*.

Immonen, R., Kanbur, R., Keen, M., & Tuomala, M. (1998). Tagging and Taxing: The Optimal Use of Categorical and Income Information in DesigningTax/Transfer Schemes. *Economica*, *65*(258), 179–192.

Jacquet, L. (2006). Optimal disability assistance when fraud and stigma matter. *Queen's Economics Department Working Paper*, (1098).

Jacquet, L. (2013). Tagging and redistributive taxation with imperfect disability monitoring. *Social Choice and Welfare*, *42*(2), 403–435. doi:10.1007/s00355-013-0738-y

Kaplow, L. (2010). Concavity of utility, concavity of welfare, and redistribution of income. *International Tax and Public Finance*, *17*(1), 25–42. doi:10.1007/s10797-008-9097-5

Kleven, H. J., & Kopczuk, W. (2011). Transfer Program Complexity and the Take-Up of Social Benefits. *American Economic Journal: Economic Policy*, *3*(February), 54–90.

Kreiner, C. T., & Tranaes, T. (2005). Optimal Workfare with Voluntary and Involuntary Unemployment*. *Scandinavian Journal of Economics*, *107*(3), 459–474.

Mankiw, N. G., & Weinzierl, M. (2010). The Optimal Taxation of Height: A Case Study of Utilitarian Income Redistribution. *American Economic Journal: Economic Policy*, *2*(1), 155–176.

Mirrlees, J. (1971). An Exploration in the Theory of Optimal Income Taxation. *The Review of Economic Studies*, *38*(2), 175–208.

Mkandawire, T. (2005). Targeting and universalism in poverty reduction. *Social Policy and Development , United Nations Research Institute for Social Development*, (Program Paper Number 23).

Nichols, A. L., & Zeckhauser, R. J. (1982). Targeting Transfers through Restrictions on Recipients. *American Economic Review*, *72*(2), 372–377.

Parsons, D. (1980). The decline in male labor force participation. *The Journal of Political Economy*, *88*(1), 117–134.

Parsons, D. (1991). The health and earnings of rejected disability insurance applicants: comment. *The American Economic Review*, *81*(5), 1419–1426.

Parsons, D. (1996). Imperfect "tagging"in social insurance programs. *Journal of Public Economics*, *62*, 183–207.

Salanié, B. (2002). Optimal demogrants with imperfect tagging. *Economics Letters*, *75*, 319–324.

Sheshinski, E. (1972). The Optimal Linear Income Tax. *The Review of Economic Studies*.

Skocpol, T. (1991). Targeting within universalism: Politically viable policies to combat poverty in the United States. In C. Jencks & P. Peterson (Eds.), *The urban underclass*. Washington, D.C.: The Brookings Institution.

Slack, S.,& Ulph.D.(2014) Optimal Universal and Categorical Benefits with Imperfect Enforcement and Classification Errors, *University of St. Andrews Discussion Paper*

Stern, N. (1976). On the specification of models of optimum taxation. *Journal of Public Economics*, *6*, 123–162.

Stern, N. (1982). Optimum taxation with errors in administration. *Journal of Public Economics*, *17*, 181–211.

Van Parijs, P. (2004). Basic Income: A Simple and Powerful Idea for the Twenty-First Century. *Politics & Society*, *32*(1), 7–39.

Viard, A. D. (2001a). Optimal Categorical Transfer Payments: The Welfare Economics of Limited Lump-Sum Redistribution. *Journal of Public Economic Theory*, (December 2000), 483–500.

Viard, A. D. (2001b). Some Results on the Comparative Statics of Optimal Categorical Transfer Payments. *Public Finance Review*, *29*(2), 148–180.

Wachter, T. Von. (2011). Trends in Employment and Earnings of Allowed and Rejected Applicants to the Social Security Disability Insurance Program. *American Economic Review*, *101*(7).