NBER WORKING PAPER SERIES

TRANSFER PROGRAM COMPLEXITY AND THE TAKE UP OF SOCIAL BENEFITS

Henrik Jacobsen Kleven
Wojciech Kopczuk

Working Paper 14301
http://www.nber.org/papers/w14301

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2008

Transfer Program Complexity and the Take Up of Social Benefits
Henrik Jacobsen Kleven and Wojciech Kopczuk
NBER Working Paper No. 14301
September 2008
JEL No. H53,I3,I38

ABSTRACT

This paper models complexity in social programs as a byproduct of efforts to screen between deserving and undeserving applicants. While a more rigorous screening technology may have desirable effects on targeting efficiency, the associated complexity introduces transaction costs into the application process and may induce incomplete take up. The paper integrates the study of take up with the study of classification errors of type I and type II, and argue that incomplete take up can be seen as a form of type I error. We consider a government interested in ensuring a minimum income level for as many deserving individuals as possible, and characterize optimal programs when policy makers can choose the rigor of screening (and associated complexity) along with a benefit level and an eligibility criterion. It is shown that optimal program parameters reflect a trade-off at the margin between type I errors (including non-takeup) and type II errors. Optimal programs that are not universal always feature a high degree of complexity. Although it is generally possible to eliminate take up by the undeserving (type II errors), policies usually involve eligibility criteria that make them eligible and rely on complexity to restrict their participation. Even though the government is interested only in ensuring a minimum benefit level, the optimal policy may feature benefits that are higher than this target minimum. This is because benefits generically screen better than either eligibility criteria or complexity. We present numerical simulations on comparative statics with respect to budget size, ability distribution, complexity costs, and stigma. Our results are discussed in light of empirical findings for public programs in the United States.

Henrik Jacobsen Kleven
Department of Economics & STICERD
LSE
Houghton Street
London WC2A 2AE
United Kingdom
h.j.kleven@lse.ac.uk

Wojciech Kopczuk
Columbia University
420 West 118th Street, Rm. 1022 IAB
MC 3323
New York, NY 10027
and NBER
wk2110@columbia.edu

1 Introduction

The United States operates a large number of social programs offering support to those in need. This includes cash assistance to the poor, food stamps, health insurance, housing programs, child care support, and social security to the aged, blind and disabled. We observe several striking differences in the design and outcomes of these programs. One difference lies in the degree of *targeting* to selected groups of individuals viewed as 'deserving'. Although the U.S. welfare state in general relies on a much higher degree of targeting than most other countries, there is substantial variation in targeting across different programs within the U.S. At one end of the spectrum, the Medicare program is almost universal, whereas at the other end of the spectrum, disability insurance programs serve a relatively small population satisfying very stringent eligibility criteria. A second difference lies in the way social programs are administered and in their degree of *complexity*. Targeted programs tend to be characterized by a substantial amount of complexity and administrative hassle, whereas universal programs are simpler and more transparent. A third difference lies in the *take up* of social benefits. Incomplete take up among intended recipients is an important issue in all means-tested programs in the U.S., but there is huge variation in participation across different programs. This is shown in Table I, which compares 10 public programs in the U.S. along the three dimensions just described.

The large variance in program design and outcomes reflects a number of underlying differences in factors such as funding, benefit generosity, ability distribution among potential applicants, observability of true eligibility, and costs of screening. We set out a theoretical framework that accounts for all of these underlying differences in order to facilitate an analysis of targeting, complexity and take up in social programs. We contribute to the existing literature along three dimensions. First, we take an initial step towards modeling and analyzing complexity in public programs. We go beyond viewing complexity as a negative side-effect of targeted programs, and treat it instead as a policy instrument which is chosen alongside benefit levels and eligibility rules in the design of a program. Second, we explain why governments may want to design a program with high complexity and incomplete take up by eligibles even though they have access to policy instruments which could increase take up. Third, we integrate the study of take up with the study of classification errors of type I (false rejections) and type II (false awards) in benefit award processes. In fact, we argue that non-enrollment in social programs can be seen as a form of type I error, and that it has to be understood by considering the trade-off with the usual type I and II errors.

Empirical economists have long been concerned with the issue of incomplete take-up rates in public programs. The empirical literature hypothesizes three possible explanations for incomplete take up: welfare stigma, transaction costs, and imperfect information. The seminal work in this area is the Moffitt (1983) model of welfare stigma, suggesting that eligibles may find non-participation in a welfare program optimal because it is viewed as demeaning and shameful. But the stigma hypothesis is consistent with other more concrete costs associated with taking up social benefits. Indeed, a substantial amount of evidence have documented that applying for welfare benefits involves large transaction costs arising from application processes being complex, tedious and time-consuming (Moffitt, 2003; Currie, 2004).

The complexity of welfare programs may arise from detailed eligibility criteria, rigorous documentation requirements, difficult and time-consuming forms, or requiring multiple trips to the program office for interviewing and testing. Moreover, some programs involve frequent re-certification to continue to receive the benefit, and applicants are frequently rejected because they fail to fulfill the administrative requirements within the required time. Notice that these forms of complexity reflect, at least in part, an attempt of program administrators to monitor true eligibility accurately, and hence complexity may have some desirable effects on the magnitude of classification errors. At the same time, these monitoring activities introduce hassle into the application process, which may hurt take up. Indeed, empirical research has shown that complexity and administrative hassle do reduce program enrollment (Currie and Grogger, 2001; Bitler et al., 2003; Daly and Burkhauser, 2003; Aizer, 2007), and that such effects may be more important than stigma (Currie, 2004).

Despite the fact that complexity and administration seem to be very important for the effects of public policies in general, and for the take up of social benefits in particular, we are not aware of theoretical work modeling the complexity of public policy. Instead, the literature on mechanism design has focused on the generosity and structure of benefits and the incentives for ineligibles to reveal themselves truthfully. The key assumption in this literature is that innate ability is unobservable at any cost, whereas earnings are perfectly observable at no cost. The government has no access to a monitoring technology to assess true eligibility and therefore has to rely on limitations in earnings-based benefits to induce self-revelation. Our paper goes beyond this extreme assumption about information by modeling the information collection process — the monitoring technology — used to elicit true eligibility for social benefits. The empirical work by Benítez-Silva et al. (2004) on disability insurance programs in the U.S. demonstrates that the monitoring technology can be a very important aspect of program design.

The way we model the complexity — or the rigor — of a monitoring technology is consistent with the evidence discussed above. It is an instrument used by program administrators to increase the intensity of screening in order to extract a better signal of true eligibility, but which makes the application process more costly and therefore may induce non-participation by eligibles. Our model also accounts for imperfect information about eligibility on part of the potential welfare applicants. The role of imperfect information for non-enrollment into social programs is well-documented (e.g. Daponte et al., 1999; Heckman and Smith, 2004), and it serves to reinforce the importance of complexity in the decision to apply for welfare. It is exactly because of imperfect information about eligibility that an individual may be reluctant to incur the transaction costs associated with applying.

We characterize program characteristics in equilibrium when policy makers can choose standard policy instruments — a benefit level and an eligibility rule — along with the additional instrument capturing program complexity. We model a government interested in income maintenance, i.e. ensuring a minimum income level for as many truly poor ('deserving') individuals as possible, being constrained by a limited budget. We consider income maintenance rather than social welfare maximization, because the former is more consonant with real-world policy debates (Besley and Coate, 1992, 1995). In other words, the primary purpose of the paper is to understand observed program design rather than exploring the design chosen by a welfarist social planner.

We show that optimal program parameters reflect a trade-off at the margin between type I errors (including non-takeup) and type II errors. Optimal programs that are not universal always feature a high degree of complexity. Although it is generally possible to eliminate take up by the undeserving (type II errors), policies usually involve eligibility criteria that make them eligible and rely on complexity to restrict their participation. These policies feature incomplete take up by the deserving along with classification errors of both type I and II in the benefit award process. Even though the government is interested only in ensuring a minimum benefit level, the optimal policy may feature benefits that are higher than this target minimum. This is because benefits generically screen better than either eligibility criteria or complexity. We present numerical simulations on comparative statics with respect to budget size, ability distribution, complexity costs, and stigma. Our results are discussed in light of empirical findings for public programs in the United States.

The rest of the paper is organized as follows. Section 2 defines the different classification errors in public programs and discusses how they have been studied in the literature. Section 3 presents our model of transfer program complexity, and derives a number of results on program design, complexity and take up. Section 4 presents numerical simulations, and Section 5 offers a discussion of applications, assumptions, and extensions.

2 Classification Errors in Social Programs

We view non-participation by eligibles in social programs as a result of program parameters chosen by policy makers. Viewed in this way, it is natural to think of incomplete take up as a form of classification error of type I — a false negative. We introduce the following terminology:

Definition 1 (classification errors)

- *Type Ia errors (incomplete take up) occur if a program design results in some truly eligible individuals not applying for benefits.*
- *Type Ib errors (rejection errors) occur if a program design results in some truly eligible individuals applying for benefits and being rejected.*
- *Type II errors (award errors) occur if a program design results in some truly ineligible individuals applying for benefits and being accepted.*

For a government wanting to alleviate poverty among those who are truly eligible, being constrained by a limited budget, it is desirable to minimize all sources of error. The occurrence of type Ia and type Ib errors undermine the goal of poverty alleviation, whereas the occurrence of type II errors make the program more expensive and divert government revenues away from other productive uses. Hence, the choice of parameters in a welfare program — benefits, eligibility rules and the complexity of the screening process — reflects the effect of each parameter on the different kinds of error. Indeed, a central message in this paper is that public programs have to be understood by integrating the treatment of all three types of classification error and considering the trade-off between them.

While a large empirical literature has analyzed the occurrence of incomplete take up and hence the occurrence of type Ia error, much fewer papers have attempted to estimate the occurrence of type Ib and type II errors.[1] A small literature looking at classification error rates in U.S. social security disability award processes suggests that both award and rejection errors are very common. For example, the recent and interesting paper by Benítez-Silva et al. (2004) estimates the award error rate of about 20% and the rejection error rate of about 60%.[2]

Opposite the empirical literature, we have a theoretical literature analyzing optimal mechanism design in transfer programs. A large set of papers have studied the relationship between benefit structure and the incentives self-revelation.[3] Assuming that there exists no monitoring technology to assess true eligibility, these papers deal exclusively with type II errors and how to avoid them by restricting benefits in different ways.

A smaller set of papers, starting with the important contribution by Akerlof (1978), introduced a simplified monitoring technology into the mechanism design problem. This monitoring technology — labelled 'tagging' by Akerlof — can identify perfectly a given subset of eligibles. The screening process is imperfect because some eligibles are not tagged (a type Ib error) and it is exogenous to policy makers. While Akerlof did not allow for type II errors, a few subsequent papers extended his work to incorporate two-sided classification error (Stern, 1982; Diamond and Sheshinski, 1995; Parsons, 1996) assuming fixed award and rejection error rates.

Our paper is different from the existing literature in two important respects. First, the monitoring technology is not exogenous in our model. We analyze the choice of monitoring technology — the rigor of screening and associated complexity — as a policy instrument. Second, we integrate the treatment of all three types of classification error, taking into account that the magnitude of errors are endogenous to parameters chosen by policy makers. This includes allowing for reduced take up in response to increased program complexity. In order to zoom in on the participation decision in public programs, our framework abstracts from other types of behavioral responses such as labor supply responses.

[1]A qualification is in place here. Empirically, it is often difficult to distinguish between type Ia and type Ib errors. What we observe in survey data is that some eligible individuals do not receive benefits, and this may in principle reflect either form of type I error. This implies that estimates of take-up rates may in part capture rejection errors as well.

[2]An early study by Nagi (1969) reached broadly similar conclusions.

[3]This literature includes Diamond and Mirrlees (1978) Nichols and Zeckhauser (1982), Blackorby and Donaldson (1988), Bruce and Waldman (1991), Besley and Coate (1992, 1995) and Saez (2002).

3 A Model of Social Program Complexity

3.1 Individuals

We assume that each individual is characterized by two parameters: an innate characteristic a and the precision by which this characteristic can be observed by outsiders σ. The characteristic a may reflect market productivity, or it may reflect other types of characteristics — say health or disability — depending on the program being considered. In the following, we refer to a simply as 'ability' or 'skill'. These skills are private information and cannot be ascertained directly by anyone else. Instead, if the individual attempts to claim welfare benefits, the government can test the individual and obtain a signal of true ability, $\tilde{a} = a + \varepsilon/\alpha$. The noise term ε reflects that program testing is imperfect, whereas the parameter α is a policy choice capturing the rigor of the test. We will come back to the interpretation and implications of α below.

Based on empirical analyses of benefit award processes (see Benítez-Silva et al., 2004), we assume that $\tilde{a}$ is a noisy but unbiased indicator of true ability so that ε is distributed with mean zero and variance σ^2. We assume that the normalized distribution of ε/σ (which has mean zero and variance one) is characterized by a c.d.f. $P(.)$, which is identical for everyone. We allow for the fact that the precision of measured skill, σ, may vary across individuals even if they have identical abilities. The heterogeneity in σ reflect that equally eligible individuals may test with more or less uncertainty in the welfare program. Aspects such as language barriers, unfamiliarity with the administrative procedures, inability to understand the formal requirements of the test, etc., would all contribute to creating more uncertainty in the test.

As for the α-parameter, one possible interpretation is to view it as the number of tests. Under this interpretation, the government can subject an applicant to different tests in order to obtain indicators of skill. Each test leads to an indicator given by $a_i = a + \varepsilon_i$ where $\varepsilon_i \sim N(0, \sigma^2)$ — i.e., each indicator is a normally distributed unbiased indicator of the true skill level with variance σ^2. Examples of tests are interviews with case workers, a requirement to provide supporting documents, an opinion of a medical commission regarding disability, etc. The common denominator of these types of indicators is that they are costly to an individual and the outcome is not known a priori. Under this interpretation, the government estimates the skill of an individual using the arithmetic mean of α^2 tests, the distributional properties of which is exactly identical to $\tilde{a} = a + \varepsilon/\alpha$ with $\varepsilon \sim N(0, \sigma^2)$.[4]

More generally, the policy parameter α determines the extent of randomness in the application process. We maintain the assumption that reducing randomness imposes a burden on individuals, because of increased out-of-pocket application costs and more time spent on forms, interviewing and providing documentation, etc. We represent the cost to the individual of complexity α by a function $f(\alpha)$.

We assume that the government sets an eligibility criterion for receiving benefits denoted by $\bar{a}$. When the government relies on complexity α, benefits are granted to applicants who

[4]Alternatively, if we had specified $\tilde{a} = a + \varepsilon/\sqrt{\alpha}$, then α (rather than α^2) would be the number of tests. We use the specification above because it is notationally simpler.

satisfy

$$\tilde{a} = a + \varepsilon/\alpha < \bar{a}. \tag{1}$$

The probability that an applicant with skill level a and precision σ receives benefits is therefore given by $P\left(\frac{\alpha(\bar{a}-a)}{\sigma}\right)$. We come back to the properties of $P(.)$ below.

When making the participation decision, an individual knows the probability of being granted the benefit and trades off the potential utility gain from welfare payments against the cost of applying. We assume that utility depends on consumption C — equal to the sum of ability a and the (potential) welfare benefit B — and on application costs $f(\alpha)$. The utility level is given by $u(C - Af(\alpha))$, with A being an indicator variable for having applied. We make the standard assumption that $u(.)$ is increasing and weakly concave (allowing for the possibility of risk neutrality). We also assume that $\lim_{C\to\infty} u(C) = \infty$ and $\lim_{\alpha\to\infty} f(\alpha) = \infty$. An individual chooses to apply when

$$P\left(\frac{\alpha(\bar{a}-a)}{\sigma}\right) u(a + B - f(\alpha)) + \left(1 - P\left(\frac{\alpha(\bar{a}-a)}{\sigma}\right)\right) u(a - f(\alpha)) > u(a), \tag{2}$$

and, conditional on applying, will receive benefits with the probability of $P\left(\frac{\alpha(\bar{a}-a)}{\sigma}\right)$.

Ceteris paribus, a higher probability of receiving benefits increases the expected utility from applying. The probability of receiving benefits conditional on applying depends on the complexity parameter α, eligibility criterion $\bar{a}$, ability level a, and personal precision of ability signals σ. A higher $\bar{a}$ unambiguously increases the probability, whereas a higher a unambiguously decreases it. The effect of complexity α and precision σ depends on the sign of $\bar{a} - a$. When $\bar{a} > a$, higher complexity and better precision both increase the probability of receiving benefits. This is intuitive: when the individual is eligible under perfect information, reducing the noise in the eligibility metric is helpful. When $\bar{a} < a$, we have the opposite situation. While greater complexity may increase or decrease the likelihood of receiving benefits depending on the sign of $\bar{a} - a$, its effect on the ex post utility level is unambiguously negative regardless of whether benefits are received or not.

Using the participation constraint (2), we may solve for the minimum probability, $\tilde{P}$, consistent with applying for benefits:

$$\tilde{P}_a(\alpha, B) \equiv \frac{u(a) - u(a - f(\alpha))}{u(a + B - f(\alpha)) - u(a - f(\alpha))}. \tag{3}$$

Individuals with a probability of receiving benefits above this critical value choose to apply for benefits, while the rest choose not to apply. In general, the threshold probability depends on the skill level a. In particular, it may be shown to be decreasing or increasing in ability depending on whether the utility function features decreasing or increasing absolute risk aversion. We do not have a strong prior as to whether higher ability individuals are willing to accept lower odds when applying for benefits, but the realistic case of decreasing absolute risk aversion would imply that this is the case. There are a number of other factors not modeled here that would have implications for this issue. For example, we restrict attention to a flat benefit although in practice the size of the benefit could depend on the realization of the indicator $\tilde{a}$. Letting the benefit depend negatively on $\tilde{a}$ would increase the minimum odds acceptable to the higher-ability individuals. Application costs may also vary with the ability level. On the one hand, if it is easier for high-ability applicants to file an application, their minimum acceptable probability would be lower. On the other hand, high-ability applicants tend to face higher opportunity cost of time spent applying, which would make their threshold probability higher.

We will simplify the analysis by restricting attention to the class of preferences that eliminates the dependence of $\tilde{P}_a$ on a:

Assumption 1 *The utility function has the Constant Absolute Risk Aversion (CARA) form,* $u(C) = \frac{1-e^{-\beta C}}{\beta}$, *where* $\beta \geq 0$ *(this specification reduces to risk-neutrality* $u(C) = C$ *for* $\beta = 0$*).*

Assumption 1 is not trivial and we elaborate on its consequences in the final section.[5] Under this assumption, the threshold probability level for applying is given by

$$\tilde{P}(\alpha, B) = \begin{cases} \frac{1-e^{-\beta f(\alpha)}}{1-e^{-\beta B}}, & \text{when } \beta > 0, \\ \frac{f(\alpha)}{B}, & \text{when } \beta = 0, \end{cases} \tag{4}$$

which is no longer a function of the ability level. It is straightforward to show that $\frac{\partial \tilde{P}}{\partial \alpha} > 0$ and $\frac{\partial \tilde{P}}{\partial B} < 0$: a higher level of complexity increases the minimum acceptable probability of receiving benefits, whereas higher benefits decrease it.

Expressing the participation constraint as

$$P\left(\frac{\alpha(\bar{a} - a)}{\sigma}\right) > \tilde{P}(\alpha, B), \tag{5}$$

it can be solved for the precision level corresponding to indifference between applying and not applying:

$$\bar{\sigma}_a(\alpha, \bar{a}, B) \equiv \frac{\alpha(\bar{a} - a)}{P^{-1}\left(\tilde{P}(\alpha, B)\right)}. \tag{6}$$

When $\bar{a} > a$, individuals with σ lower than $\bar{\sigma}_a$ (high precision) apply for benefits. When $\bar{a} < a$, only individuals with σ greater than $\bar{\sigma}_a$ (low precision) choose to apply.

3.2 Population

We assume that there are two levels of ability: a low level a_L and a high level a_H. At each ability level, individuals are heterogeneous with respect to σ: for some, their ability level may be easily observable while for others it may be very difficult to ascertain without extensive testing. We note the following:

Remark 1 *At each ability level, let the precision of measured skill* σ *be distributed on* $[0, \infty)$. *There are three qualitative cases for the distribution of the probability of receiving benefits in the population:*

[5]The only reason for making the CARA-assumption is that it eliminates the dependence of $\tilde{P}_a$ on a. We argue in the final section that this implication of CARA is a realistic description of the real world, although it may reflect dimensions of heterogeneity not incorporated in our model. Generalization of the model to non-CARA preferences that preserves the central property of $\tilde{P}_a$ are possible by adding additional dimensions of heterogeneity.

1. *$\bar{a} \leq a_L < a_H$. Probabilities are in $(0, P(0)]$ and increasing in σ (always strictly increasing for high-ability individuals, strictly increasing for low-ability individuals only if $\bar{a} < a_L$).*

2. *$a_L < \bar{a} \leq a_H$. Probabilities for low-ability individuals are in $(P(0), 1)$ and strictly decreasing in σ; probabilities for high-ability individuals are in $(0, P(0)]$ and increasing in σ (strictly increasing if $\bar{a} < a_H$).*

3. *$a_L < a_H < \bar{a}$. Probabilities are in $(P(0), 1)$ for both types and are strictly decreasing in σ.*

Whenever $a \neq \bar{a}$, any probability in the appropriate open interval, $(0, P(0))$ or $(P(0), 1)$, can be attained for some $\sigma \in [0, \infty)$.

This remark implies a "non-monotonicity" in committing Type II errors: they have to be committed when either $\bar{a} < a_L$ or $\bar{a} > a_H$, but not for intermediate values of $\bar{a}$. In the former cases, because the intervals of probabilities of receiving benefits are identical for the low- and high-ability populations, it will be impossible to avoid type II errors altogether. Note that $P(0)$ reflects a property of the normalized distribution of ε and therefore it is a constant independent of policy parameters or individual characteristics. In the natural case where the likelihoods of over- and understating true ability are identical such that median(ε) is zero, we have $P(0) = 1/2$. Recall that the threshold probability for applying $\tilde{P}$ depends on the policy parameters α and B and, if these parameters are not constrained, $\tilde{P}$ can take any value. As a result,

Remark 2 *There exist policy parameters that result in no Type II errors ("full separation"); only low ability individuals apply. Such policies are characterized by $a_L < \bar{a} \leq a_H$ and $\tilde{P}(\alpha, B) \geq P(0)$.*

Moreover, there also exist policy parameters that additionally result in no Type Ia errors. They are characterized by $a_L < \bar{a} \leq a_H$ and $\tilde{P}(\alpha, B) = P(0)$.

One of the objectives of our analysis will be to determine whether policies with no type Ia and type II errors are optimal and, despite their apparent attractiveness, we will show that in the most interesting cases they are not. In particular, note that a government implementing a policy of full separation where $\tilde{P}(\alpha, B) = P(0)$ — i.e., no type II or type Ia errors — will continue to make Type Ib errors. That is, despite that only low-ability individuals are applying, some of them will be rejected. In fact, in the case of a symmetric distribution for ε where $P(0) = 1/2$, some low-ability applicants will face probabilities of receiving benefits as low as 1/2. Reducing the number of Type Ib errors can be accomplished by increasing the rigor of screening α, but in order to avoid Type Ia errors, the government must increase benefits correspondingly. Such increases are costly and, at the same time, constrained in their size when one wants to simultaneously discourage high-ability individuals from applying. As a result, the government faces serious constraints in pursuing policies that reduce the number of type Ib errors without introducing other types of classification error.

As we will demonstrate, these constraints may be severe enough to justify committing all three types of error.

To complete the characterization of the assumptions about the population, we need to specify the distribution of σ. We will denote the c.d.f. of the distribution of σ for ability-type a by G_a and the corresponding density function by g_a. The support of both distributions is assumed to be $[0, \infty)$. We assume that $g_a(0) = 0$, the density of individuals with perfectly observable skill is zero. The number of individuals of type a is given by $\bar{N}_a \equiv \int_0^\infty dG_a(\sigma)$, with both $\bar{N}_L$ and $\bar{N}_H$ assumed to be positive and finite.

Some of our results will depend on the following regularity assumptions:

Assumption 2 (thin tail for low ability) $\lim_{\sigma \to \infty} \sigma^2 g_L(\sigma) = 0.$

Assumption 3 (finite slope of density at zero for high ability) $\lim_{\sigma \to 0} g_H'(\sigma) < \infty.$

The first assumption states that the distribution of σ has no thick tail. In particular, it rules out the Pareto distribution, but it allows for distributions that have thinner tails such as the log-normal distribution. Intuitively, it will allow for the number of low-ability applicants to respond smoothly to policy changes that just discourage applying by everyone. The second assumption will guarantee that small changes in policy that make it beneficial for the high ability individuals to apply will result in only a small influx of them.

To summarize the model so far, Figure 1 illustrates the distribution of $P\,(.)$ and classification errors for a particular program. Both panels show the P-distribution for low- and high-ability individuals, with Panel A highlighting the results for the low-types and Panel B highlighing the results for the high-types. The figure illustrates a program with $\bar{a} > a_H$, implying that $P\,(.)$ is distributed on the interval $(P(0), 1)$ for both types and is strictly decreasing in σ. We focus on this type of program, because it turns out to be interesting later on. The graphs are based on an actual numerical simulation that we discuss in detail in Section 4. The distribution of $P\,(.)$ is determined by the distribution of the noise term ε/σ (which is assumed to be normal so that $P\,(0) = 1/2$) along with the distribution of the precision of measured skill σ (which is assumed to be log-normal). Given $\bar{a} > a_H$, densities of $P\,(.)$ are positive everywhere in the open interval $\left(\frac{1}{2}, 1\right)$ for both types, because any probability in this interval can be attained for some $\sigma \in [0, \infty)$. At a given σ, low-ability applicants have a higher probability of being awarded benefits, and hence the P-distribution for low-ability individuals is shifted to the right compared to the distribution for high-ability individuals. The two types have the same threshold probability $\tilde{P}$ ($= 0.736$ in the simulation), and individuals with higher Ps than this (corresponding to those with low σs) apply for benefits. The program is associated with all three types of classification error. In the low-ability distribution, type Ia (take-up) errors are committed in the region to the left of $\tilde{P}$, while type Ib errors occur in the region to the right because probabilities of acceptance are lower than 1. In the high-ability distribution, type II errors occur in the region to the right of $\tilde{P}$ because probabilities of acceptance are greater than zero (in fact, greater than 0.736). An interesting question is whether a program outcome of this kind can be an equilibrium outcome. To study this question, we turn to the final piece of the model: the specification of the government's objective.

3.3 Government

We consider a problem of income maintenance extending the specification of Besley and Coate (1992, 1995). They considered the design of income maintenance programs ensuring that each individual obtains a target minimum benefit at a minimum fiscal cost. In our model — as in reality — we do not necessarily have full participation, because eligible individuals may choose not to apply for the benefit and because eligible applicants may be rejected by program administrators due to imperfect testing. Hence, the objective becomes to provide a minimum benefit for as many low-ability (truly deserving) individuals as possible, being constrained by a limited budget.[6]

Denoting by $\bar{B}$ the target minimum benefit and by R the exogenously given budget size, the government's problem may be written as

$$\max_{\alpha,\bar{a},B} \quad N_L(\alpha,\bar{a},B) \tag{7}$$

subject to

$$[N_L(\alpha,\bar{a},B) + N_H(\alpha,\bar{a},B)]\,B \leq R \tag{8}$$

and

$$B \geq \bar{B}, \tag{9}$$

where $N_a(\alpha,\bar{a},B)$ is the number of successful applicants of type a as a function of policy parameters.

There are several aspects of our policy objective that deserve mentioning. First, even though policy makers are interested in providing low-income support, they are not concerned with the utility cost that program complexity imposes on individuals. As discussed in the beginning, this is not a welfarist framework. While an extension of the model to social welfare maximization would be interesting, the approach adopted here fits better with actual political debates on poverty relief, which tends to be based on the notion that being poor means having too little income, not having too low utility. In other words, we view our modeling strategy primarily as a piece of positive economics, even though it may of course also be seen as normative if one subscribes to the view that income maintenance rather than utility maximization is normatively justifiable.

Second, we assume that benefits cannot fall below some minimum value despite that, in general, not all of the low-ability individuals are going to receive benefits (note though that the government can increase benefits above $\bar{B}$). Reducing benefits to a small enough value would allow for providing benefits to everyone, and therefore allowing for unrestricted benefits is incompatible with a non-trivial problem of maximizing the number of deserving recipients. Absent a direct welfarist objective, providing a target minimum income to successful recipients is a natural way of modeling the goal of poverty alleviation.

Third, we do not model the revenue side of the system. While the distortions introduced are undoubtedly important, our model does not necessarily describe the full society. Rather,

[6] Interestingly, as noted by Besley and Coate (1992), this policy objective fits Mill's (1848) characterization of the poverty-alleviation problem as "how to give the greatest amount of needful help, with the smallest encouragement to undue reliance on it."

our "high-ability" individuals should be viewed as still relatively poor but not poor enough to be in need of social welfare. Under this interpretation, benefits are financed by a wealthier (and not modeled) segment of the society.

Fourth, the government pursues policies that are horizontally inequitable. Some low-ability individuals are going to receive benefits while others will not. The point we make is that it is not possible to pursue a horizontally equitable policy unless one is able to provide benefits to everyone — rich and poor. This is a property of this model and, likely, of the real world: in order to reach every poor individual we would have to accept a very large number of Type II errors.

In the following section, we characterize social programs that solve the problem specified above. We show that the solution depends, among other things, on the size of the program budget R. We restrict attention to program budgets satisfying $R < \bar{B}\left(\bar{N}_L + \bar{N}_H\right)$. If the budget were larger than this, the government's problem has a simple solution: the number of low-ability applicants reaches its theoretical maximum $\bar{N}_L$ by giving a universal benefit $B \geq \bar{B}$ to everybody, which is an affordable policy when $R \geq \bar{B}\left(\bar{N}_L + \bar{N}_H\right)$. A universal benefit would be implemented by letting the eligibility criterion $\bar{a}$ tend to infinity, in which case the probability of receiving benefits tends to 1 for everybody.

3.4 Results

We begin our analysis of the model by specifying the first-best allocation that the government would pursue under full information.

Definition 2 (first best) *Suppose that it is possible to observe both a and σ. Then the optimal policy provides benefits $\bar{B}$ to $\min\left(R/\bar{B}, \bar{N}_L\right)$ individuals with ability a_L. The choice of these individuals is undetermined (there are many first-best policies).*

The requirement that a first-best program must reach $\min\left(R/\bar{B}, \bar{N}_L\right)$ recipients amounts to saying that the program either spends the entire budget R or, if there are unused funds, this is because there are no low-ability individuals left who have not received benefits. Notice that this definition of a first-best policy is conditional on the *exogenous* funds R allocated to the program, and therefore does not account for the fact that the amount of revenue allocated to a program may in itself depend on the information available to policy makers. In particular, if information were perfect, it would not make sense to allocate funds to a program such that $R > \bar{B} \cdot \bar{N}_L$ given the problem specified in (7)-(9). The purpose of the above definition is not to specify a "global" first best, but to specify the best possible outcome for a program designed under imperfect information at any given budget size R. Given the presence of imperfect information, even if a program has a large budget, $R > \bar{B} \cdot \bar{N}_L$, it will not be able to reach all low-ability individuals, and this is therefore an interesting case to consider. We come back to this point below.

As noted in Remark 2, there exist policies that result in providing benefits only to low-ability individuals. In certain cases, it is possible to achieve one of the first-best allocations despite the lack of perfect information.

Proposition 1 (first best) *First-best programs always involve full separation. For R small enough, first-best is feasible and the optimal program is characterized by $a_L < \bar{a} \leq a_H$, $B = \bar{B}$, and $\tilde{P}(\alpha, \bar{B}) \geq P(0)$. The optimum is not necessarily unique.*

Proof. Setting policy instruments such that $a_L < \bar{a} \leq a_H$, $B = \bar{B}$, and $\tilde{P}(\alpha, \bar{B}) \geq P(0)$ ensures that (i) benefits are provided only to low-ability individuals and (ii) each recipient receives only the target minimum. The final requirement for a program to be first best is that $N_L = \min\big(R/\bar{B}, \bar{N}_L\big)$. To see that this is only possible if the budget is "small", notice that the class of programs specified above can never reach all low-ability individuals. The number of low-ability recipients within this class of programs is maximized by setting $\bar{a} = a_H$. Given $\bar{a} = a_H$ and $B = \bar{B}$, it is not possible to set α such that $N_L = \bar{N}_L$. This is because, at any finite α, the probability of rejection for each low-ability applicant, $1 - P\left(\frac{\alpha(a_H - a_L)}{\sigma}\right)$, is greater than zero, and α cannot be increased without bound because the associated increase in $\tilde{P}(\alpha, \bar{B})$ would ultimately discourage all applications. Hence, there is a maximum number of low-ability individuals $N_L^* < \bar{N}_L$ that can be reached within the class of programs we have specified. Define $R^* = \bar{B} \cdot N_L^* < \bar{B} \cdot \bar{N}_L$ as the largest budget that can be spent on this type of program, and denote by α^* the level of α that achieves N_L^*. Now, for $R > R^*$, we always have $N_L < \min\big(R/\bar{B}, \bar{N}_L\big)$ and therefore not first best. Conversely, for any $R < R^*$, we can always ensure $N_L = \min\big(R/\bar{B}, \bar{N}_L\big)$ by increasing α beyond α^* (because a higher α increases $\tilde{P}$ and low-ability recipients fall to zero for high enough α because complexity costs outweigh benefits). Finally, because first-best programs are always associated with $R/\bar{B} < \bar{N}_L$, and because programs that provide benefits to any high-ability individual imply $N_L < R/\bar{B}$, first-best policies always involve full separation. ■

The proposition shows that, if the budget is small enough, we can spend all of the money providing the target minimum $\bar{B}$ to low-ability individuals only, which is the first-best outcome a the given budget. In particular, this is the case for $R \in (0, R^*]$ where $R^* < \bar{B} \cdot \bar{N}_L$. At the other extreme, if the budget is very large, $R \geq \bar{B}\left(\bar{N}_L + \bar{N}_H\right) \equiv \bar{R}$, we pointed out above that the optimal program offers a universal benefit $B \geq \bar{B}$ to everybody. This type of program is also first best, because the government can never do better than this given the specified policy objective.[7] The most interesting case is the one in between the small-budget case $R \in (0, R^*]$ and the very-large-budget case $R \in \left(\bar{R}, \infty\right]$, i.e. where $R \in \left(R^*, \bar{B}\left(\bar{N}_L + \bar{N}_H\right)\right)$, in which case first best is not feasible. The rest of this section is devoted to characterizing optimal social programs in this intermediate range.

We have to consider both full separation (but non-first best) programs and non-full separation programs. We start by noting that, at any budget size, full separation policies are always feasible:

Lemma 1 (type II errors can be avoided) *At any budget size, there exists a policy that satisfies the budget constraint and involves full separation.*

Proof. We just need to consider the situations where the first-best policy is not feasible. Let us consider policies involving $\tilde{P}(\alpha, B) = P(0)$ and $\bar{a} = a_H$. For such policies, no high-ability individuals apply, whereas all of the low-ability individuals apply. Consider increasing α while simultaneously increasing B to keep $\tilde{P}(\alpha, B) = P(0)$. By construction, this policy will retain full separation. As $\alpha \to \infty$, $P\left(\frac{\alpha(\bar{a} - a_L)}{\sigma}\right)$ increases and tends to one for any σ and therefore, because all of the low-ability individuals apply, the number of low-ability individuals receiving benefits increases and tends to $\bar{N}_L$. Simultaneously, we need to have $B \to \infty$ and therefore spending will be tending to ∞. Hence, at some point the full budget will be spent. ■

[7]Proposition 1 do not account for this type of first-best program. As mentioned above, the results in this section restrict attention to program budgets satisfying $R < \bar{B}\left(\bar{N}_L + \bar{N}_H\right)$.

Notice that, when the budget is large, full-separation policies that exhaust the entire budget feature benefits that are higher than $\bar{B}$. Higher benefits tend to attract high-ability individuals, but they can be discouraged from applying by having a high degree of complexity.

To characterize the optimal policy under full separation, we will need the following lemma:

Lemma 2 *Consider $a < \bar{a}$ and $\tilde{P}(\alpha, B) = P(0)$. Under assumption 2,*

1. *a small increase in α increases the number of individuals receiving benefits $N_a(\alpha, \bar{a}, B)$ (even though it reduces the number of applicants)*
2. *a small decrease in B has no effect on the number of individuals receiving benefits $N_a(\alpha, \bar{a}, B)$*
3. *$N_a(\alpha, \bar{a}, B)$ is continuously differentiable in α and B (despite switching from everyone applying to non-full take-up).*

Proof. See the appendix. ■

Assumption 2 guarantees smoothness of the number of beneficiaries when α and B change so as to just stop some people from applying: these are the people with the highest variances and the thin-tail assumption implies that there are not 'many' of them. We can then characterize the optimal full-separation, non-first best program as follows:

Proposition 2 (best policy avoiding type II errors) *Under assumption 2, the best policy implementing full separation when the first-best allocation is not feasible is characterized by $B > \bar{B}$, $\bar{a} = a_H$, $\tilde{P}(\alpha, B) > P(0)$, and $\frac{\partial N_L}{\partial \alpha} = 0$.*

Proof. First, since the first-best allocation is not feasible, a full separation policy that spends all of the budget must have $B > \bar{B}$. By Lemma 1, there exist full separation policies that satisfy the budget constraint.

Second, the best full separation policy involves $\bar{a} = a_H$. To see this, suppose instead that $\bar{a} < a_H$ in the optimum. Then we can increase $\bar{a}$ to a_H, which would imply more low-ability people receiving benefits. Now, we are spending too much money, but we can reduce B until the budget is satisfied (this is possible because initially $B > \bar{B}$ and at $\bar{B}$ not everything is spent). In the new equilibrium, we have $N_L B = R$ and a lower B, so that N_L must be higher, contradicting that $\bar{a} < a_H$ was optimal.

Third, the optimal policy involves $\tilde{P}(\alpha, B) > P(0)$. Conversely, suppose that $\tilde{P}(\alpha, B) = P(0)$. Consider increasing α slightly so that $\tilde{P}(\alpha, B) > P(0)$. By Lemma 2, the number of low-ability recipients increases and spending increases over R. Therefore, we may now reduce B until spending falls to R (we may do so because $B > \bar{B}$ to begin with). We end up with all of the budget spent, lower benefits and therefore more low-ability individuals receiving benefits — a contradiction.

Finally, having established that $B > \bar{B}$, $a = a_H$, and $\tilde{P}(\alpha, B) > P(0)$, the problem is to maximize $N_L(\alpha, a_H, B)$ with respect to α and B, subject to $N_L(\alpha, a_H, B)B = R$. The latter equation can be solved for $B = B(\alpha)$ where $\frac{\partial B}{\partial \alpha} = -\frac{N_L + B\frac{\partial N_L}{\partial B}}{B}$. The problem we solve is now equivalent to maximizing $N_L(\alpha, a_H, B(\alpha))$ with respect to α. The first-condition is $\frac{\partial N_L}{\partial \alpha} + \frac{\partial N_L}{\partial B}\frac{\partial B}{\partial \alpha} = 0$. Substituting for $\frac{\partial B}{\partial \alpha}$ and simplifying yields $\frac{\partial B}{\partial \alpha} N_L = 0$. Inserting this into the original first-order condition $\frac{\partial N_L}{\partial \alpha} + \frac{\partial N_L}{\partial B}\frac{\partial B}{\partial \alpha} = 0$, we obtain $\frac{\partial N_L}{\partial \alpha} = 0$. We are guaranteed that such a point exists because N_L is positive and increasing in α at $\tilde{P}(\alpha, B) = P(0)$ (by Lemma 2), N_L is equal to zero when $\tilde{P}(\alpha, B) = 1$, and $\tilde{P}(\alpha, B)$ itself increases with α (and attains the value of one for a sufficiently high α). ■

This proposition has several implications. First, because $\tilde{P}(\alpha, B) > P(0)$ both kinds of Type I error are made:

Corollary 1 *The best policy that avoids Type II errors involves both Type Ia and Ib errors.*

Although the objective is to maximize the number of low-ability recipients and the government is able to discourage high-ability individuals from applying, the optimal policy is associated with incomplete take up. The reason is that discouraging high-ability individuals from applying makes it impossible to provide benefits to all of the low-ability individuals who do apply. Given that some Type Ib errors are being made, it is always optimal to reduce their number somewhat at the cost of introducing some Type Ia errors.

Second, the optimal policy features benefits that are higher than the minimum required level $\bar{B}$. This is a mechanical result. Given that full separation imposes a restriction on $\bar{a}$ and given that a sufficiently high α discourages applications, the only way to spend all of the budget while retaining full separation is by increasing B.

Third, the optimal full separation policy involves setting complexity α such that it has no effect at the margin on the number of low-ability recipients. This implies that the additional discouragement of low-ability applicants from a higher complexity cost (operating through $\tilde{P}(\alpha, B)$) is exactly offset by a higher probability of receiving benefits conditional on applying, $P\left(\frac{\alpha(\bar{a}-a_L)}{\sigma}\right)$. As we shall see below, this result does not carry over to optimal non-separation policies.

Finally, observe that full-separation policies (whether first-best is feasible or not) may be very costly in terms of the complexity burden that they impose on welfare recipients. As an example, consider the case of risk-neutrality where $\tilde{P}(\alpha, B) = \frac{f(\alpha)}{B}$. At the optimum, we have $\tilde{P}(\alpha, B) \geq P(0)$ and therefore $f(\alpha) \geq P(0)B$. Hence, the cost of complexity consumes at least a fraction $P(0)$ of welfare transfers. Recall that $P(0)$ is the probability that an individual will test below his true ability level. Under the natural assumption that the distribution of tests is symmetric, i.e. $P(0) = \frac{1}{2}$, complexity consumes at least one-half of the income surplus for those who get the benefit. Since some applicants are rejected in the process, aggregate complexity costs may then constitute more than half of the surplus to all welfare applicants.

So far, we have imposed the rigid restriction that the policy maker attempts to keep high-ability individuals from applying. This must be the best policy if one can simultaneously set $B = \bar{B}$ because the number of the low-ability recipients then reaches its theoretical maximum. However, as we have shown, this is possible only if the budget is small enough (Proposition 1). For greater budgets, the best full separation policy requires overpaying benefits (Proposition 2), and therefore it is possible that allowing some high-ability individuals to apply while simultaneously reducing benefits will result in a higher number of low-ability recipients. Indeed, we can show

Lemma 3 *Under assumption 3, we can improve upon the policy characterized in Proposition 2 by increasing $\bar{a}$ slightly.*

Proof. In the appendix. ■

This result follows because a small increase in the eligibility threshold above a_H has only a second-order effect on the number of high-ability recipients who are just becoming eligible while having a first-order effect on the number of low-ability recipients. This allows for reducing the benefit below the level prevailing under the optimal full-separation policy (where $B > \bar{B}$), and therefore financing a higher number of low-ability recipients. Hence,

Corollary 2 (type II errors are optimal) *When the first-best allocation cannot be implemented, the second-best policy always involves non-separation.*

This is an important result. Even though it is possible to discourage high-ability individuals from applying, it is not optimal. The optimal policy will therefore involve both Type I and Type II errors.

The rest of this section will be devoted to characterizing the optimal policy under non-full separation. While Lemma 3 establishes that there exists non-separation programs with $\bar{a} > a_H$ that dominate the best full separation program, we have to consider the possibility that the *optimal* non-separation program is associated with a "stringent" eligibility criterion, i.e. $\bar{a} \leq a_H$. All else equal, a stringent eligibility criterion will of course discourage high-ability applicants from applying, but we can bring them back in by having a low complexity so that $\tilde{P}(\alpha, B) < P(0)$. However, we can show that non-separation policies combining a stringent eligibility criterion with low complexity ($\bar{a} \leq a_H$ and $\tilde{P}(\alpha, B) < P(0)$) are always dominated by non-separation policies that combine a lenient eligibility criterion with high complexity ($\bar{a} > a_H$ and $\tilde{P}(\alpha, B) \geq P(0)$):

Proposition 3 (eligibility criterion is "lenient") *When the first-best allocation is not feasible, setting $\bar{a} \leq a_H$ is never optimal.*

Proof. Lemma 3 implies that, if the first best is not feasible, the full separation policy is not optimal. Therefore, we want to consider non-full separation policies where $\bar{a} \leq a_H$ and $\tilde{P} < P(0)$. Suppose a policy of this kind, denoted by $(\alpha^*, \bar{a}^*, B^*)$, is optimal. Consider then an alternative policy that keeps $B = B^*$, set $\bar{a}$ to satisfy $\max\{\bar{a}^*, a_L\}$, and adjusts α to obtain $\tilde{P} = P(0)$. The number of high-ability applicants drops to zero, whereas all of the low-ability applicants will apply with the probability of receiving benefits increasing for each of them.[8] This change therefore increases the value of the objective function. If this policy results in a reduction in the total number of beneficiaries (note that all of the previous high-ability recipients drop out), it is affordable and therefore it is an improvement — contradiction. Otherwise, if the policy increases the total number of recipients, it is not affordable. If $B^* > \bar{B}$, we can then reduce benefits. Such an adjustment will maintain full separation and if it yields an affordable policy it must be an improvement because full budget will be spent on lower benefits paid to low-ability individuals only. This again contradicts the optimality of the original policy. When reducing benefits to $\bar{B}$ results in a policy that is still unaffordable, that implies that it is possible to spend more than the full budget on a first-best allocation and therefore the first-best allocation can be implemented as in the proof of Proposition 1, thereby contradicting the assumption that first-best allocation is not feasible. ■

The intuition for the result in Proposition 3 is that stringent programs with low complexity are associated with a lot of type Ib errors. In particular, if $\bar{a} \leq a_L$, all low-ability

[8]For the case where $\max\{\bar{a}^*, a_L\} = a_L$, there is a technical qualification due to the fact that the participation constraint (5) is written with strict inequality. Because of this, if $\bar{a} = a_L$ and $\tilde{P} = P(0)$, the low-ability individuals would not apply. To be precise, the government would instead have to set $\bar{a} = a_L + \delta$ where δ can be arbitrarily small, in which case all low-ability individuals would apply.

applicants would face probabilities of receiving benefits below $P(0)$. If instead $a_L < \bar{a} \leq a_H$, low-ability applicants would face probabilities of receiving benefits distributed on $(P(0), 1]$, but the rigor of screening is low and therefore, even though low-ability applicants are formally eligible, the distribution of the Ps is concentrated toward the lower end of the support $P(0)$.

We next demonstrate that the optimal policy changes smoothly from the first-best region to the non-full separation region.

Proposition 4 *Denote by R^* the maximum budget that allows for implementing the first-best allocation and let $(\alpha, \bar{a}, B) = (\alpha^*, a^H, \bar{B})$ be the corresponding optimal policy. Denote by $x(R) = (\alpha(R), \bar{a}(R), B(R))$ the optimal policies as a function of R, $R \geq R^*$. The function $x(R)$ is right-continuous at R^*.*

Proof. In the appendix. ■

Recall the structure of our problem: we maximize the number of low-ability recipients N_L subject to the constraints $(N_L + N_H)B = R$ and $B \geq \bar{B}$. Lemma 2 guarantees that N_L and N_H are both continuously differentiable at $\tilde{P}(\alpha, B) = P(0)$ which is the only point where it is not immediately obvious. Therefore, the maximum satisfies the following first-order conditions (where λ is the Lagrange multiplier associated with the government budget):

$$\frac{\partial N_L}{\partial \alpha} - \lambda \left[\frac{\partial N_L}{\partial \alpha} + \frac{\partial N_H}{\partial \alpha} \right] B = 0, \tag{10}$$

$$\frac{\partial N_L}{\partial \bar{a}} - \lambda \left[\frac{\partial N_L}{\partial \bar{a}} + \frac{\partial N_H}{\partial \bar{a}} \right] B = 0, \tag{11}$$

$$\left[\frac{\partial N_L}{\partial B} - \lambda \left(\frac{\partial N_L}{\partial B} + \frac{\partial N_H}{\partial B} \right) B - \lambda (N_L + N_H) \right] (B - \bar{B}) = 0, \tag{12}$$

where the first bracketed term in equation (12) is non-positive and the second is non-negative. When one considers a restricted problem of selecting α and $\bar{a}$ holding B constant, condition (12) need not hold but equations (10) and (11) remain valid. Consequently, as long as α and $\bar{a}$ are selected optimally given B, we must have:

$$\frac{\partial N_L / \partial \bar{a}}{\partial N_H / \partial \bar{a}} = \frac{\partial N_L / \partial \alpha}{\partial N_H / \partial \alpha} = \frac{\lambda B}{1 - \lambda B}. \tag{13}$$

Neither eligibility criterion $\bar{a}$ nor the intensity of screening/complexity α have a direct revenue cost. Therefore, intuitively, what matters in comparing them is how well each of them screens low- from high-ability individuals. This is summarized by the marginal change in the number of low-ability recipients relative to the marginal change in the number of high-ability recipients. At the optimum, the two instruments screen equally well. It is also straightforward to show that when $\frac{\partial N_L / \partial \bar{a}}{\partial N_H / \partial \bar{a}} > \frac{\partial N_L / \partial \alpha}{\partial N_H \partial \alpha}$, $\bar{a}$ should be increased and/or α reduced, with the opposite implication when the sign of this inequality is reversed.

In general, the effect of α on the number of recipients of each type, $\frac{\partial N_L}{\partial \alpha}$ and $\frac{\partial N_H}{\partial \alpha}$, may be either positive or negative. But because $\frac{\partial N_L / \partial \bar{a}}{\partial N_H / \partial \bar{a}}$ is always positive, any program satisfying eq. (13) must be associated with complexity such that $\frac{\partial N_L}{\partial \alpha}$ and $\frac{\partial N_H}{\partial \alpha}$ have the same sign. In the next section, we present numerical simulations showing that the optimal solution is typically associated with $\frac{\partial N_L}{\partial \alpha}$ and $\frac{\partial N_H}{\partial \alpha}$ being negative.

Before continuing, we state the following very useful identity that links derivatives of the number of recipients with respect to the three instruments (the proof is in the appendix):

$$\frac{\partial N_a}{\partial \alpha} = \frac{\bar{a} - a}{\alpha} \frac{\partial N_a}{\partial \bar{a}} + \frac{\partial \tilde{P}/\partial \alpha}{\partial \tilde{P}/\partial B} \frac{\partial N_a}{\partial B}, \quad \text{where} \quad \frac{\partial \tilde{P}/\partial \alpha}{\partial \tilde{P}/\partial B} = -\frac{e^{\beta B} - 1}{e^{\beta f(\alpha)} - 1} \cdot f'(\alpha) \tag{14}$$

This result follows because all three instruments operate through two margins. First, instruments can affect $\tilde{P}(\alpha, B)$, the minimum acceptable probability of receiving benefits consistent with applying. Second, instruments can affect the maximum realization of the individual error term that results in obtaining benefits, $\alpha(\bar{a} - a)$. Complexity works through both margins, whereas benefits work only through the first one and the eligibility criterion works only through the second one.

We can now show that the government pursues social policies associated with incomplete take up:

Proposition 5 (Type Ia errors are optimal) *Under assumption 2, when the first-best allocation is not feasible, for any value of B, the optimal choice of $\bar{a}$ and α implies $\tilde{P}(\alpha, B) > P(0)$.*

Proof. Conversely, suppose that $\tilde{P}(\alpha, B) \leq P(0)$. Then, we have $\frac{\partial N_L}{\partial B} = 0$ and $\frac{\partial N_H}{\partial B} = 0$ (Lemma 2 shows that this holds at $\tilde{P}(\alpha, B) = P(0)$, while it obviously holds at $\tilde{P}(\alpha, B) < P(0)$). As a consequence, identity (14) becomes $\frac{\partial N_a}{\partial \alpha} = \frac{\bar{a} - a}{\alpha} \frac{\partial N_a}{\partial \bar{a}}$, which implies

$$\frac{\partial N_L / \partial \alpha}{\partial N_H / \partial \alpha} = \frac{\bar{a} - a_L}{\bar{a} - a_H} \frac{\partial N_L / \partial \bar{a}}{\partial N_H / \partial \bar{a}} > \frac{\partial N_L / \partial \bar{a}}{\partial N_H / \partial \bar{a}}$$

That, however, implies that the original policy could not have been optimal because it violates the optimality condition (13) (and, in fact, α should be increased). ■

Corollary 2 and Proposition 5 together implies that, when the budget is not small, social programs feature both type Ia and type II errors. Optimal programs of course also feature type Ib errors because, given that α and $\bar{a} > a_H$ are finite, low-ability applicants face probabilities of receiving benefits $P\left(\frac{\alpha(\bar{a} - a_L)}{\sigma}\right)$ distributed on $\left(\tilde{P}, 1\right)$, and hence face non-zero probabilities of rejection.[9] We have therefore shown that large-budget programs are associated with all three types of classification error as illustrated in Figure 1.

An interesting issue regarding the optimal setting of instruments is the choice of the optimal level of benefits. Given that the government cares only about the number of recipients, it may seem obvious that benefits should be set at the lowest possible level. But recall that the best full-separation policy did not have this property: according to Proposition 2 benefits should be increased above their minimum level. In that context, this was a mechanical result driven by the inability to otherwise spend all of the budget on eligibles. However,

[9] Notice that α cannot increase without bound because in that case, to prevent everybody from dropping out of the program due to prohibitive application costs, B would also have to increase without limit, which requires an unlimited budget, $R = \infty$. The eligibility criterion $\bar{a}$ also cannot increase without bound unless the budget is large enough to give benefits to everyone, i.e. $R \geq \left[\bar{N}_L + \bar{N}_H\right] \bar{B}$, in which case a universal program would be optimal.

notice that benefits also play a screening role by potentially attracting high- and low-ability applicants at different rates. Intuitively, if benefits are sufficiently good at screening, this may warrant increasing them despite their budgetary cost. This possibility is reinforced by the next proposition.

Proposition 6 (B screens better than α and $\bar{a}$) *Suppose that α and $\bar{a}$ are set optimally given B. Then,*

$$\frac{\partial N_L/\partial B}{\partial N_H/\partial B} > \frac{\partial N_L/\partial \bar{a}}{\partial N_H/\partial \bar{a}} = \frac{\partial N_L/\partial \alpha}{\partial N_H/\partial \alpha} = \frac{\lambda B}{1-\lambda B}\,. \tag{15}$$

Proof. Equalities in the statement of the proposition repeat equation (13). To show that the inequality is valid, recall identity (14) and note that $\tilde{P}(\alpha, B)$ does not depend on the type. Therefore,

$$\frac{\frac{\partial N_L}{\partial B}}{\frac{\partial N_H}{\partial B}} = \frac{\frac{\bar{a}-a_L}{\alpha}\frac{\partial N_L}{\partial \bar{a}} - \frac{\partial N_L}{\partial \alpha}}{\frac{\bar{a}-a_H}{\alpha}\frac{\partial N_H}{\partial \bar{a}} - \frac{\partial N_H}{\partial \alpha}} = \frac{\frac{a_H-a_L}{\alpha}\frac{\partial N_L}{\partial \bar{a}}}{\frac{\bar{a}-a_H}{\alpha}\frac{\partial N_H}{\partial \bar{a}} - \frac{\partial N_H}{\partial \alpha}} + \frac{\frac{\bar{a}-a_H}{\alpha}\frac{\partial N_L}{\partial \bar{a}} - \frac{\partial N_L}{\partial \alpha}}{\frac{\bar{a}-a_H}{\alpha}\frac{\partial N_H}{\partial \bar{a}} - \frac{\partial N_H}{\partial \alpha}}$$

The denominator of the first term is equal to $-\frac{\partial \tilde{P}/\partial \alpha}{\partial \tilde{P}/\partial B}$ times $\frac{\partial N_H}{\partial B}$ and it is positive because $\partial \tilde{P}/\partial \alpha > 0$ while $\partial \tilde{P}/\partial B < 0$. Therefore, the first term of the expression is unambiguously positive. When α and $\bar{a}$ are set optimally given B, we have $\frac{\partial N_L/\partial \bar{a}}{\partial N_H/\partial \bar{a}} = \frac{\partial N_L/\partial \alpha}{\partial N_H/\partial \alpha} = \frac{\lambda B}{1-\lambda B}$ and it is straightforward to show that in this case the second term is equal to $\frac{\lambda B}{1-\lambda B}$. Therefore, we have

$$\frac{\frac{\partial N_L}{\partial B}}{\frac{\partial N_H}{\partial B}} = \frac{\frac{a_H-a_L}{\alpha}\frac{\partial N_L}{\partial \bar{a}}}{-\frac{\partial \tilde{P}/\partial \alpha}{\partial \tilde{P}/\partial B}\frac{\partial N_H}{\partial B}} + \frac{\lambda B}{1-\lambda B} > \frac{\lambda B}{1-\lambda B}\,. \tag{16}$$

■

This is an important result: on the margin, benefits are better at screening low- from high-ability individuals than any of the other instruments. This is a global result that holds for any value of B as long as the other instruments (complexity and eligibility) are set optimally. Therefore, the only reason not to increase benefits beyond $\bar{B}$ is the revenue cost. In other words, the question is whether the advantage from using benefits to screen is big enough to compensate for the extra revenue cost. From the screening point of view, increasing benefits is preferred to the other two instruments.

When $\bar{a}$ and α are set optimally, benefits should be increased from some level B if

$$\frac{\partial N_L}{\partial B} - \lambda B\left(\frac{\partial N_L}{\partial B} + \frac{\partial N_H}{\partial B} + \frac{N_L+N_H}{B}\right) > 0 \quad \Rightarrow \quad \frac{\frac{\partial N_L}{\partial B}}{\frac{\partial N_H}{\partial B}} > \frac{\lambda B}{1-\lambda B} + \frac{\lambda B}{1-\lambda B}\frac{N_L+N_H}{B\frac{\partial N_H}{\partial B}}$$

Recall equation (13): $\frac{\lambda B}{1-\lambda B}$ reflects the optimal extent of screening performed by the other instruments. Benefits should be used beyond their minimal level only if they are sufficiently better than the other instruments at screening by a factor identified in the last term — this is a correction for the budgetary cost of increasing benefits. It is difficult for benefits to satisfy this condition if the other instruments are already good at screening ($\frac{\lambda B}{1-\lambda B}$ is high), when there are a lot of individuals whose benefits will have to be increased ($N_L + N_H$ is high) and when $B\frac{\partial N_H}{\partial B}$ is small. Substituting for $\frac{\partial N_L/\partial B}{\partial N_H/\partial B}$ using the equality in (16) yields

$$\frac{a_H - a_L}{-\alpha\frac{\partial \tilde{P}/\partial \alpha}{\partial \tilde{P}/\partial B}}\frac{\partial N_L}{\partial \bar{a}} > \frac{\lambda B}{1-\lambda B}\frac{N_L+N_H}{B}.$$

Recalling that $\frac{\partial N_L/\partial \bar{a}}{\partial N_H/\partial \bar{a}} = \frac{\lambda B}{1-\lambda B}$, we can rewrite this to

$$\frac{a_H - a_L}{\alpha}\frac{\partial N_H}{\partial \bar{a}} > -\frac{\partial \tilde{P}/\partial \alpha}{\partial \tilde{P}/\partial B}\frac{N_L + N_H}{B} = -\frac{\partial \tilde{P}/\partial \alpha}{\partial \tilde{P}/\partial B}\frac{R}{B^2}, \tag{17}$$

where the last equality uses the budget identity $R = B(N_L + N_H)$. It is optimal to increase B beyond $\bar{B}$ if this condition holds when evaluated at $\bar{B}$ and the optimal α and $\bar{a}$ (at $\bar{B}$). How should we interpret this condition? B should be used if changes in eligibility $\bar{a}$ bring too many high-ability individuals, where the inequality gives the specific meaning to "too many". Alternatively, note from identity (14) that $\frac{\bar{a}-a_L}{\alpha}\frac{\partial N_H}{\partial \bar{a}}$ is the effect of α on the number of high-ability individuals receiving benefits while holding the number of applicants constant (i.e. holding $\tilde{P}$ constant). Thus, the same condition can be expressed in terms of extra complexity bringing in "too many" high-ability individuals. The following proposition characterizes the optimal choice of benefits.

Proposition 7 (optimal benefits) *Suppose that the first-best allocation is not feasible,* $R \geq R^*$. *Denote by* N_L^* *the number of low-ability recipients under the best full separation policy identified by Proposition 2. Let* $\bar{a}^* = \inf_{\bar{a}} \left\{ \max_{\alpha} N_L(\alpha, \bar{a}, \bar{B}) \geq N_L^* \right\}$, $a_H < \bar{a}^* < \infty$.[10] *Then,*

1. *For* R *sufficiently close to* R^*, *setting* $B = \bar{B}$ *is optimal.*
2. *For* R *sufficiently large, setting* $B = \bar{B}$ *is optimal.*
3. *A sufficient condition for* $B > \bar{B}$ *is given by*

$$\frac{\bar{a}^* - a_H}{\bar{a}^* - a_L} G_L^{-1}(N_L^*) \geq G_H^{-1}\left(\frac{1}{P(0)}\left(\frac{R}{\bar{B}} - N_L^*\right)\right). \tag{18}$$

Proof. In the appendix. ■

The intuition for the first result is straightforward. When the budget is small (but large enough to make first best infeasible), the eligibility threshold $\bar{a}$ will be very close to a_H. As we cross a_H with $\bar{a}$, initially we are still mostly providing benefits to low-ability individuals (on the margin, the share of high-ability recipients is close to zero when $\bar{a}$ is close to a_H). Given the presence of such a good instrument that does not have a direct revenue cost, it must dominate any instrument that does have a revenue cost (such as B). That is, as the number of high-ability applicants is initially small, any screening benefits of using high benefits have to be dominated by the costly nature of this instrument. The second part is also intuitive: as the budget size increases, the number of individuals served increases as well and therefore increasing benefits becomes more costly.

Part 3 is the most interesting. It gives the sufficient condition for $B > \bar{B}$. Moreover, note that this condition can be satisfied by varying G_H without affecting any of the other variables in eq. (18). In particular, the definitions of N_L^* and $\bar{a}^*$ are based solely on the

[10] We know that $\bar{a}^* > a_H$ because, from the proof of Proposition 3, if the first-best allocation is not feasible, any non-full separation policy with $\bar{a} \leq a_H$ is dominated by a full separation policy and therefore also dominated by the best full separation policy N_L^*.

low-ability distribution and parameters — they do not depend on G_H. Furthermore, the argument of G_H^{-1} on the right-hand side depends on constants R, $\bar{B}$ and again on N_L^* so that it is independent of G_H. Finally, note that both the left-hand side and the argument of G_H^{-1} are positive due to the fact that $\bar{a}^* > a_H$ and N_L^* delivers fewer low-ability recipients than $R/\bar{B}$ which is what the first-best policy would deliver. Thus, given parameters and low-ability distribution, we will be able to find some distributions G_H that satisfy the condition identified in the above proposition. All that is required is selecting the distribution so that there are more than $\frac{1}{P(0)}(R/\bar{B} - N_L^*)$ high-ability individuals with σ smaller than $\frac{\bar{a}^*-a_H}{\bar{a}^*-a_L}G_L^{-1}(N_L^*)$. This is a requirement imposed on G_H at a particular strictly positive point.[11] The corollary below is a consequence of this reasoning and it highlights that the case $B > \bar{B}$ cannot be dismissed as being irrelevant because it will apply when the number of high-ability applicants is sufficiently large.

Corollary 3 *Fix the parameters of the problem other than the high-ability distribution. Select some distribution of high-ability individuals $G_H^0(\sigma)$ (with the corresponding number of high-ability individuals N_H^0) and consider a class of distributions $G_H^\eta(\sigma) = \eta G_H^0(\sigma)$ (with the corresponding number of high-ability individuals ηN_H^0). For high enough η, setting $B > \bar{B}$ is optimal.*

Proof. For sufficiently high η,

$$G_H^\eta\left(\frac{\bar{a}^* - a_H}{\bar{a}^* - a_L}G_L^{-1}\left(N_L^*\right)\right) = \eta G_H^0\left(\frac{\bar{a}^* - a_H}{\bar{a}^* - a_L}G_L^{-1}\left(N_L^*\right)\right) \geq \frac{1}{P\left(0\right)}\left(\frac{R}{\bar{B}} - N^*\right).$$

and this condition is equivalent to the inequality in Part 3 of Proposition 7. ■

4 Numerical Simulations

To establish a benchmark simulation, we set ability levels at $a_L = 1$, $a_H = 2$, the number of low- and high-ability individuals at $\bar{N}_L = 1000$, $\bar{N}_H = 1000$, and the target minimum benefit at $\bar{B} = 1$. The distribution of the noise term ε/σ is assumed to be normal (so that $P\left(0\right) = \frac{1}{2}$), and the precision of measured skill σ is assumed to be log-normal with a mean and variance equal to 1. Notice that the log-normal distribution satisfies Assumptions 2 and 3. The coefficient of absolute risk aversion β is set equal to 2. Finally, to solve the model numerically, we specify the functional form for the complexity cost function as $f\left(\alpha\right) = c_0 + c_1 \cdot \alpha^{c_2}$. In the case of 'no complexity', we have $\alpha = 0$ and hence $f\left(0\right) = c_0$, so that c_0 captures non-complexity related costs (such as stigma or other fixed costs) associated with participating in a social program. In the benchmark simulation, we set $c_0 = 0$, $c_1 = 0.5$, and $c_2 = 1$. This is the calibration underlying illustration in Figure 1 and discussed in Section 3.2.

Figures 2-5 shows comparative statics with respect to budget size, ability distribution, complexity costs, and 'stigma' (c_0). Each figure consists of six panels: eligibility criterion $\bar{a}$ in Panel A, screening intensity α and benefit B in Panel B, complexity costs as a share of benefits $f\left(\alpha\right)/B$ in Panel C, the elasticity of low-ability recipients N_L with respect to each of

[11] In particular, there is no restriction imposed on the properties of $G_H(.)$ around zero so that we can pick a distribution satisfying Assumption 3.

the three instruments in Panel D, and the share of individuals (L- and H-types, respectively) applying for benefits ('take-up rate') and the share of individuals being awarded benefits in Panels E and F. Figure 2 shows numerical results as a function of the program budget R based on the benchmark calibration just described. The horizontal axis in the figure starts at a budget size equal to R^* — the largest budget where first best is feasible ($R^* \simeq 810$ in the benchmark) — and then increases the budget toward $\bar{R} \equiv \bar{B}\left(\bar{N}_L + \bar{N}_H\right)$ — the budget where a universal benefit giving $\bar{B}$ to everyone becomes feasible ($\bar{R} = 2000$ in the benchmark). The bold black curve in Figure 2 shows the benchmark results, whereas the other curves show results for alternative ability distributions by changing the number of high-ability individuals relative to low-ability individuals, $\bar{N}_H/\bar{N}_L$ (keeping $\bar{N}_L = 1000$). In particular, the think black curve is associated with $\bar{N}_H/\bar{N}_L = 1.5$, and the gray curve is associated with $\bar{N}_H/\bar{N}_L = 2$. The vertical dashed line indicates $R^* + 500$, which is the level of budget that we are going to rely on in the subsequent experiments.

In the benchmark where $\bar{N}_H/\bar{N}_L = 1$, the figure shows that benefits are always kept at their target minimum. The eligibility criterion $\bar{a}$ starts at 2 where high-ability individuals are *just* eligible. It then increases monotonically with the program budget R and tends to infinity as R converges to $\bar{R}$. In other words, a larger program budget (at a given $\bar{B}$) is always associated with more universalism in benefit provision. Interestingly, any program that is not universal is associated with a substantial degree of complexity. In the simulations reported in Figure 2, complexity costs as a proportion of the (potential) benefit vary between 35% and 65%. Moreover, complexity α is increasing in the size of the budget. In other words, as a program becomes better funded, it is optimal to spend the extra funds making the program more lenient (increasing $\bar{a}$) rather than slacking on the rigor of the screening process by reducing α. Although a high degree of complexity may in itself discourage some low-ability individuals from applying, the combination of a high α and a high $\bar{a}$ allows the program to keep a high number of L-applicants (few type Ia errors) and identify them with high precision (few type Ib errors). Another way to gauge the degree of complexity is to consider the elasticity in equilibrium of the number of L-recipients with respect to α shown in Panel D. What we see is that, at any budget size, complexity is taken to a point where it has a negative effect on the number of L-recipients at the margin. As discussed above, this is entirely consistent with an optimum given that the marginal effect on H-recipients is also negative. As expected, as the size of the budget increases, all instruments become less effective in increasing participation (elasticities get closer to zero).

Panels E and F shows the extent of classification errors as a function of budget size. The type II error rate (the share of H-types receiving benefits) is initially zero, increases as a function of R, and converges to 1 as $R \to \bar{R}$. The basic intuition is that, for a poorly funded program, classification errors of type II is a luxury that it cannot afford, whereas a well-funded program can afford making type II errors in order to reduce the amount of type I errors. The type I error rate (i.e., type Ia + type Ib) equals 1 minus the share of L-types receiving benefits. This rate is decreasing in R and tends to zero as $R \to \bar{R}$. The type Ib error rate equals the take-up rate minus the share of L-individuals receiving benefits, and is also decreasing with R. However, the type Ia error rate (1 - take-up rate) is not monotonically decreasing in R, which may seem surprising. Although the type Ia error rate does decrease in R at large enough Rs, at lower budget levels the government find it optimal to trade-off type Ib errors for type Ia errors. This effect occurs because a larger budget and the associated relaxation of the eligibility criterion invite more H-applicants into the program, which makes testing more important. The implied increase in transaction costs deter some L-applicants from applying but substantially lower the amount of type Ib errors.

Figure 2 also explores our theoretical finding that $B > \bar{B}$ may be optimal. According to Proposition 7, this may occur when the budget R is neither too close to R^*, nor too close to $\bar{R}$. We showed that the requirement for $B > \bar{B}$ in this intermediate budget range is that the number of high-ability applicants is sufficiently large. The numerical simulations confirm this result. As we increase $\bar{N}_H/\bar{N}_L$ so as to get more H-applicants, it becomes optimal to increase benefits at intermediate budget sizes. The range over which $B > \bar{B}$ is optimal and the optimal size of B in this interval are both monotonically increasing in $\bar{N}_H/\bar{N}_L$.

The next three figures focus on a fixed budget size, $R = R^* + 500$, and then vary the three parameters of the complexity cost function (c_0, c_1, and c_2). Figure 3 considers the implications of increasing the utility cost of complexity by increasing c_1 from 0 to 2. Notice that a higher c_1 implies both a higher level of the complexity cost $f(\alpha)$ and a higher marginal complexity cost $f'(\alpha)$, both of which influence the effect of α on the number of recipients $\frac{\partial N_a}{\partial \alpha}$ (cf. eq. 14). It is intuitive that an increase in c_1 makes α less effective as an instrument, which is confirmed by the figure showing α as a strongly declining function of c_1. In fact, the reduction of α in response to a higher c_1 is strong enough that the total complexity costs as a share of benefits, $f(\alpha)/B$, is reduced. At the same time, the government makes the program more lenient by increasing $\bar{a}$. Hence, as c_1 increases, the optimal social program moves from a highly complex, highly targeted program to a less complex and less discriminating program, with increased likelihood of all types of errors.

Figure 4 shows the effects of increasing c_0 (pure stigma or other fixed costs of applying). A higher c_0 has a non-monotonic effect on the degree of complexity. At low levels of c_0, increasing stigma leads to increased complexity. The intuition for this result is as follows. Higher stigma results in fewer applicants and those who apply have a high likelihood of being awarded benefits. It then becomes harder to discriminate between low- and high-ability applicants, which makes it necessary to increase the rigor of testing. As the level of stigma becomes large, convincing anyone to apply becomes very hard and this requires a compensating reduction in complexity α (to restrict the reduction in take up) along with a lenient eligibility criterion (to ensure a high probability of receiving benefits for those who do apply). While stigma has a strong negative impact on equilibrium take up, it has an indirect desirable effect on the amount of type Ib errors. The reduction in type Ib errors may seem beneficial by itself, but it is the consequence of the difficulty of screening in the presence of stigma: at very high levels of stigma, virtually all applicants are approved. Moreover, higher stigma increases the number of high-ability applicants (via the relaxation of the eligibility criterion) and thereby creates more type II errors. Overall, stigma is bad for targeting efficiency, because it reduces the number of deserving recipients *and* increases the number of undeserving recipients.

Figure 5 shows the effect of introducing convexity in $f(\alpha)$ by raising c_2 above 1. Notice that, besides introducing curvature in $f(\alpha)$, the parameter c_2 also influences $f(\alpha)$ and $f'(\alpha)$ and that the effects are ambiguous depending on whether α is greater than or less than 1. It is therefore not a priori clear how c_2 affects program design. The results in the figure shows that the effect on equilibrium complexity α is ambiguous, whereas the effect on the eligibility criterion $\bar{a}$ is negative (other numerical configurations confirm this qualitative effect).

5 Interpretations, Applications, and Extensions

This paper stresses the importance of transaction costs and imperfect information for incomplete take up in public programs. The standard economic model of welfare program participation, starting with Moffitt (1983), focuses instead on welfare stigma as the reason for incomplete take up. Moffitt's model distinguishes between a flat component of stigma — a fixed cost associated with program participation — and a variable component depending on the size of the benefit. Under flat stigma, an increase in the size of the welfare benefit imply a higher take-up rate, whereas under variable stigma there is no such effect of higher benefits on take-up. Moffitt showed empirically that higher benefits are indeed associated with higher take-up, consistent with the presence of flat stigma. However, his results are also consistent with the presence of other fixed transaction costs from program participation such as those arising from the complexity and bureaucracy of the application process. Indeed, a large number of empirical studies have documented that complexity and hassle constitute important barriers to program enrollment. Moreover, the positive effect of benefits on take up is also consistent with the presence of imperfect information about eligibility. When benefits are higher, it is more attractive for the imperfectly informed individuals to enter the lottery for welfare benefits, thereby giving rise to a higher take-up rate.

While complexity and stigma are in many ways consistent and complementary explanations for low take up, they are also different in a very fundamental way: one is a direct policy instrument and the other is not. Although there may be ways for policy to influence stigma, the effects are indirect and involve a great deal of uncertainty.[12] If stigma is the ultimate reason for low take up, perhaps all we can do to increase take up is to make the programs more generous by increasing benefits or relaxing eligibility criteria. The problem is that this would make the programs more attractive to the undeserving and therefore increase the amount of type II errors. On the other hand, if the complexity of welfare programs is an important determinant of take up, it follows that complexity is just as important an instrument as the size of the benefit and the stringency of the eligibility criterion. It also

[12]For example, stigma may be affected by the terminology of the program (say, whether it is called welfare or a tax credit), the way the program is advertised, whether or not administrative procedures are demeaning to the applicant, the possibilities of applying by mail or through the Internet as opposed to going to a program office, whether the transfer is paid in cash or in kind (the latter being more conspicuous), and so on. While these aspects of policy making may affect the extent of stigma associated with a given program, we should bear in mind that stigma reflects individual *preferences* for receiving a social benefit and is therefore not a policy instrument per se. Our understanding of the effect of different policies on the preferences for social benefits is still lacking.

follows that we should view high complexity and incomplete take-up as equilibrium outcomes of policy making under imperfect information, instead of simply a flaw of practical program design that calls for remedial policy action. Hence, our paper presents a model to explain the existence of complexity and incomplete take-up as equilibrium outcomes. It has long been recognized in the theoretical literature that the appropriate design of welfare programs reduces their cost by limiting take-up by non-deserving recipients. Our model fills the gap in this literature by recognizing the trade-off due to the fact that the same policies adversely affect take up by those who are the intended recipients.

More generally, our paper represents a first attempt to incorporate administrative complexity as a choice variable in policy analysis. Although several authors have suggested that complexity is an important aspect of policy design, for example in the context of tax policy (Slemrod, 1990; Slemrod and Yitzhaki, 2002), little has been done in terms of actual modeling. We have emphasized the application to the design of transfer programs and the take up of social benefits, but our model can also be applied to the analysis of tax policy and tax evasion. This includes cash benefits which are provided through the tax system such as the Earned Income Tax Credit (EITC) and the Child Tax Credit, and it may include tax deductions and exemptions more generally. As elaborated by Holtzblatt and McCubbin (2003), tax rules affecting low-income filers involve a great deal of complexity. A prima facie evidence of the difficulties in dealing with the tax code is the high reliance on tax preparers even among low-income tax payers despite the significant fees charged for such services (Kopczuk and Pop-Eleches, 2007). It is conceivable that this complexity of the tax code reflects the kind of trade-off between type I and type II errors studied in this paper.

Our model and results are consistent with several features of observed social policy and take-up behavior within the United States. First, the model may be able to shed light on the observation of large differences in complexity and participation rates across different programs. We find that, in equilibrium, program characteristics and take up can vary substantially depending on a number of key variables. This includes the size of the program budget (R), the size of the minimum benefit ($\bar{B}$), the distribution of true skills (a_L, a_H, $\bar{N}_L$, and $\bar{N}_H$), the distribution of estimated skills (distribution of σ, $G_a(.)$), and the size and structure of complexity costs (properties of $f(\alpha)$). In other words, different types of social programs will be designed differently and be associated with different take-up rates because they serve different populations, because they involve different kinds of eligibility tests, and because of differences in the size and generosity of programs. Other things being equal, we have seen that programs with a small budget relative to the target benefit, i.e. a small $R/\bar{B}$, are predicted to have a stringent eligibility rule, high complexity, a high occurrence of type I errors, and no or very few type II errors. This may perhaps reflect a program such as the Special Supplemental Nutrition Program for Women, Infants and Children (WIC) in the U.S. For larger programs, the model predicts that eligibility rules are more lenient while complexity is still high, and that both types of classification error occur. This would seem to reflect a program like the Disability Insurance (DI) program, which is a large program involving complex and rigorous testing, and where both type I and type II errors occur (Benítez-Silva et al., 2004). Finally, our model predicts that very large programs will feature universal benefits and no complexity. This fits with the Medicare program — by far the largest transfer program in the United States — which is a universal program (conditional on being old) with a very low degree of complexity (default coverage) and close to full take up.

Another potential difference across programs is the degree of stigma. Our model includes the possibility of pure stigma as a constant in the complexity cost function, and we explored numerically the implications of stigma for program design and outcomes. We find that high-stigma programs have a low incidence of type Ib errors and low take-up rates. This pattern arguably fits the pattern observed in practice: welfare programs tend to be associated with relatively high stigma, imperfect take up and low false rejection rates, whereas programs such as DI have low stigma, high take up and high false rejection rates.

Second, the model is consistent with the occurrence of different take-up behavior across equally eligible individuals resulting from heterogeneity in the observability of skills (the variance of the error term). In the model equilibrium, those with low variances face high acceptance rates in the application process making it optimal for them to apply for benefits, whereas equally eligible individuals with high variances face high risks of being rejected by program administrators and hence find it not optimal to apply. The latter group of individuals are those who, despite that they are truly deserving of the benefit, test with a high degree of uncertainty and may easily fail to live up to the requirements of the program. Language barriers, unfamiliarity with the administrative procedures, inability to understand the formal requirements of the test, etc., would all contribute to creating more uncertainty in the test. This is consistent with the large amount of evidence showing considerable variation in program participation across eligible individuals depending on race, immigration status, fluency in English, age, education, gender, and family status (Borjas and Hilton, 1996; Currie, 2000, 2003; Heckman and Smith, 2004; Duggan and Kearney, 2005). For example, our model can account for why newly arrived immigrants are characterized by a lower take-up of social benefits, conditional on being eligible, than immigrants who are more assimilated into society (Borjas and Hilton (1996)). Our results may also explain why some ethnic groups (such as Hispanics and Asians) have lower take-up than others (such as African Americans) as shown by Currie (2003).

Third, we find that complexity and administrative costs can be very high in equilibrium, especially for governments eager to avoid giving money to the 'undeserving'. Under certain simplifying assumptions — low risk aversion and the noise of observed skill being distributed with zero median — we find that transaction costs constitute more than 50 percent of the amount of benefits granted. In the more realistic case of risk averse agents, transaction costs as a share of total benefits are smaller than 50 percent but are nevertheless substantial.

Our model abstracts from several aspects that we would like to discuss briefly. First, our model ignores the costs of administering a social program. In practice, there will be administrative costs from the processing of applications due to the paper work involved, the time spent by administrators and specialists for interviewing and testing, etc. Realistically, these administrative costs depend positively on the amount of complexity and on the total number of applications that have to be processed. A positive relationship between complexity and administrative costs obviously makes complexity less effective as a policy instrument. On the other hand, administrative costs that are increasing in the number of applications favor policies capable of increasing the number of deserving recipients with a relatively small (or no) accompanying increase in the number of applicants. This tends to improve the effectiveness of complexity as a policy instrument, because it identifies the truly eligible applicants with more precision while making it more costly for would-be applicants to claim the benefit. Given our result that benefits are particularly good at screening, it may also increase the likelihood that higher benefits are paid in equilibrium.

Second, in our model, policy makers do not concern themselves with the negative effect of complexity and hassle on individuals' utility. As in the well-known Besley and Coate (1992, 1995) model, we assume that politicians care only about designing an efficient system of income maintenance. Our paper should therefore be seen as being positive rather than normative: it presents a model to explain and understand the behavior of imperfectly informed policy makers engaging in poverty alleviation, not a model to identify the most desirable policy from the point of view of a *welfarist* social planner. If politicians cared about utility instead of income, it may seem obvious that complexity would be a less effective instrument. However, even in a model of social welfare maximization, given that complexity is modeled as a byproduct of efforts to screen between deserving and undeserving applicants, it continues to be associated with some desirable effects. For example, a situation with no complexity $\alpha = 0$ — corresponding to program administrators flipping a coin to determine eligibility — is an unlikely candidate for the program chosen by a social planner interested in redistributing money from rich to poor. Second, complexity simultaneously serves as an ordeal for program applicants. As shown by Nichols and Zeckhauser (1982) and others, ordeals may be desirable in their own right provided that (i) the utility gain from transfers is lower for the undeserving and/or (ii) the utility cost of the ordeal is higher for the non-deserving.

Third, an important technical assumption in the model is constant absolute risk aversion. This assumption guarantees that the minimum acceptable odds in the benefit award process is the same for low- and high-ability individuals, and this is the only purpose for making it. An assumption of increasing absolute risk aversion would imply that high-ability individuals require better odds in order to apply. This would make it feasible to always implement the first-best outcome by pushing complexity to a level where the minimum odds acceptable to high-ability individuals are above one, while the minimum odds acceptable to low-ability individuals remain below one. A more realistic assumption of decreasing absolute risk aversion would not allow for such a counter-intuitive outcome, but it would make analysis substantially more complicated.

We believe that the central implication of constant absolute risk aversion in the context of the model has an intuitive and realistic economic content: it implies that, given odds at which some low-ability individuals apply, we can always find a high-ability individual who would also apply given the same odds. We believe that this is a realistic description of the real world, although it may reflect dimensions of heterogeneity not incorporated in our model. In our view, a generalization of the model to non-CARA preferences should preserve this central property by adding additional dimensions of heterogeneity.

Finally, a possibility mentioned by Bertrand et al. (2004) and Currie (2004) is that non-participation in social programs may be explained, in part, by individuals being boundedly rational. While we have not yet discussed psychological reasons for low take-up, our framework may be reinterpreted to capture a form of bounded rationality. In this interpretation, the parameter σ (precision of observed skill) reflects the degree of irrationality. An individual with a high σ face a lot of uncertainty when being tested, even if he is truly eligible, because he tends to make mistakes, say stupid things at the interview, appear to be lying when he is really not, and so on. On the other hand, low-σ individuals do everything right and respond well at the interview so as to give a very precise test. In the model equilibrium, the individuals who are not taking up benefits, despite being eligible, are the least rational ones (those with a high σ). This interpretation of our model has the flavor of the Luce (1959) approach to bounded rationality in which individuals make random decision errors and outcomes are the result of a probabilistic process. Clearly, there are other forms of bounded rationality such as present bias and framing effects that may also be relevant to the take-up of social benefits. So far, research on the effects and design of public policy under bounded rationality is relatively unexplored, and is an interesting topic for future research.

A Proofs

Proof of Lemma 2. The total number of individuals of ability a receiving benefits is given by

$$N_a(\bar{a}, \alpha, B) = \int_0^{\bar{\sigma}_a(\alpha,\bar{a},B)} P\left(\frac{\alpha(\bar{a}-a)}{\sigma}\right) dG_a(\sigma) \tag{A.1}$$

and we want to evaluate $\frac{\partial N_a}{\partial \alpha}$ and $\frac{\partial N_a}{\partial B}$ when $\bar{a} > a$ and $\tilde{P}(\alpha, B) = P(0)$. Begin with the first of these. Differentiating explicitly yields

$$\frac{\partial N_a}{\partial \alpha} = \int_0^{\bar{\sigma}_a(\alpha,\bar{a},B)} p(\cdot)\frac{\bar{a}-a}{\sigma} dG_a(\sigma) + \frac{\partial \bar{\sigma}_a}{\partial \alpha}\tilde{P}(\alpha, B) g_a(\bar{\sigma}_a)\,. \tag{A.2}$$

The first term is unambiguously positive. We will show that the second term is zero when $\bar{a} > a$ and $\tilde{P}(\alpha, B) = P(0)$. Evaluating $\frac{\partial \bar{\sigma}_a}{\partial \alpha}$ yields

$$\frac{\partial \bar{\sigma}_a}{\partial \alpha} = \frac{(\bar{a}-a)P^{-1} - \alpha(\bar{a}-a)\left(p\left(\frac{\alpha(\bar{a}-a)}{\bar{\sigma}_a}\right)\right)^{-1}\frac{\partial \tilde{P}}{\partial \alpha}}{\left(P^{-1}\left(\tilde{P}(\alpha,B)\right)\right)^2} = \frac{\bar{\sigma}_a}{\alpha} - \frac{\bar{\sigma}_a^2}{\alpha(\bar{a}-a)}\left(p\left(\frac{\alpha(\bar{a}-a)}{\bar{\sigma}_a}\right)\right)^{-1}\frac{\partial \tilde{P}}{\partial \alpha}\,. \tag{A.3}$$

Note that $\lim_{\tilde{P}(\alpha,B)\to P(0)} \bar{\sigma}_a = \infty$: as the threshold probability of receiving benefits approaches $P(0)$, the number of individuals not applying goes to zero. All other terms in expression (A.3): α, $\bar{a}-a$, $\frac{\partial \tilde{P}}{\partial \alpha}$ and $p(\cdot)$ are positive and finite (the argument of $p(\cdot)$ goes to 0 as $\bar{\sigma}_a$ goes to infinity and density at 0 is positive). Consequently, as we approach $P(0)$ with $\tilde{P}$, expression (A.3) tends to $-\infty$ at the rate of $\bar{\sigma}^2$. Consequently, the behavior of the second term of (A.2) depends on the behavior of $\bar{\sigma}_a^2 g(\bar{\sigma}_a)$ and, by assumption 2, $\lim_{\bar{\sigma}_a\to\infty} \bar{\sigma}_a^2 g(\bar{\sigma}_a) = 0$.

Now consider $\frac{\partial N_a}{\partial B}$. It is equal to $\frac{\partial \bar{\sigma}_a}{\partial B}\tilde{P}(\alpha, B)g_a(\bar{\sigma}_a)$ and we have $\frac{\partial \bar{\sigma}_a}{\partial B} = -\frac{\alpha(\bar{a}-a)}{(P^{-1}(\tilde{P}))^2}\frac{dP^{-1}(\tilde{P})}{dB} = -\bar{\sigma}_a^2 \frac{1}{\alpha(\bar{a}-a)}\frac{dP^{-1}(\tilde{P})}{dB}$. It is straightforward to show as before that all terms but $\bar{\sigma}_a$ and $g(\bar{\sigma}_a)$ are bounded away from zero and infinity. Therefore $\frac{\partial N_a}{\partial B}$ behaves as $g(\bar{\sigma}_a)\bar{\sigma}_a^2$ and thus it is zero in the limit by assumption 2.

Finally, the second term of (A.2) is uniformly zero when $\tilde{P}(\alpha, B) < P(0)$ and the first term is continuous. Therefore, the whole expression in (A.2) is continuous, which proves the third part of the lemma. Similarly, $\frac{\partial N_a}{\partial B}$ is uniformly zero when $\tilde{P}(\alpha, B) < P(0)$, so that it is also continuous at $\tilde{P}(\alpha, B) = P(0)$. ■

Proof of Lemma 3. Denote by (α^*, a_H, B^*) the best policy under full-separation characterized in Proposition 2, and consider increasing $\bar{a}$ above a_H. We will show first that the right-derivative of N_H with respect to $\bar{a}$ is equal to zero at (α^*, a_H, B^*): $\frac{\partial N_H(\alpha^*, a_H, B^*)}{\partial \bar{a}_+} = 0$. Differentiating (A.1) with respect to $\bar{a}$ yields

$$\frac{\partial N_a}{\partial \bar{a}} = \int_0^{\bar{\sigma}_a(\alpha,\bar{a},B)} p\left(\frac{\alpha(\bar{a}-a)}{\sigma}\right)\frac{\alpha}{\sigma} g_a(\sigma) d\sigma + \frac{\partial \bar{\sigma}_a}{\partial \bar{a}}\tilde{P}(\alpha, B) g_a(\bar{\sigma}_a), \tag{A.4}$$

such that

$$\frac{\partial N_H(\alpha^*, a_H, B^*)}{\partial \bar{a}_+} = \lim_{\substack{\bar{a}\to a_H \\ \bar{a}>a_H}} \left\{ \int_0^{\bar{\sigma}_H(\alpha^*,\bar{a},B^*)} p\left(\cdot\right)\frac{\alpha^*}{\sigma} g_H(\sigma) d\sigma + \frac{\partial \bar{\sigma}_H}{\partial \bar{a}_+}\tilde{P}(\alpha^*, B^*) g_H\left(\bar{\sigma}_H(\alpha^*, \bar{a}, B^*)\right) \right\}.$$

Note that $\lim_{\bar{a}\to a_H} \bar{\sigma}_H = 0$, and therefore the second term is zero: we have $g(0) = 0$ while the other components $\tilde{P}$ and $\frac{\partial \bar{\sigma}_H}{\partial \bar{a}_+} = \frac{\alpha}{P^{-1}(\cdot)}$ tend to finite limits (the limit of $P^{-1}(\cdot)$ is positive because we are considering a point with $\tilde{P} > P(0)$). In the first term, the integrand $p\left(\cdot\right)\frac{\alpha^*}{\sigma} g_H(\sigma)$ is bounded from above in the neighborhood of $\sigma = 0$ by assumption 3 and because $p(\cdot)$ is bounded from above. Therefore, the first term tends to zero.

On the other hand, for the low-ability individuals we have that $a_L < \bar{a}$, $\bar{\sigma}_L > 0$, and $\frac{\partial \bar{\sigma}_L}{\partial \bar{a}} > 0$, so that the derivative $\frac{\partial N_L}{\partial \bar{a}}$ given by A.4 is strictly positive.

Hence, starting at the best full-separation policy (α^*, a_H, B^*), we can increase the threshold $\bar{a}$ slightly above a_H so as to give benefits to more low-ability people, while bringing in only an infinitesimal number of high-ability people. We would then be spending too much money, but we can reduce B below B^* until the revenue constraint is satisfied. At this new equilibrium, since B is lower, the total number of recipients, $N_L + N_H$, is higher. Moreover, since the number of high-ability recipients is infinitesimal, the number of low-ability recipients is higher than before. ■

Proof of proposition 4. Consider $R = R^* + \varepsilon$. Denote the optimal policy at R^* by $(\alpha^*, a^H, \bar{B})$ and note that it involves $\tilde{P} > P(0)$ (if $\tilde{P} = P(0)$, increasing the number of low-ability recipients could be increased by increasing α by Lemma 2 and therefore implement the first-best for an even greater budget). To simplify notation, denote the optimal policy given ε by $(\alpha(\varepsilon), \bar{a}(\varepsilon), B(\varepsilon))$.

We will show first that $\lim_{\varepsilon \to 0} \bar{\sigma}_H = 0$. To see that note that the proof of Proposition 2 implies that we can achieve $\frac{R^*+\varepsilon}{B} > N_L > \frac{R^*}{B}$ by sticking to the full separation policies. Note that we must have $N_H < \frac{\varepsilon}{B}$ because otherwise $N_L \geq \frac{R^*}{B}$ and the optimal non-separation policy would be preferred. By definition $N_H = \int_0^{\bar{\sigma}_H} P\left(\frac{\alpha(\bar{a}-a_H)}{\sigma_H}\right) dG_H(\sigma_H)$. We know that $P\left(\frac{\alpha(\bar{a}-a_H)}{\sigma_H}\right) > P(0)$ for everyone because $\bar{a} > a_H$. Therefore, $N_H > G_H(\bar{\sigma}_H)P(0)$ and consequently $G_H(\bar{\sigma}_H) < \frac{N_H}{P(0)} < \frac{\varepsilon}{BP(0)}$, implying that $\lim_{\varepsilon \to 0} \bar{\sigma}_H = 0$.

Now observe that $\lim_{\varepsilon \to 0} \bar{\sigma}_H = 0$ implies $\lim_{\varepsilon \to 0} \bar{a} = a_H$. Recall that $\bar{\sigma}_H = \frac{\alpha(\bar{a}-a_H)}{P^{-1}(\tilde{P}(\alpha,B))}$ and $\bar{\sigma}_L = \frac{\alpha(\bar{a}-a_L)}{P^{-1}(\tilde{P}(\alpha,B))}$ so that $\bar{\sigma}_H = \frac{\bar{a}-a_H}{\bar{a}-a_L}\bar{\sigma}_L$. Therefore, $\bar{a} - a_H = \frac{\bar{\sigma}_H}{\bar{\sigma}_L - \bar{\sigma}_H}(a_H - a_L)$. Note also that $G_L(\bar{\sigma}_L) > \frac{R^*}{B}$ (the number of low-ability applicants which is still greater than the number of recipients must be at least as high as in the full-separation optimum). Consequently, as $\bar{\sigma}_H \to 0$, $\frac{\bar{\sigma}^H}{\bar{\sigma}^L - \bar{\sigma}^H} \to 0$ and therefore $\bar{a} - a_H \to 0$.

Consider what happens when $\varepsilon \to 0$. The resulting value of the objective is $N_L(\varepsilon)$. N_L is a continuous function of policy parameters in the relevant region.[13] Suppose that $\lim_{\varepsilon \to 0}(\alpha, \bar{a}, B) \neq (\alpha^*, \bar{a}^*, B^*)$. In that case, $\lim_{\varepsilon \to 0} N_L(\varepsilon) = N_L(\lim_{\varepsilon \to 0} \alpha(\varepsilon), \lim_{\varepsilon \to 0} \bar{a}(\varepsilon), \lim_{\varepsilon \to 0} B(\varepsilon)) < N_L(\alpha^*, \bar{a}^*, B^*)$ — the last inequality follows from the fact that the limiting point $(\lim_{\varepsilon \to 0} \alpha(\varepsilon), \lim_{\varepsilon \to 0} \bar{a}(\varepsilon), \lim_{\varepsilon \to 0} B(\varepsilon))$ implements full separation because $\tilde{P}(\alpha(\varepsilon), B(\varepsilon)) > P(0)$ (by Proposition 5), $a(\varepsilon) \to a_H$ as demonstrated earlier and $(\alpha^*, \bar{a}^*, B^*)$ was the optimal point under full separation. This is however a contradiction because it implies that for sufficiently small ε we would have been better off using the full separation policy (and not using all of the money). Consequently, $\lim_{\varepsilon \to 0}(\alpha, \bar{a}, B) = (\alpha^*, a^H, \bar{B})$. ■

Proof of identity 14. Recall the definition of N_a, equation (A.1) and the definition of $\bar{\sigma}_a$ in equation (6). α affects N_a through two channels. First, $\alpha(\bar{a} - a)$ is the maximum realization of the individual error term that results in receiving benefits — it enters both the integrand in N_a and the limit of integration. Second, the minimum acceptable probability threshold $\tilde{P}(\alpha, B)$ influences who applies. The effect of α on N_a is the sum of these two effects. Instrument $\bar{a}$ works only on the first margin, while instrument B works only on the second margin. Recognizing that allows for writing the effect of α as a combination of the effects with respect to the other two probabilities. ■

Proof of Proposition 7. Part 1.

Let α^* be the optimal value of α at the full separation policy with maximum budget R^*. Denote by $\tilde{\lambda}$ the Lagrange multiplier from the problem of maximizing the objective function with respect to α and $\bar{a}$ while setting $B = \bar{B}$. It can be easily shown that we will want to increase B over $\bar{B}$ if and only if

$$\frac{\partial N_L}{\partial B} > \frac{\tilde{\lambda}B}{1-\tilde{\lambda}B}\frac{\partial N_H}{\partial B} + \frac{\tilde{\lambda}B}{1-\tilde{\lambda}B}\frac{N_L + N_H}{B} \tag{A.5}$$

The left-hand side is non-negative and we don't have to worry about it increasing without bounds

[13] N_a is continuous when $\bar{a} > a$ and has a discontinuity at $\bar{a} = a$ when $\tilde{P} = P(0)$. In this case, the discontinuity is at a_L, but we are considering $\bar{a} \geq a_H > a_L$.

as ε changes — it converges to a finite limit of $\frac{\partial N_L}{\partial B}(\alpha^*, a_H, \bar{B})$. All terms on the right-hand side are non-negative and $\frac{N_L+N_H}{B}$ is finite and bounded away from zero ($N_L > \frac{R^*}{\bar{B}}$ and $B = \bar{B}$). We will show that $\frac{\tilde{\lambda}B}{1-\tilde{\lambda}B} \to \infty$ as $\varepsilon \to 0$ and thus this inequality is violated for small enough ε. To see that, recall that $\frac{\tilde{\lambda}B}{1-\tilde{\lambda}B} = \frac{\partial N_L/\partial \bar{a}}{\partial N_H/\partial \bar{a}}$. We will show that the numerator is finite while the denominator falls to zero. To see that, write explicitly $\partial N_a/\partial \bar{a}$:

$$\frac{\partial N_a}{\partial \bar{a}} = \int_0^{\bar{\sigma}_a} \alpha p\left(\frac{\alpha(\bar{a}-a)}{\sigma}\right)\frac{g_a(\sigma)}{\sigma}d\sigma + \frac{\alpha}{P^{-1}(\tilde{P}(\alpha,B))}\tilde{P}(\alpha,B)g_a(\bar{\sigma}_a) \tag{A.6}$$

All terms here are non-negative. The first-term vanishes for the high-types as $\varepsilon \to 0$, because $\lim_{\varepsilon\to 0}\bar{\sigma}_H = 0$ while the integrand is bounded away from infinity in the neighborhood of $\sigma = 0$ by Assumption 3. It remains positive for low-ability individuals because $\bar{\sigma}_L$ remains bounded away from zero as $\varepsilon \to 0$. By Proposition 4, $\lim_{\varepsilon\to 0}\tilde{P}(\alpha,B) = \tilde{P}(\alpha^*,\bar{B}) > P(0)$, so that $P^{-1}(\tilde{P}(\alpha,B))$ has non-zero limit. Consequently, for the high-ability types the second term disappears because $g_H(0) = 0$ while it remains positive for the low-ability types. As a result $\lim_{\varepsilon\to 0}\frac{\tilde{\lambda}B}{1-\tilde{\lambda}B} = \lim_{\varepsilon\to 0}\frac{\partial N_L/\partial \bar{a}}{\partial N_H/\partial \bar{a}} = \infty$, implying that the inequality (A.5) must be violated for sufficiently small ε and therefore $B = \bar{B}$ is optimal for sufficiently small εs.

Part 2.

This is an implication of condition (17) evaluated at $\bar{B}$ and the optimal α and $\bar{a}$. To see this, hold $\bar{B}$ constant and increase R. The parameter α is bounded by $P(0) < \tilde{P}\left(\alpha,\bar{B}\right) < 1$, so that $\frac{\partial\tilde{P}/\partial\alpha}{\partial\tilde{P}/\partial B} = -\frac{e^{\beta B}-1}{e^{\beta f(\alpha)}-1}f'(\alpha)$ is bounded away from zero when evaluated at the optimal α and $\bar{B}$.

Moreover, we can show that $\frac{\partial N_H}{\partial \bar{a}} \to 0$ as we keep increasing R. To see this, start by noting that, as R increases, $\bar{a}$ has to increase. There exists a finite budget size $\bar{R} \geq \bar{B}\left(N_L^* + N_H^*\right)$ at which everyone receives benefits, and at that budget size we have $\bar{a} = \infty$ and $\bar{\sigma}_a = \infty$. As R approaches this value, we have $\bar{a} \to \infty$ and $\bar{\sigma}_a \to \infty$.

Now recall eq. (A.6) and consider what happens as $\bar{a}$ and $\bar{\sigma}_a$ increases. The second term can be written as $\tilde{P}(\alpha,B)\frac{\bar{\sigma}_a g_a(\bar{\sigma}_a)}{(\bar{a}-a)}$. We must have $\lim_{\bar{\sigma}_a\to\infty}\bar{\sigma}_a g_a(\bar{\sigma}_a) = 0$ (if the limit exists), because otherwise $g_a(\cdot)$ would not be a distribution function. Moreover, $\bar{a}-a$ tends to ∞ so that the second term disappears in the limit. For the first term, integration by parts yields

$$\begin{aligned}\int_0^{\bar{\sigma}_a} \alpha p\left(\frac{\alpha(\bar{a}-a)}{\sigma}\right)\frac{g_a(\sigma)}{\sigma}d\sigma &= \frac{1}{\bar{a}-a}\left\{- P\left(\frac{\alpha(\bar{a}-a)}{\sigma}\right)\sigma g_a(\sigma)\Big|_0^{\bar{\sigma}_a} + \int_0^{\bar{\sigma}_a} P\left(\frac{\alpha(\bar{a}-a)}{\sigma}\right)\left[g_a(\sigma)+\sigma g_a'(\sigma)\right]d\sigma\right\} \\ &= \frac{1}{\bar{a}-a}\left\{\left(1-\tilde{P}\left(\alpha,B\right)\right)\bar{\sigma}_a g_a(\bar{\sigma}_a) + \int_0^{\bar{\sigma}_a} P\left(\frac{\alpha(\bar{a}-a)}{\sigma}\right)\left[g_a(\sigma)+\sigma g_a'(\sigma)\right]d\sigma\right\}.\end{aligned}$$

We have that $\frac{1}{\bar{a}-a}$ tends to zero. The first-term in the bracket disappears as $\bar{\sigma}_a$ tends to infinity. Because $P(\cdot)$ is a c.d.f., it can be bounded from above by 1, so that the second term is smaller than $\int_0^{\bar{\sigma}_a} g_a(\sigma) + \sigma g_a'(\sigma)d\sigma = \sigma g_a(\sigma)\Big|_0^{\bar{\sigma}_a} = \bar{\sigma}_a g_a(\bar{\sigma}_a)$ and therefore also disappears as $\bar{\sigma}_a$ gets large. Consequently, $\frac{\partial N_a}{\partial \bar{a}}$ tends to zero as $\bar{a}$ and $\bar{\sigma}_a$ — and budget size R — become large (for both L- and H-types).

By implication, as we keep increasing the budget size R, the left-hand side of (17) goes to zero, whereas the right-hand side increases without bound. As a result, for a large enough R the inequality has to be violated.

Part 3.

Suppose that the optimal policy satisfies $B = \bar{B}$. Because a full-separation policy delivers N_L^* individuals, an optimal policy must provide benefits to more than N_L^* individuals. By definition of $\bar{a}^*$, it must therefore satisfy $\bar{a} \geq \bar{a}^*$. Furthermore, it must satisfy $G_L(\bar{\sigma}_L) > N_L^*$ (the number of low-ability applicants which is greater than the number of recipients must be greater than N_L^*), such that $\bar{\sigma}_L > G_L^{-1}(N^*)$. Recall the identity $\bar{\sigma}_H = \frac{\bar{a}-a_H}{\bar{a}-a_L}\bar{\sigma}_L$. This formula is increasing in $\bar{a}$ and therefore $\bar{\sigma}_H \geq \frac{\bar{a}^*-a_H}{\bar{a}^*-a_L}\bar{\sigma}_L > \frac{\bar{a}^*-a_H}{\bar{a}^*-a_L}G_L^{-1}(N_L^*)$. Note that this lower limit is strictly positive because $\bar{a}^* > a_H$.

Now, note that we also have an upper bound for $\bar{\sigma}_H$: We need to have at least N_L^* low-ability recipients and, with the budget R, we can then have no more than $\frac{R}{\bar{B}} - N_L^*$ high-ability recipients. Consequently, $N_H < \frac{R}{\bar{B}} - N_L^*$ while $N_H > P(0)\, G_H(\bar{\sigma}_H)$ (at least a share $P(0)$ of the high-ability applicants receive benefits, because $\tilde{P}(\alpha, B) > P(0)$). Consequently, $\bar{\sigma}_H < G_H^{-1}\left(\frac{1}{P(0)}\left(\frac{R}{\bar{B}} - N_L^*\right)\right)$. Putting it together we have $\frac{\bar{a}^* - a_H}{\bar{a}^* - a_L} G_L^{-1}(N_L^*) < \bar{\sigma}_H < G_H^{-1}\left(\frac{1}{P(0)}\left(\frac{R}{\bar{B}} - N^*\right)\right)$. If the upper bound is lower than the lower bound, there is no $\bar{\sigma}_H$ that satisfies this condition and therefore our original assumption that $B = \bar{B}$ is optimal must be false.[14] ■

[14]Remarks:

1. There is no inconsistency with $B = \bar{B}$ for small R — as we reduce R, $a^* \to a^H$ and therefore the lower bound goes to zero (I think the upper bound also goes to zero, but apparently our assumption of a finite slope of density guarantees that it does not go to zero that fast).
2. If we can choose $B > \bar{B}$, a^* would fall — this is the same argument as the one we made to show that we can always spend all of the money on a full separation policy, higher B allows for setting higher α while holding $\tilde{P}$ constant. With the same $\tilde{P}$ but higher α, probability of receiving benefits for everyone goes up because $\bar{a} > a^L$ and screening is better — there is therefore more applicants and they are more successful. Consequently, for high enough B we can guarantee the existence of σ^H that would satisfy the inequality.

References

Aizer, Anna. (2007). 'Public health insurance, program take-up and child health', *Review of Economics and Statistics* .

Akerlof, George A. (1978). 'The economics of "tagging" as applied to optimal income tax, welfare programs, and manpower planning', *American Economic Review* 68(1), 8–19.

Benítez-Silva, Hugo, Buchinsky, Moshe and Rust, John. (2004), How large are the classification errors in the Social Security award process?, Working Paper 10219, National Bureau of Economic Research.

Bertrand, Marianne, Mullainathan, Sendhil and Shafir, Eldar. (2004). 'A behavioral-economics view of poverty', *American Economic Review Papers and Proceedings* 94(2), 419–423.

Besley, Timothy and Coate, Stephen. (1991). 'Public provision of private goods and the redistribution of income', *American Economic Review* 81(4), 979–84.

Besley, Timothy and Coate, Stephen. (1992). 'Workfare versus welfare incentive arguments for work requirements in poverty-alleviation programs', *American Economic Review* 82(1), 249–61.

Besley, Timothy and Coate, Stephen. (1995). 'The design of income maintenance programmes', *Review of Economic Studies* 62(2), 187–221.

Bitler, Marianne, Currie, Janet and Scholz, John Karl. (2003). 'WIC participation and eligibility', *Journal of Human Resources* 38, 1139–79.

Blackorby, Charles and Donaldson, David. (1988). 'Cash versus kind, self-selection and efficient transfers', *American Economic Review* 78(4), 691–700.

Blank, Rebecca M. (2002). 'Evaluating welfare reform in the United States', *Journal of Economic Literature* 40(4), 1105–1166.

Borjas, George J. and Hilton, Lynette. (1996). 'Immigration and the welfare state: Immigrant participation in means-tested entitlement programs', *Quarterly Journal of Economics* 111(2), 576–604.

Bruce, Neil and Waldman, Michael. (1991). 'Transfers in kind: Why they can be efficient and nonpaternalistic', *American Economic Review* 81(5), 1345–1351.

Currie, Janet. (2000), Do children of immigrants make differential use of public health insurance?, *in* **George J. Borjas**., ed., 'Issues in the Economics of Immigration', The University of Chicago Press, Chicago; London.

Currie, Janet. (2003), U.S. food and nutrition programs, *in* **Robert A. Moffitt,**, ed. (2003), *Means-Tested Transfer Programs in the United States*, University of Chicago Press.

Currie, Janet. (2004), The take up of social benefits, Working Paper 10488, National Bureau of Economic Research.

Currie, Janet and Grogger, Jeffrey. (2001). 'Explaining recent declines in Food Stamp program participation', *Brookings-Wharton Papers on Urban Affairs* pp. 203–244.

Daly, Mary C. and Burkhauser, Richard V. (2003), The Supplemental Security Income Program, *in* **Robert A. Moffitt,**, ed. (2003), *Means-Tested Transfer Programs in the United States*, University of Chicago Press.

Daponte, Beth Osborne, Sanders, Seth and Taylor, Lowell. (1999). 'Why do low-income households not use Food Stamps? evidence from an experiment', *Journal of Human Resources* 34(3), 612–628.

Diamond, Peter A. and Mirrlees, James A. (1978). 'A model of social insurance with variable retirement', *Journal of Public Economics* 10(3), 295–336.

Diamond, Peter A. and Sheshinski, Eytan. (1995). 'Economic aspects of optimal disability benefits', *Journal of Public Economics* 57(1), 1–23.

Duggan, Mark G. and Kearney, Melissa S. (2005), The impact of child SSI enrollment on household outcomes: Evidence from the Survey of Income and Program Participation, Working Paper 11568, National Bureau of Economic Research.

Gruber, Jonathan. (2003), Medicaid, *in* **Robert A. Moffitt,**, ed. (2003), *Means-Tested Transfer Programs in the United States*, University of Chicago Press.

Heckman, James J. and Smith, Jeffrey A. (2004). 'The determinants of participation in a social program: Evidence from a prototypical job training program', *Journal of Labor Economics* 22(2), 243–98.

Holtzblatt, Janet and McCubbin, Janet. (2003), Complicated lives: Tax administrative issues affecting low-income filers, mimeo, Department of the Treasury.

Kopczuk, Wojciech and Pop-Eleches, Cristian. (2007). 'Electronic filing, tax preparers and participation in the Earned Income Tax Credit', *Journal of Public Economics* 91(7-8), 1351–1367.

Luce, Robert Duncan. (1959), *Individual Choice Behavior; a theoretical analysis*, Wiley, New York.

Mill, John Stuart. (1848), *Principles of Political Economy*, Vol. Books IV and V, Penguin, Harmondsworth, UK.

Moffitt, Robert. (1983). 'An economic model of welfare stigma', *American Economic Review* 73(5), 1023–35.

Moffitt, Robert A., ed. (2003), *Means-Tested Transfer Programs in the United States*, University of Chicago Press.

Nagi, Saad Z. (1969), *Disability and Rehabilitation: Legal, Clinical, and Self-Concepts and Measurement*, Ohio State University Press, Columbus, OH.

Nichols, Albert L. and Zeckhauser, Richard. (1982). ‘Targeting transfers through restrictions on recipients’, *American Economic Review* 72(2), 372–77.

Olsen, Edgar O. (2003), Housing programs for low-income households, *in* **Robert A. Moffitt,**, ed. (2003), *Means-Tested Transfer Programs in the United States*, University of Chicago Press.

Parsons, Donald O. (1996). ‘Imperfect ’tagging’ in social insurance programs’, *Journal of Public Economics* 62(1-2), 183–207.

Saez, Emmanuel. (2002). ‘Optimal income transfer programs: Intensive versus extensive labor supply responses’, *Quarterly Journal of Economics* 117(3), 1039–1073.

Scholz, John Karl. (1994). ‘The Earned Income Tax Credit: Participation, compliance and anti-poverty effectiveness’, *National Tax Journal* 48(1), 59–81.

Slemrod, Joel. (1990). ‘Optimal taxation and optimal tax systems’, *Journal of Economic Perspectives* 4(1), 157–78.

Slemrod, Joel and Yitzhaki, Shlomo. (2002), Tax avoidance, evasion and administration, *in* **Alan Auerbach and Martin S. Feldstein**., eds, ‘Handbook of Public Economics’, Vol. 3, Elsevier/North Holland, Amsterdam; New York.

Stern, Nicholas. (1982). ‘Optimal taxation with errors in administration’, *Journal of Public Economics* 17(2), 181–211.

TABLE 1. SOCIAL PROGRAMS IN THE UNITED STATES

PROGRAM	TAKE UP[1]	TARGETING[2]	COMPLEXITY[2]
Medicaid	73%	Medium	High
Medicare Part B[3]	96%	Low	Low
Supplemental Security Income Program (SSI)	60%	High	High
Social Security Disability Insurance (DI)	No estimate	High	High
The Earned Income Tax Credit (EITC)	80%-86%	Medium	High
Temporary Assistance for Needy Families (TANF)[4]	60%-90%	Medium	High
Housing Programs	below 50%	Medium	High
Food Stamps	69%	Medium	Medium
The Special Supplemental Nutrition Program for Women,Infants and Children (WIC)	67%, 73%, 38%	High	High
Child Care Subsidies	40%	Medium	High

1) Estimated take-up rates are averages for the entire eligible populations except for Medicaid (children), SSI (elderly), and WIC (women, infants, children). Where an interval is provided, this reflects different estimation methods and/or data. Take-up estimates are taken from following sources. Medicaid: Gruber (2003), Medicare: Currie (2004), SSI: Daly and Burkhauser (2003), EITC: Scholtz (1994), TANF: Blank (2001), Housing Programs: Olsen (2003), Food Stamps: Currie (2003), WIC: Bitler et al. (2003), Child Care Subsidies: Witte (2002).

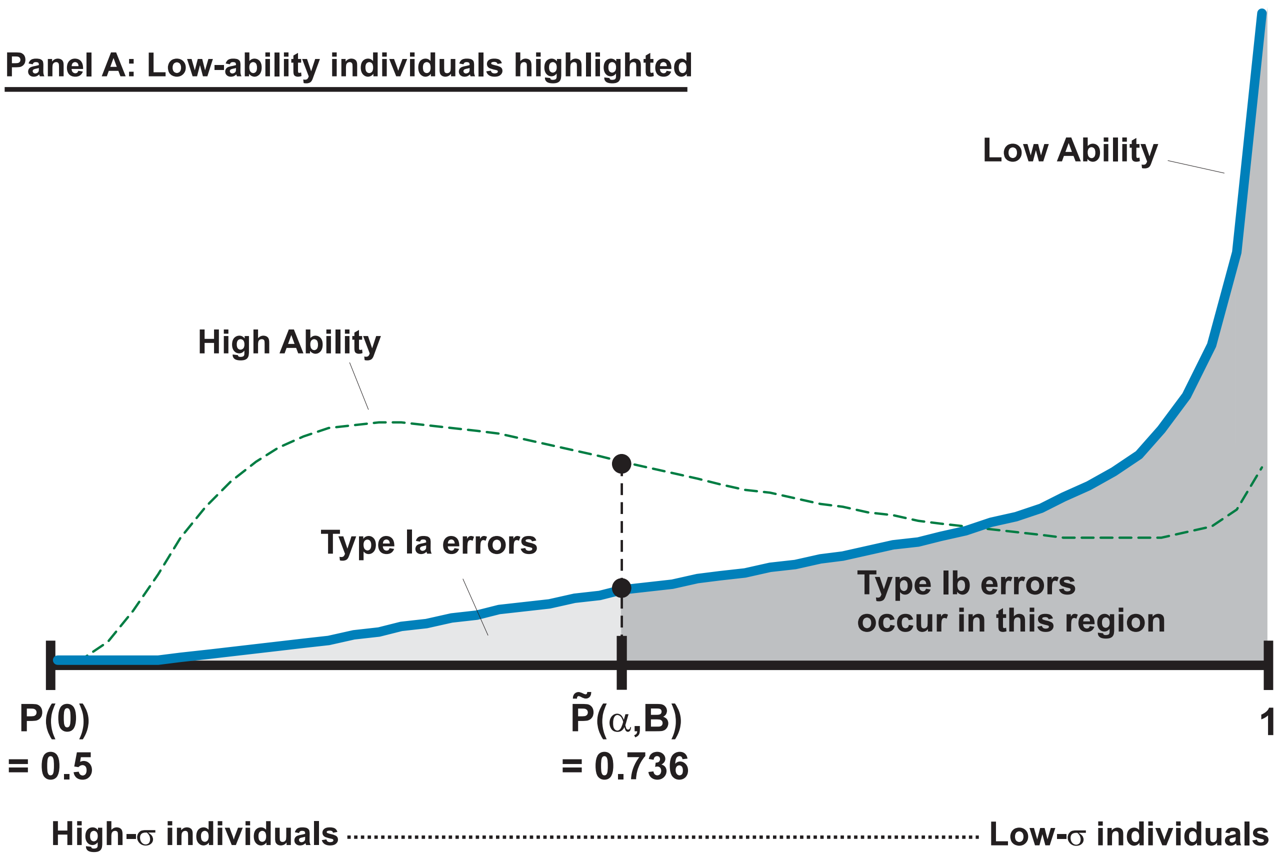
2) Targeting is interpreted as the strictness of elibility criteria (the reverse of universalism), while complexity is interpreted broadly to include all transaction costs. Targeting and complexity classifications (low, medium, high) are based on Moffitt (2003) and Currie (2004).

3) Medicare Part A is mandatory and therefore has a take-up rate equal to 100%.

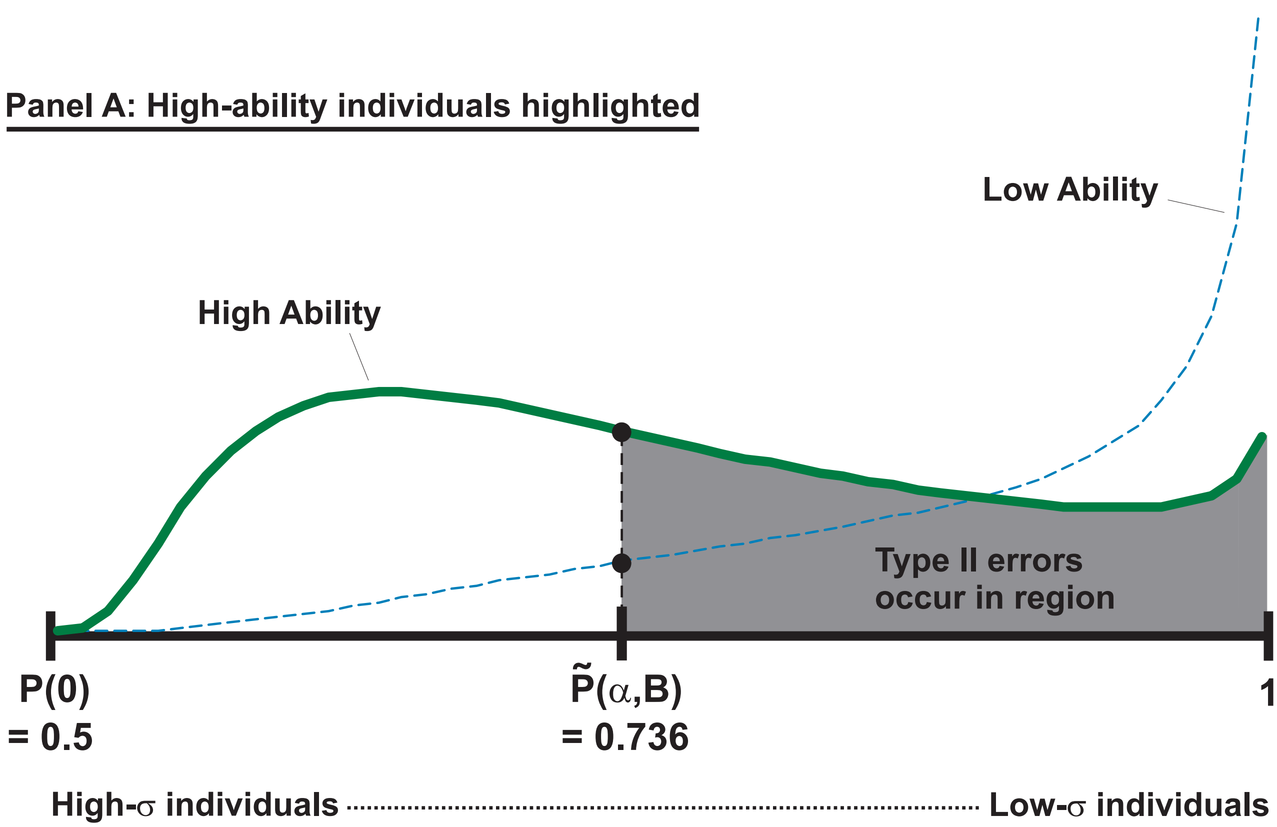
4) Replaced Aid to Families with Dependent Children (AFDC) in 1996. Take-up estimates are based on the AFDC program.

FIGURE 1. DENSITIES OF P(.) AND CLASSIFICATION ERRORS

Panel A: Low-ability individuals highlighted



Panel A: High-ability individuals highlighted



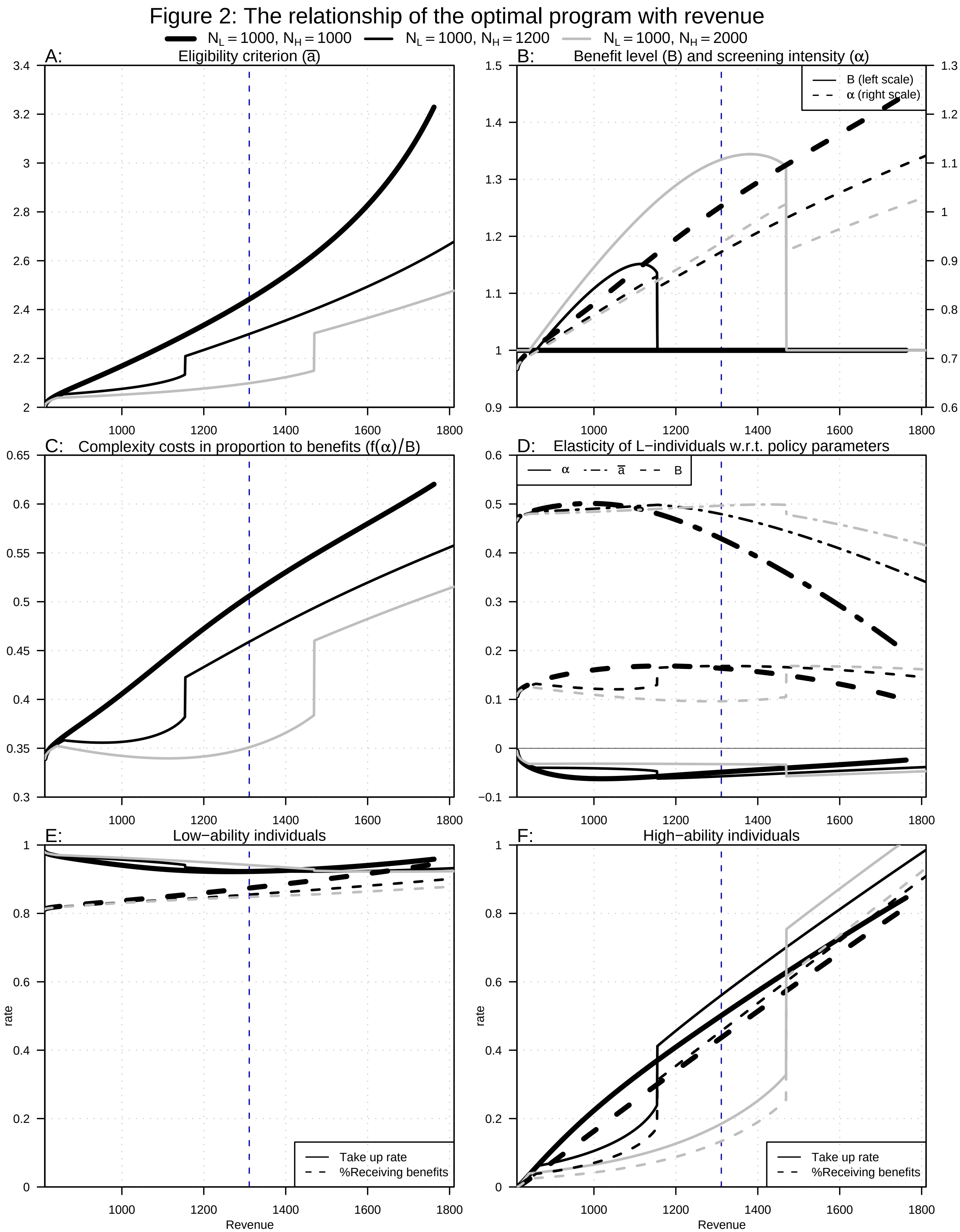


Figure 2: The relationship of the optimal program with revenue
N_L = 1000, N_H = 1000
N_L = 1000, N_H = 1200
N_L = 1000, N_H = 2000
A: Eligibility criterion (ā)
B: Benefit level (B) and screening intensity (α)
B (left scale)
α (right scale)
C: Complexity costs in proportion to benefits (f(α)/B)
D: Elasticity of L−individuals w.r.t. policy parameters
α
ā
B
E: Low−ability individuals
F: High−ability individuals
Take up rate
%Receiving benefits
rate
Revenue

Figure 3: The relationship of the optimal program with c_1, where $f(\alpha) = c_0 + c_1 \cdot \alpha^{c_2}$, at $\overline{R} + 500$

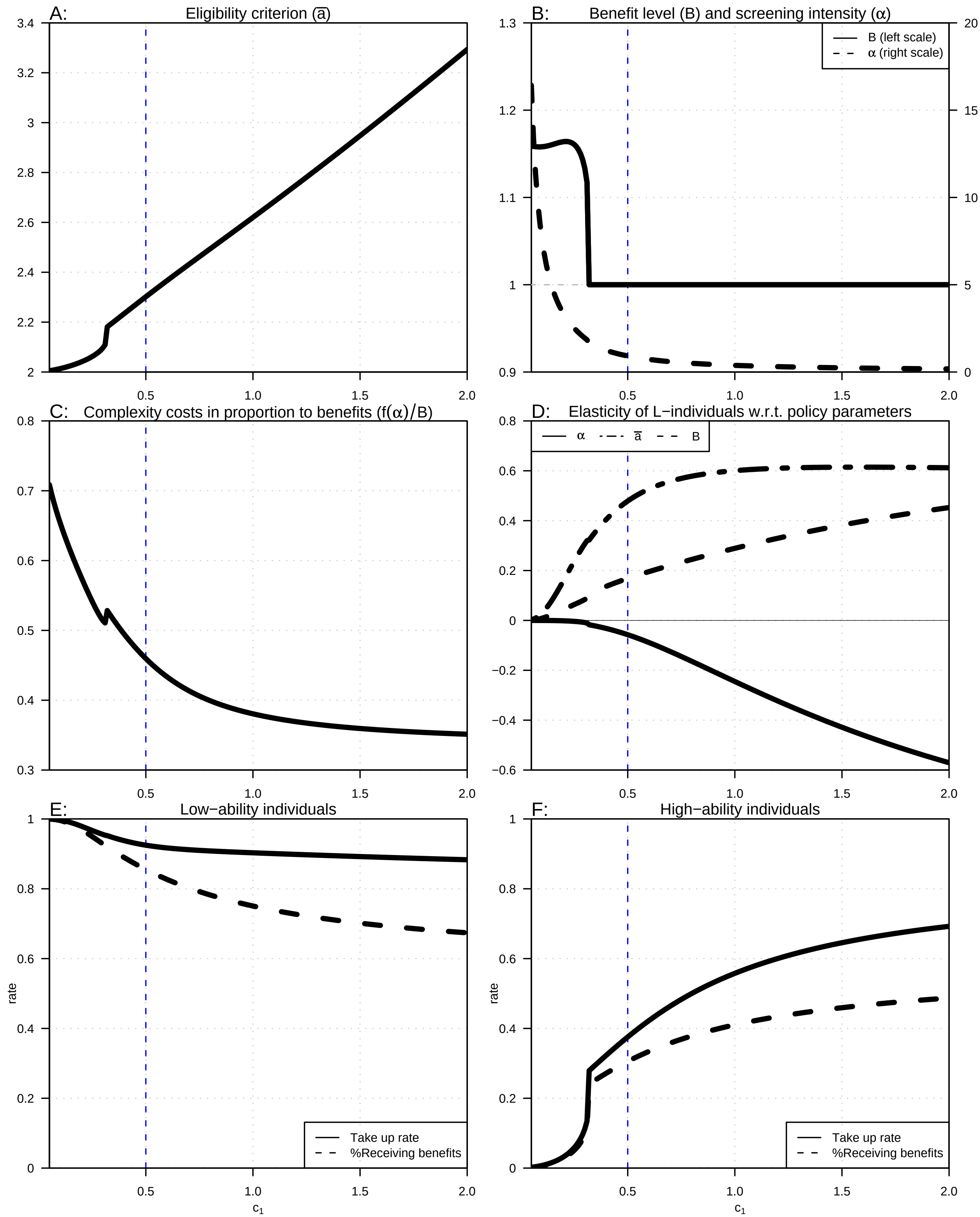


Figure 4: The relationship of the optimal program with c_0, where $f(\alpha) = c_0 + c_1 \cdot \alpha^{c_2}$, at $\overline{R} + 500$

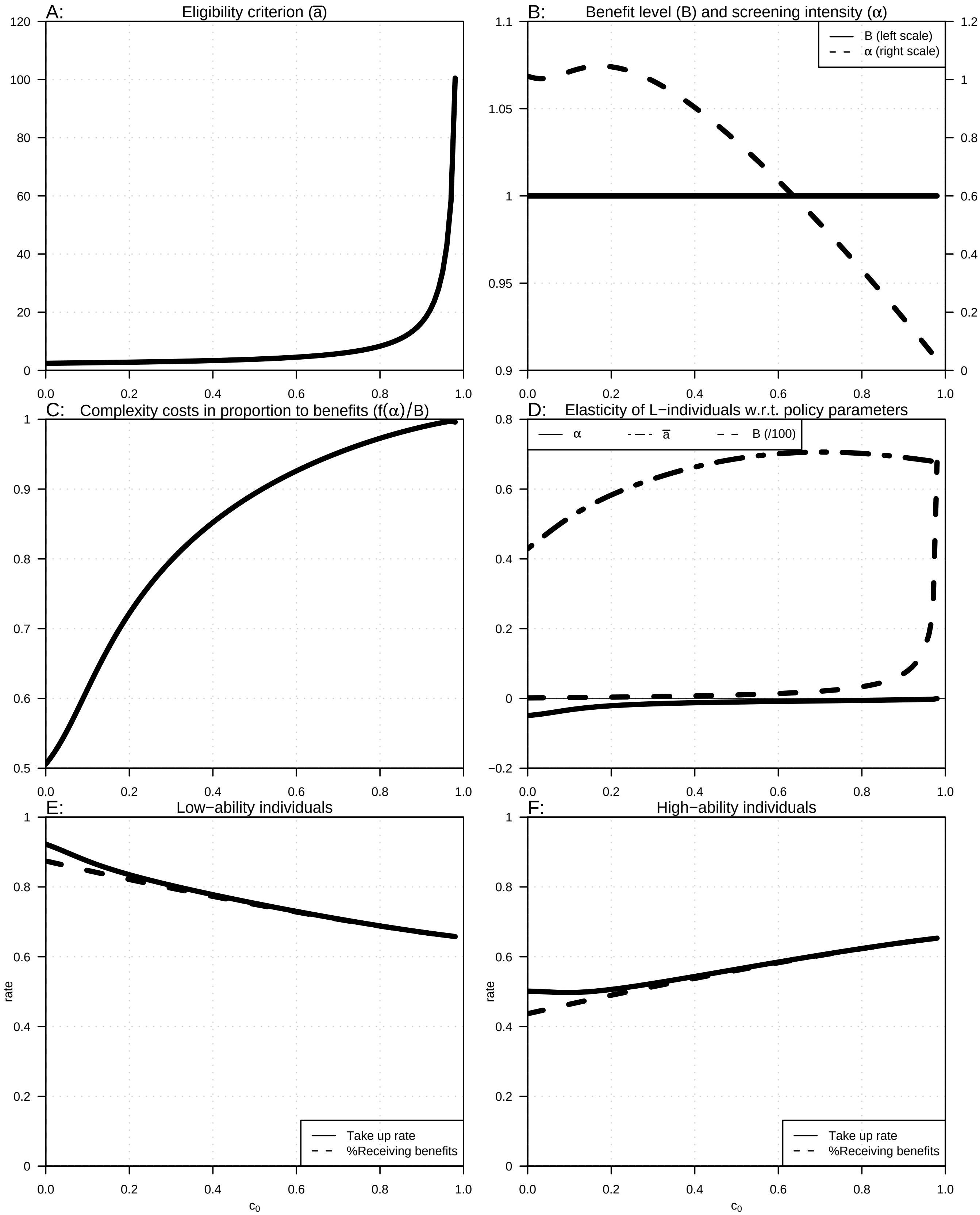


Figure 5: The relationship of the optimal program with c_2, where $f(\alpha) = c_0 + c_1 \cdot \alpha^{c_2}$, at $\overline{R} + 500$