Welfare Program Complexity and the Take Up of Social Benefits

Henrik Jacobsen Kleven
University of Copenhagen, EPRU, and CEPR

Wojciech Kopczuk
Columbia University and NBER

September 23, 2005
Abstract

We present a model of the design of social welfare programs that allows for reduced take-up of benefits in response to program complexity. We consider policy makers that are interested in providing a minimum level of benefits to as many “deserving” individuals as possible. The optimal programs that are not universal always feature a large degree of complexity (defined precisely in the paper). Although it is generally possible to eliminate take up by the undeserving, optimal policies usually involve eligibility criteria that make them eligible and rely on complexity to reduce their participation. Even though the government is only interested in providing the minimum level of benefits and is unable to reach all deserving individuals, we find that the optimal policy can feature benefits that are higher than this minimum level. This is because benefits generically screen better than either eligibility criteria or direct tests.
1 Introduction

The coverage and generosity of social benefits have increased substantially in the United States and European countries since World War II. Despite the cross-country differences that exist in the size and design of welfare programs, most of these countries spend sizeable amounts of public spending on various forms of social benefits. The programs include unemployment insurance, social assistance, in-work benefits, health insurance, food and nutrition programs, housing programs, child care support, and cash benefits for the aged, blind, and disabled. The programs have been motivated by a desire to alleviate poverty by providing a minimum bundle of goods to those who are truly needy. However, a concern about many of the programs is that not all intended recipients take them up. This is especially the case in the United States and the United Kingdom where the take up of some welfare benefits is very low. The presence of low take up is problematic because it undermines the basic goal of the program to alleviate poverty among the eligible population.

Empirical economists have long been concerned with the issue of low participation rates in public programs. The empirical literature hypothesizes three possible explanations for low take up: welfare stigma, transaction costs, and imperfect information. The most well-known work on take up behavior is the Moffitt (1983) analysis of welfare stigma. He modeled welfare program participation as a utility maximization problem where it may be suboptimal for the individual to take up a benefit that she is eligible for because of stigmatizing effects from doing so. In this story, the costs of taking up welfare benefits arises from this action being demeaning and shameful to the recipient. The stigma story may be consistent with other types of costs associated with welfare program participation. 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 has to do with detailed eligibility criteria, rigorous documentation requirements, difficult and time-consuming forms, interactions between programs, and the fact that the applicant may have to make several 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. Empirical research has shown that complexity and administrative hassle constitute important barriers to program enrollment (e.g., Currie and Grogger, 2001; Bitler et al., 2003; Aizer, 2004, 2003), and that such effects may be more important than stigma (Currie, 2004). The third explanation for low take up — imperfect information — 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 expected utility maximizer may be reluctant to incur the transaction costs associated with applying. If a potential applicant had perfect information about being eligible, it would always be worthwhile applying and getting the benefit with certainty as long as transaction costs are lower than the benefit itself.

Despite the fact that issues pertaining to complexity and administration seem to be
tremendously important to the effects of public policies in general, and to the take up of social benefits in particular, we are not aware of theoretical work modeling the complexity of public policy. Instead, policy-oriented research has focused on statutory rates (the size of benefit and tax rates) along with simple eligibility criteria (typically the size of earned income). If complexity is discussed at all, it tends to be mentioned as a negative side-effect of certain types of public policy programs. For example, in Akerlof’s (1978) analysis of redistributional policies involving tagging of those who are needy versus universal policies, he points out that a disadvantage of tagging — which must be weighed against its higher target efficiency — is its complexity and high costs of administration. In the take up literature, high transaction costs and the associated low take up are often seen as negative aspects of inadequately designed programs and something that calls for remedial policy action.

In this paper, we argue that it is important to formulate an explicit model of the complexity of welfare policies in order to understand their effects and design. We go beyond arguing that complexity is just a negative side-effect of targeted welfare policies. Because complexity can have an effect on participation, it is a policy instrument that may be used to screen welfare applicants in an environment of imperfect information, and hence complexity itself may have some desirable effects. At the same time, it limits the effectiveness of otherwise superior means of screening and therefore has implications for the optimal setting of alternative instruments that do not cause complexity by themselves such as benefits and eligibility criteria. By implication, high transaction costs and low take up should be viewed as equilibrium outcomes of situations where imperfectly informed policy makers, constrained by a limited budget, wants to alleviate poverty among the truly needy.

We set up a model where the government has imperfect information about the true eligibility of individuals applying to a welfare program. That is, an individual may or may not be eligible according to some innate characteristic (ability, health, disability, etc.) which is not directly observable. Instead, program administrators can subject each applicant to a test (involving documentation, interviews, testing by specialists, etc.) providing an imperfect indicator of true eligibility, which is then used by the administrators to grant or deny the benefit. Since the indicator is an imperfect one, errors generally occur. By increasing the complexity of the test (more documentation, more interviews, etc.), the precision of the signal is improved but at the same time it is made more costly for individuals to apply which may hurt take up.

The individuals know their own innate characteristic (and hence true eligibility), but they have imperfect information about the outcome of the test used by the program office to decide on eligibility and has to form an expectation about it. Individuals decide to apply for the benefit if the expected utility from doing so is at least as great as the expected utility from not applying. In this maximization problem, potential applicants weigh complexity-related transaction costs against the size of the benefit and the probability of getting it. Even ineligibles may find it worthwhile applying, because they may get lucky due to the fact that testing is not perfect. In fact, we find that the government may find it optimal to explicitly make “non-deserving” individuals eligible but rely on complexity to discourage
The literature has distinguished between two types of error in the provision of welfare benefits. Errors can occur because some of those who are eligible for a benefit do not take it up (so-called type I errors), and errors can occur because ineligibles end up getting the benefit because they are able to disguise themselves as deserving (so-called type II errors). In this paper, we want to make an additional distinction between two different kinds of type I error. In our model, this type of error can occur either because eligible individuals do not find it worthwhile applying (type Ia) and it can occur because eligible applicants are denied benefits by the program administrators due to the imperfection of the test (type Ib). For a government wanting to alleviate poverty among those who are truly needy, being constrained by a limited budget, it is desirable to minimize all sources of error. A high occurrence of type Ia and type Ib errors undermine the goal of poverty-alleviation, while a high occurrence of type II errors make the program too expensive. Hence, the choice of parameters in the welfare program — complexity, the size of the benefit, and the eligibility rule — reflects the ability of each parameter to reduce the different kinds of error. Increasing the complexity of the welfare program has two effects: (i) it implies higher costs of applying which, ceteris paribus, makes less people apply. This reduces type II errors but increases type Ia errors. (ii) It makes for a more precise test so that eligible individuals have a higher probability of getting the benefit conditional on applying, whereas ineligible applicants have a lower probability of getting the benefit. This reduces both type Ia, type Ib, and type II errors. Hence, the choice of complexity reflects trading-off the positive effects from a more precise test, along with the hassle it creates for ineligibles, against the negative effects on take up for truly deserving individuals due to higher transaction costs.

We show that there are three possible equilibria for welfare policy, corresponding to a small government budget, an intermediate budget, and a large budget. If the government budget is sufficiently small, policy makers implement a policy involving full separation between those who are truly eligible and those who are not. This means that no type II errors are being made, which is achieved through a combination of a relatively strict eligibility rule and a high degree of complexity/hassle in the application process making it suboptimal for the ineligibles to apply. The equilibrium degree of complexity depends among other things on the individuals’ attitudes toward risk: if risk aversion is sufficiently small, we show that total transaction costs may constitute more than 50% of total welfare benefits. This implies that a lot of type I errors are being made (take up rates are low). We may think of this equilibrium as capturing the case of a right-wing government (say, the Anglo-Saxon case), where high complexity and low take up is optimal due to the fact that policy makers, constrained by a small budget for redistribution, are very keen to avoid type II errors.

For an intermediate range of budgets, optimal welfare policy no longer involves full separation between the deserving and the non-deserving, so that both type I and type II errors are being made. With more money available for redistribution, the government is willing to give money to some of the ineligibles in order to increase take up among the truly needy. Hence, the threshold for eligibility is less strict in this case. The degree of complexity
continues to be high, however, because this is still an effective instrument to restrict the number of ineligibles applying, and to sort between those who do apply.

As we continue to increase the budget size, the threshold for being eligible to the benefit is made less strict – the policy is made more universal – such that we are giving money to more people. Ultimately, for a sufficiently large budget, we enter a third range where the government can afford to give a universal benefit to all citizens. In this equilibrium, since everybody takes the benefit, a lot of type II errors are being made while no type I errors are being made, and there is no need for complexity to screen individuals. This case may be interpreted as the case of a left-wing government (say the Scandinavian case) choosing simple and universal welfare programs with full take up. In between the extreme-left-wing case (large budget, no type I errors, universalism, and low complexity) and the right-wing case (small budget, no type II errors, targeting, and high complexity), the intermediate range reflects a “moderate” government implementing policies with some type I and some type II errors, and where complexity is still high. This intermediate case corresponds to the most interesting case of large but non-universal programs for which very strict screening would limit participation below acceptable levels.

Our work may be seen as a first step to incorporating complexity as a choice variable in the analysis of public policy, and to view it as an equilibrium outcome of situations where governments operate in environments of imperfect information. We contribute to the literature studying the optimal design of transfer programs and to the theory of optimal redistribution more generally. Of particular interest here are the papers showing that in-kind instruments — including workfare — can serve a screening role and hence improve upon the target efficiency of poverty-alleviation programs (Nichols and Zeckhauser, 1982; Blackorby and Donaldson, 1988; Besley and Coate, 1992, 1995). Interestingly, Nichols and Zeckhauser (1982) mentioned the possibility that ordeals such as demeaning qualification tests and tedious administrative procedures may be able to sort between those who are truly needy and those who are not. The concept of ordeals is closely related to the type of workfare studied by Besley and Coate (1992, 1995). In our model, the complexity of the application process may indeed be seen as a form of ordeal used to screen individuals.

At the same time, there are important differences between our paper and the existing literature. First of all, in our model there is a better screening instrument and its use is hampered by the presence of complexity. If the government was able to decouple screening from complexity, there would be no reason for introducing complexity as an ordeal. This is not possible, and the government has to account for both the impact that complexity has on screening through a standard instrument (where it contributes to reduced take-up) and its role as an ordeal mechanism. Second, the literature did not model the process by which program administrators obtain indicators of innate characteristics (e.g. productivity or health) on the basis of which benefits are granted or denied. Instead, the literature simply assumed that innate characteristics are unobservable, whereas market transactions (earned income, health expenditures, etc.) can be perfectly observed. In our framework, the precision of the indicator of true ability is related to the complexity of the system, which is a policy instrument. By implication, complexity can be used as a screening device both
because of the hassle it creates for applicants and because it improves the quality of the signal to program administrators. By contrast, the type of ordeals/workfare discussed in the previous papers is a pure hassle (a negative transfer); it may serve a sorting function only if it imposes different utility costs on eligibles and ineligibles.

Another contribution of our paper is to integrate the analysis of type I and type II errors into a single framework. So far, the two types of error have been studied separately, with the take up literature focusing exclusively on the occurrence of type I errors and the optimal screening literature considering the occurrence of type II errors and the incentives created for the non-deserving to reveal themselves truthfully. The possibility of eligible individuals not claiming the transfer, or being rejected in the application process, is not considered in the screening literature. From the perspective of this literature, it is always desirable to implement policies that minimize type II errors, whereas from the perspective of the take up literature it is always good to reduce type I errors. Instead, we argue that policies can be better understood by considering the trade-off between the different types of error. It is exactly because policy makers wish to avoid type II errors that they may want to design a system with high transaction costs (complexity) and low take up.

The organization of the paper is as follows. In Section 2, we review the empirical evidence on take-up behavior and its determinants, and we present evidence that complexity and administrative hassle constitute important barriers to program enrollment. In Section 3, we formulate a model of welfare program complexity, and derive a number of results on program design, complexity and take up consistent with the empirical evidence presented in Section 2. Finally, Section 4 offers a discussion of implications and model assumptions.

2 The empirical literature

2.1 Observed take-up behavior and some possible explanations

Non-participation in public programs by eligible individuals can be substantial. The issue has received the strongest attention in the United States where the take up of some social benefits is quite low. Moffitt (2003) and Currie (2004) survey a number of transfer programs in the United States, showing that there is a tremendous degree of variation in take-up rates across different programs, over time for each program, and across different states and demographic groups.

At the low end, the recently enacted State Children’s Health Insurance Program (SCHIP) serves only 9 percent of children meeting the income eligibility criteria defined in the program (Lo Sasso and Buchmueller, 2004). The take up of Medicaid, a program offering health insurance to the poor, is higher. In his survey of the medicaid program, Gruber (2003) reports that 22.6 percent of all low-income children were enrolled in the program in 1996, with a total of 31 percent of low-income children being eligible, implying a take-up rate for children equal to 73 percent. While this number reflects a relatively high degree of participation on average, it has been shown that marginal take up following expansions in Medicaid eligibility tend to be substantially lower. Currie and Gruber (1996a,b) estimated that only 23 percent of the children made newly eligible under the medicaid expansions in the 1980s actually enrolled in the program, while Card and Shore-Sheppard (2002) estimate...
marginal take-up rates as low as 5-13 percent. The take up of Medicaid by pregnant women is somewhat higher, with take-up rates for the newly-available Medicaid in the 1980s being 34 percent (Currie and Gruber, 1996b). In contrast to Medicaid, the Medicare program — a non means-tested program providing health coverage for the elderly and the disabled — has participation rates close to 100 percent (Currie, 2004).

Other social programs characterized by high degrees of non-participation include public housing programs with take up rates being far below 50 percent (Olsen, 2003), and the Child Care Subsidy Programs which serve only 15 percent of eligible children (Currie, 2004). The Supplemental Security Income Program (SSI) providing cash assistance for aged, blind and disabled individuals with low incomes has low take up by the poor elderly. For this group, Daly and Burkhauser (2003) calculate that participation in the SSI program declined steadily from 79 percent in 1974 to 54 percent in 1982 after which it has fluctuated a great deal. By 1998, 60 percent of all poor people above the age of 65 received SSI. For the Temporary Assistance for Needy Families Program (TANF), previously Aid to Families with Dependent Children (AFDC), take up has been estimated to be in the range from 60 to 90 percent of those eligible for the benefit (Blank, 2001), while the Food Stamp Program (FSP) has an average take-up rate at 69 percent (Currie, 2003). Finally, the Earned Income Tax Credit (EITC), a refundable tax allowance for working families with low income and kids, has a take-up rate at around 80 percent or less (Scholz, 1994; Holtzblatt and McCubbin, 2003).

In addition to this variation across different programs, there is a great deal of variation across individuals who are “equally” eligible. For example, several studies have documented large racial differences in the take up of social benefits, with welfare program participation being relatively low for Hispanic and Asian people and relatively high for African Americans. For the Food Stamp Program (FSP), for example, Currie (2003) reports that 92 percent of eligible African Americans participate, whereas only 61 percent of eligible Hispanics and 59 percent of eligible white non-Hispanics participate in that program. Similarly, Duggan and Kearney (2005) find that, conditional on being poor, Hispanic children are less likely to be enrolled in the SSI program. Currie (2000) finds that immigrant children, many of whom are Hispanic, are less likely to be enrolled in Medicaid conditional on being eligible. In general, immigrants tend to have low take up rates, other things being equal, although their tendency to take up social benefits increase as they become more assimilated in society (e.g Borjas and Hilton, 1996).

In addition to these differences across different ethnic groups and for immigrants and non-immigrants, many papers have documented a large variation in participation rates across age, gender, family structure, and state of residency. For example, take up among those eligible for food stamps is 86 percent for children, only 1/3 for the elderly, close to 100 percent for single mothers, and 78 percent for families with two or more adults (Currie, 2003). A similar degree of variation is found in the Special Supplemental Nutrition Program for Women, Infants and Children (WIC) with a take-up rate at 73 percent for eligible infants, a take-up rate at 38 percent for children at the ages one to four, and a take-up rate at 67 percent for pregnant and postpartum women. Finally, there is a large variation across
different states, with participation being low in a state such as Alaska and high in states such as Vermont and Maine (Currie, 2003).

A theory of welfare program participation should be consistent with the observed take-up behavior described above. In particular, such a theory should be able to explain the large observed differences in take up across different programs, and it should explain the observed differences in take up for the program as a whole and for marginal expansions of the program. It should also be able to explain different take up by equally eligible individuals. The empirical literature hypothesizes three possible explanations for low take up: (i) stigma associated with participating in public programs, (ii) transaction costs due to complexity and administrative hassle associated with applying for benefits, and (iii) lack of information about the existence of some programs, about the eligibility criteria defined in the programs, and about the outcome of the test used by administrators to determine eligibility. This paper formulates a theory of take up behavior based on a combination of the second and third explanations, i.e. complexity/hassle and imperfect information. Indeed, there is a substantial amount of evidence suggesting that these two aspects constitute important barriers to program participation. Our model can also be viewed as a non-standard case of the stigma story, where stigma increases with intensity of screening. We start by describing the importance of complexity and administrative hassle in applying for benefits in the United States.

2.2 Evidence on the complexity of public programs

The presence of complexity and bureaucracy in transfer programs in the United States is well-documented (Moffitt, 2003; Currie, 2004). The high degree of complexity is related both to the legislation and to the application processes. The eligibility rules vary from program to program, but they tend to be quite detailed and involve criteria relating to aspects such as income, assets, employment, marital status, pregnancy, infants, children, age, state of residence, health, nutritional status, and disability. In addition to the detailed rules for each program, there can be interactions of eligibility in one program with participation in other programs (“adjunctive eligibility”). The application process itself can be lengthy, cumbersome and demeaning; it tends to involve rigorous testing, documentation, interviewing, and there may be frequent re-certification for some programs. Moreover, many states have adopted welfare diversion policies that may increase transaction costs associated with applying for welfare benefits (Currie and Grogger, 2001). These are procedures intended to divert would-be applicants away from applying by, for example, telling them that they must satisfy job search requirements or offering lump sum financial assistance for those who decide not to apply.

The extent of administrative hassle in the Medicaid program has been documented by e.g. Gruber (2003) and Currie (2004). For the nonelderly and non-disabled, eligibility depends on family size and structure along with three separate income tests that must be passed. Some states also employ asset tests, while others do not. For the elderly and disabled, there exists a separate set of rules involving four different routes to Medicaid with different criteria related to income, assets, family structure, health expenditures, and
the participation in other programs. The application process involves strong requirements for documentation, a stringent time-limit on the number of days the applicant can spend providing the documentation, along with several trips to the program office for interviewing. Up to a quarter of applications are denied because applicants do not produce the required documentation within the time limit or because they fail to show up for all of the interviews. Coverage may need to be re-established every six months.

Currie and Grogger (2001) consider the Food Stamp Program (FSP), where the average application takes nearly five hours to complete, and involves at least two trips to the FSP office. Re-certification for benefits typically takes $2^{1/2}$ hours and requires at least one trip to the program office. Out-of-pocket application costs average about 6 percent of the average monthly benefit. Another program providing support for food and nutrition, the Special Supplemental Nutrition Program for Women, Infants and Children (WIC), is even more complicated (Currie, 2003; Bitler et al., 2003). This program involves complicated eligibility criteria. First of all, the individual is subject to a categorical test, since a recipient must either be a pregnant, breastfeeding, or postpartum woman; an infant up to the age of one year, or a child aged one to four. In addition, applicants have to fulfill income eligibility criteria (sometimes based on monthly, sometimes on annual income), along with criteria related to household structure and state recidency, or individuals may qualify because they are adjunctively eligible through participation in other programs such as AFDC/TANF, FSP, or Medicaid. Finally, participants in the program must be certified to be at “nutritional risk,” which is determined by a screening that takes place at the applicant’s first visit to the WIC clinic. This screening involves, among other things, a blood test for anemia.

The most complicated programs of them all are perhaps those providing disability insurance for those who are physically or mentally impaired. Benítez-Silva et al. (2004) study the Social Security Disability Insurance Program (DI) and the Supplemental Security Income Program (SSI) for the blind and disabled. These programs involve very long, complicated and bureaucratic application processes. An application is potentially subject to a five-stage screening procedure, involving a large number of program administrators and specialists. In 2001, the mean time for an initial decision to be made by the Social Security Administration was about 90 days, with substantially more time devoted to the difficult cases where the determination of physical or psychological problems is less clear-cut. Moreover, it is possible for the applicant to appeal an initial rejection, with the appeal process involving four different stages. In the end, if the applicant is awarded a benefit, there is a five-month waiting period before the individual starts receiving it. While DI is a non-means tested program, SSI also requires documentation of low income and assets to be granted the benefit. According to Daly and Burkhauer (2003), for the blind and disabled applying to the SSI program, the complex and stringent medical test denies benefits to 63 percent of all applicants.

Some benefits in the U.S. are provided through the tax system. The most notable examples are the Earned Income Tax Credit (EITC) and the Child Tax Credit. Estimates of the take-up rates for EITC are of the order of 80% in the case of EITC (Scholz, 1994; Holtzblatt and McCubbin, 2003) with possibly lower rates among new labor force entrants (Hill et al.,
While these numbers are larger than in case of many other programs, they are still surprisingly low given an apparent advantage of programs administered through the tax code in reaching eligible recipients and the presence of tax preparation industry that is interested in signing up eligibles. As elaborated by Holtzblatt and McCubbin (2003), the EITC eligibility rules are complex. Applicants must have income in the appropriate range and have qualifying children (EITC benefits to childless applicants are relatively small). Only earned income counts and too large investment income disqualifies the applicant. If there is more than one person eligible to claim a child, the one with higher income should claim it. The definition of a qualifying child is relatively straightforward for a two-parent family bringing up their own children but it becomes very complicated when parents are divorced or never married, or when grandparents or other caregivers want to claim benefits. Moreover, the rules determining eligibility for EITC, Child Care Credit and dependent exemptions are different. An obvious but important element of take up involves filing a tax return: individuals who are not otherwise required to file a tax return may not be aware of benefits awaiting them if they do. A prima facie evidence of difficulties in dealing with the tax code is the high reliance on tax preparers even among very low-income individuals. Kopczuk and Pop-Eleches (2005) report that 67% of EITC recipients rely on tax preparers despite the significant fees charged for such services.

2.3 Evidence that complexity and imperfect information are important barriers to take up

Having described the tremendous amount of complexity and administrative hassle existing in public programs, we still need to document that these aspects are important determinants of the take up of social benefits by eligible individuals. Indeed, an accumulating amount of evidence suggests that these aspects are very important for non-participation in welfare programs, and that policies affecting the amount of complexity and information can have significant effects on take-up rates.

For the Food Stamp Program, Currie and Grogger (2001) use direct measures of transaction costs to estimate their effect on participation. They consider the length of recertification intervals in the program, showing that longer intervals increase the take up of food stamps, with the effect being strongest for single parent families and somewhat smaller for rural households. They also consider the recent introduction of Electronic Benefit Transfer (EBT) systems, whereby participants in the program receive the transfer on a magnetic debit card with a PIN code. The EBT system has been hypothesized to reduce stigma, but may also affect the transaction costs incurred by recipients. Currie and Grogger find small positive effects from the adoption of EBT on the participation rates of married households with no children and on rural households. Daponte et al. (1999) conduct an experiment designed to investigate the role of information for the take up of food stamps. They find that ignorance about the program does indeed contribute to non-participation.

A couple of papers suggest that transaction costs also matter for participation in the Special Supplemental Nutrition Program for Women, Infants and Children (WIC). Based on cross-sectional data, Brien and Swann (2001) find that requiring income documentation
of WIC applicants reduced participation rates. Bitler et al. (2003) show in a panel that the frequency of visits to the WIC office, income documentation requirements, and rules about adjunctive eligibility for WIC via participation in other programs all matter for take up. Daly and Burkhauser (2003) consider the changes over time in take-up rates and caseloads for the Supplemental Security Income Program (SSI). The main explanation identified for the observed changes is variation in the stringency of the medical test and other modifications of the administrative practices in the disability determination process.

Another indication that complexity and information constitute important barriers to program enrollment is provided by evidence showing that take up tends to be higher when organizations and businesses have a stake in helping people with the application processes. For example, while the Medicaid program as a whole is characterized by substantial non-participation (cf. above), the take up among pregnant women is very close to 100 percent. Quite likely, this is related to the fact that hospitals have very strong incentives to help pregnant women enrolling in the program (Currie, 2004), because hospitals are required to serve a woman in active labor whether or not she can pay. Indeed, many hospitals have established Medicaid enrollment offices on site assisting people completing the forms and telling them how to obtain the necessary documentation.

Also on the Medicaid program, a recent study by Aizer (2004, 2003) considers a state-sponsored program in California providing application assistance to reduce the process cost of applying along with a media campaign to increase awareness of the program. The application assistance was provided by nonprofit community organizations that were paid $50 per successfully completed Medicaid application. Aizer finds a large effect of this outreach campaign on participation, especially for Hispanic and Asian children who face the highest costs enrolling due to language and immigration concerns. Access to bilingual application assistants lead to substantial increases in monthly Medicaid enrollment for these groups. The evidence by Aizer lends strong support to the process cost and information stories, whereas the welfare stigma story can hardly explain the observed effects of California’s community outreach program.

Finally, Kopczuk and Pop-Eleches (2005) demonstrate a positive effect of electronic tax filing on participation in the Earned Income Tax Credit (EITC) program. They suggest that this link is driven by the impact of electronic filing on the tax preparation industry, since it provided better opportunities for commercial tax preparers to offer their services to low-income individuals eligible for the EITC. This verifies the hypothesis that welfare program participation is higher when businesses have a stake in promoting it, suggesting again that transaction costs and imperfect information associated with applying are important. If these aspects did not play a role, it would never be optimal for a would-be EITC claimant to pay a tax preparer for his services.

3 A model of welfare program complexity

We model enrollment in public programs in the presence of complex and costly application processes and imperfect information about eligibility. We characterize program characteristics in equilibrium when policy makers can choose standard policy instruments — a benefit
level and an eligibility rule — along with an additional instrument capturing complexity. In deriving the equilibrium, we assume a ‘realistic’ objective of policy makers: poverty alleviation. Our policy objective is a natural extension of Besley and Coate (1992, 1995) who considered the design of income maintenance programs providing a minimum benefit to the poor. In our model — as in reality — the government cannot guarantee full take up, since eligible individuals may choose not to apply for the benefit, or eligible applicants may be rejected by program administrators due to imperfect testing. Hence, the objective becomes to provide a minimum standard of living for as many truly eligible as possible, i.e. with as high a take up among the deserving as possible. In general, three types of characterization errors can occur in the provision of public benefits, defined as follows:

**Definition 1 (Type Ia error)** The government makes Type Ia errors if the policy results in some truly deserving individuals not applying for benefits.

**Definition 2 (Type Ib error)** The government makes Type Ib errors if some truly deserving individuals who do apply for benefits are rejected.

**Definition 3 (Type II error)** The government makes Type II errors if some truly undeserving individuals apply for benefits and receive them.

“Deserving” individuals are those whom the government would like to provide with benefits. The meaning of this term will be unambiguous in the context of our model.

### 3.1 Setup and assumptions

We model the decision of a government that wants to provide social benefits to needy individuals. We deviate from the previous literature by taking seriously the problem of complexity inherent in the real-life programs. Complexity is the outcome of attempts to screen better. However, while better screening and the associated increase in complexity can improve the efficiency of a program, it may also discourage eligible individuals from enrolling. In order to zoom in on the participation decision, our framework abstracts from other types of behavioral responses. We begin by describing individual preferences and behavior.

#### 3.1.1 Individuals

We assume that each person is characterized by two parameters: an innate characteristic $a$ and the precision of that characteristic $\sigma$. For example, the characteristic $a$ can reflect market productivity, a natural interpretation if we are thinking about an income maintenance program. But of course the parameter can reflect other types of characteristics — say, health status 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 eligibility of the individual by obtaining a signal $\tilde{a} = a + \varepsilon/\alpha$. The noise term $\varepsilon$ reflects that testing is imperfect, and we assume that the normalized distribution of $\varepsilon/\sigma$ is characterized by the c.d.f. $P(\cdot)$ which is identical for
everyone ($\varepsilon/\sigma$ is mean zero and variance one). The parameter $\alpha$ is a policy choice affecting the precision/complexity of the test.

One possible interpretation of $\alpha$ is in terms of the number of tests. Under this interpretation, the government can observe indicators of skills. Indicator $i$ is given by $a_i = a + \varepsilon_i$ where $\varepsilon_i \sim N(0, \sigma^2)$: it is a normally distributed unbiased indicator of the true skill level with standard deviation $\sigma$. Examples of indicators are outcomes of interviews with a case worker, evaluation of a complicated set of eligibility criteria, an opinion of a medical commission regarding disability status, 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 average of $\alpha$ tests.

More generally, the policy parameter $\alpha$ determines the extent of randomness in the application process. We maintain the assumption that reducing randomness imposes additional 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 $\alpha$ 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 $\alpha$, benefits are granted to individuals who satisfy

$$\bar{a} = a + \varepsilon/\alpha < a$$

The probability that an individual with skill level $a$ and precision $\sigma$ receives benefits is therefore given by $P\left(\frac{\alpha(\bar{a} - a)}{\sigma}\right)$.

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 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 $u(\cdot)$ 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

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

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 $\alpha$, eligibility criterion $\bar{a}$ and personal precision of ability signals $\sigma$. A higher $\bar{a}$ unambiguously increases the probability. The effect of complexity and precision depends on the sign of $\bar{a} - a$. When $\bar{a} > a$, higher complexity and better precision (smaller $\sigma$) both increases 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 applying for 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 probability of receiving benefits above this critical value choose to apply for benefits, while the rest decide 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 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 \( \bar{a} \). Letting the benefit depend negatively on \( \bar{a} \) 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 \( \beta = 0 \)).

Under assumption 1, the threshold probability level for applying is given by

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

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

\[
\bar{\sigma}_a(\alpha, \bar{a}, B) \equiv \frac{\alpha(\bar{a} - a)}{P^{-1}(\tilde{P}(\alpha, B))}. \tag{6}
\]
When \( \bar{a} > a \), individuals with \( \sigma \) lower than \( \bar{\sigma}_a \) (high precision) apply for benefits. When \( \bar{a} < a \), only individuals with \( \sigma \) greater than \( \bar{\sigma}_a \) (low precision) choose to apply.

### 3.1.2 Population

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

**Remark 1** There are three qualitative cases describing the distribution of the probability of receiving benefits in the population:

1. \( \bar{a} \leq a_L < a_H \). Probabilities are in \((0, P(0)]\) and increasing in \( \sigma \) (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 \( \sigma \); probabilities for high-ability individuals are in \((0, P(0)]\) and increasing in \( \sigma \) (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 \( \sigma \).

When \( a \neq \bar{a} \), 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 \( \varepsilon \) and therefore it is a constant independent of policy parameters or individual characteristics. In the “natural” case where the likelihoods of over- and under-stating true ability are identical — so that median(\( \varepsilon \)) is zero — we have \( P(0) = \frac{1}{2} \). Recall that the threshold probability for applying \( \tilde{P} \) depends on the policy parameters \( \alpha \) 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 government implementing a policy of full separation will continue to make Type Ib errors — despite only low-ability individuals applying, some of them will be rejected. Reducing the number of Type Ib errors can be accomplished by increasing complexity $\alpha$, but in order to limit the number of 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 eliminate Type II errors. As we will demonstrate, these constraints are severe enough to justify committing Type II errors.

To complete the characterization of the assumptions about the population, we need to specify the distribution of $\sigma$. We will denote the c.d.f. of the distribution of $\sigma$ 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 perfect precision is zero. The number of individuals of type $a$ is given by $N^*_a \equiv \int_0^\infty dG_a(\sigma)$, with both $N_L$ and $N_H$ assumed to be positive and finite.

Some of our results will depend on the following technical 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 $\sigma$ has no thick tail. In particular, it precisely rules out the Pareto distribution, but it allows for distributions that have thinner tails such as, for example, 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.

### 3.1.3 Government

The final piece of the model is the specification of government’s objectives. We consider the problem of poverty alleviation. The government is interested in lifting low-ability individuals out of poverty. Its objective is to pay benefits of at least $\bar{B}$ to as many low-ability individuals as possible, subject to the budget constraint. Formally,

\[
\max_{\alpha, \bar{a}, B} \quad N_L(\alpha, \bar{a}, B) \quad \text{(7)}
\]

subject to

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

and

\[
B \geq \bar{B}. \quad \text{(9)}
\]

where $N_a(\alpha, \bar{a}, B)$ is the number of successful applicants of type $a$ as a function of policy parameters and $R$ is an exogenously given budget size.
There are a few strong assumptions made here. First, while the policy makers are interested in providing benefits, they are not accounting in their objective function for the cost that complexity imposes on individuals. Therefore, this is not a welfarist framework. We view an extension of this model to a welfare-maximization framework as interesting, but we feel that real-life policies do not take into account the cost of complexity. Therefore, we view our modeling strategy as both positive (because it reflects the status quo approach to welfare program design) and normative (as long as one subscribes to the preferences evident in the current policy design).

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 maximization of the number of deserving recipients. Absent a direct welfarist objective, providing the minimum level of well-being 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, 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, 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 need to accept a large number of Type II errors.

3.2 Results

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

**Definition 4 (First-best policy)** Suppose that it was possible to observe both $a$ and $\sigma$. The optimal policy involves providing benefits of $\bar{B}$ to $R/\bar{B}$ individuals with ability $a_L$. The choice of these individuals is undetermined (there are many first-best policies).

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

**Proof.** Subject to the benefit and revenue constraints, the government can provide benefits to at most $R/\bar{B}$ individuals. Therefore, providing benefits to any high-ability individual must result in
fewer than $R/\bar{B}$ low-ability individuals receiving benefits and is therefore a non-first-best outcome. Policies resulting in $R/\bar{B}$ low-ability recipients satisfy $a_L < \bar{a} \leq a_H$, $B = \bar{B}$ and $\tilde{P}(\alpha, B) \geq P(0)$. Any choice of policy parameters satisfying these criteria implements the first-best allocation provided that expenditures ($\bar{B}$ times the number of low-ability recipients) equal the budget $R$, so that the policy spends all the money. First-best allocations can be also implemented for smaller $R$s by increasing $\alpha$ appropriately (because a higher $\alpha$ increases $\tilde{P}$ and low-ability recipients fall to zero for high enough $\alpha$ because complexity costs outweigh then benefits). Consequently, implementing the first-best allocation is possible for a sufficiently small $R$. 

While it is possible to implement the first-best policy when the budget is small enough, notice that it is not possible to do so for a large enough budget. A first-best policy requires setting $B = \bar{B}$. Because $\tilde{P}$ is independent of $\bar{a}$, maximizing the number of low-ability applicants while preserving full separation requires $\bar{a} = a_H$. Given $\bar{a} = a_H$ and $B = \bar{B}$, however, the parameter $\alpha$ cannot be set to provide benefits to all of the low-ability individuals. This is because setting eligibility at the finite level of $\bar{a} = a_H$ provides a non-zero likelihood of rejection for any individual with $\sigma > 0$ when $\alpha$ is finite and $\alpha$ cannot be increased without bound because resulting administrative costs discourage applications (holding $B = \bar{B}$). Therefore, there is a maximum number of low-ability individuals that can be reached under the first-best policy and when the government has at its disposal a budget that allows for providing benefits to more individuals than that, the first-best allocation is no longer feasible. Still, however, it is possible to consider policies that do not involve Type II errors by allowing benefits to increase beyond $\bar{B}$.

**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 $P(\alpha, B) = P(0)$ and $\bar{a} = a_H$. For such policies, no high-ability individuals apply (because $\tilde{P} = P(0)$ and $\bar{a} \leq a_H$) while all of the low-ability individuals apply. Consider increasing $\alpha$ while simultaneously increasing $B$ to keep $P(\alpha, B) = P(0)$. By construction, this policy will retain full separation. As $\alpha \to \infty$, $P\left( \frac{a(\bar{a} - a_L)}{\sigma} \right)$ increases (and tends to one) for any $\sigma$ and therefore, because all of the low-ability individuals apply, the number of low-ability individuals receiving benefits increases. Simultaneously, we need to have $B \to \infty$ and therefore spending will be tending to $\infty$. At some point, therefore, the full budget will be spent. 

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 $\alpha$ 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 $\alpha$ 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 \( \alpha \) and \( B \) change so as to stop some people from applying: these are the people with the highest variance and the thin-tails assumption implies that there are not “many” of them.

**Proposition 2** The best policy implementing full separation when the first-best allocation is not feasible (“large budget”) is characterized by \( B > \bar{B} \), \( \tilde{a} = a_H \), \( \tilde{P}(\alpha, B) > P(0) \), 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 \( \tilde{a} = a_H \). To see this, suppose instead that \( \tilde{a} < a_H \) in the optimum. Then we can increase \( \tilde{a} \) 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 \( N_LB = R \) and a lower \( B \), so that \( N_L \) must be higher, contradicting that \( \tilde{a} < a_H \) was optimal.

Third, the optimal policy involves \( \tilde{P}(\alpha, B) > P(0) \). Conversely, suppose that \( \tilde{P}(\alpha, B) = P(0) \). Consider increasing \( \alpha \) 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 \( B > \bar{B}, \alpha = a_H \), and \( \tilde{P}(\alpha, B) > P(0) \), the problem is to maximize \( N_L(\alpha, a_H, B) \) with respect to \( \alpha \) 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 \( \alpha \). Note that \( N_L \) falls to zero as either \( \bar{B} \) or \( \alpha \) increases, therefore guaranteeing a non-degenerate interior solution. 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 \( \frac{\partial N_L}{\partial \alpha} N_L = 0 \). Recalling that the first-order condition was \( \frac{\partial N_L}{\partial \alpha} + \frac{\partial N_L}{\partial B} \frac{\partial B}{\partial \alpha} = 0 \), it also implies \( \frac{\partial N_L}{\partial \alpha} = 0 \). We are guaranteed that such a point exists because by Lemma 2, \( N_L \) is increasing in \( \alpha \) at \( \tilde{P}(\alpha, B) = P(0) \), \( N_L \) is equal to zero when \( \tilde{P}(\alpha, B) = 1 \) and \( \tilde{P}(\alpha, B) \) itself increases with \( \alpha \) (and attains the value of one for a sufficiently high \( \alpha \)). ■

This proposition has three implications. First, because \( \tilde{P}(\alpha, B) > P(0) \) both kinds of Type I error will have to be made:

**Corollary 1** The best policy that avoids Type II errors involves both Type Ia and Iib 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 involves 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 Iib 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 involves setting benefits higher than the minimum required level \( \bar{B} \). This is a mechanical result. Given that full separation imposes a restriction on \( \tilde{\alpha} \) and given that sufficiently high \( \alpha \) has to discourage applications, the only way to spend all of the budget while retaining full separation is by increasing \( B \).

Third,
Corollary 2 The optimal policy that avoids Type II errors involves maximum reliance on $\alpha$: complexity should be increased until it no longer has an effect on the number of low-ability recipients.

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 $\tilde{P}(\alpha, B) = \frac{f(\alpha)}{f'}$. 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 consume 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 utility surplus for those who get the benefit. Since some applicants are rejected in the process, aggregate complexity costs may then constitute far 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. As we have shown, however, 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 3 When the first-best allocation cannot be implemented (budget is not small), the second-best policy always involves non-separation.

This is an important result. Despite that 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. First, we establish that such a policy is indeed characterized by $\bar{a} > a_H$

Proposition 3 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 \( \bar{P} < P(0) \). Suppose a policy of this kind, denoted by \((\alpha^*, \bar{a}^*, B^*)\), was optimal. Consider then an alternative policy that keeps \( B = B^* \), set \( \bar{a} \) to satisfy \( \max\{\bar{a}, a_L\} \), and adjusts \( \alpha \) to obtain \( \bar{P} = P(0) \). 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. 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 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. ■

This proposition demonstrates that despite the fact that the government will pay some benefits to the high-ability individuals, the eligibility criteria are not very stringent and in fact the optimal policy will involve eligibility criteria that make high ability applicants explicitly eligible \( \bar{a} > a_H \). Instead they will be discouraged from applying too much by having a high complexity cost \( f(\alpha) \).

We have now established that (when the first-best allocation is not feasible) we can limit attention to policies that involve \( a > a_H \) and result in \( N_H > 0 \). 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 \( \bar{P}(\alpha, B) = P(0) \) which is the only point where it is not immediately obvious. Therefore, the maximum should satisfy the following first-order conditions (where \( \lambda \) is the Lagrange multiplier associated with the government budget):

\[
\begin{align*}
\frac{\partial N_L}{\partial \alpha} - \lambda \left( \frac{\partial N_L}{\partial \alpha} + \frac{\partial N_H}{\partial \alpha} \right) B &= 0 \quad (10) \\
\frac{\partial N_L}{\partial \bar{a}} - \lambda \left( \frac{\partial N_L}{\partial \bar{a}} + \frac{\partial N_H}{\partial \bar{a}} \right) B &= 0 \quad (11) \\
\left[ \frac{\partial N_L}{\partial B} - \lambda \left( \frac{\partial N_L}{\partial B} + \frac{\partial N_H}{\partial B} \right) B - \lambda (N_L + N_H) \right] (B - \bar{B}) &= 0, \quad (12)
\end{align*}
\]

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

\[ \frac{\partial N_L}{\partial \bar{a}} \frac{\partial N_H}{\partial \bar{a}} = \frac{\partial N_L}{\partial \alpha} \frac{\partial N_H}{\partial \alpha} = \frac{\lambda B}{1 - \lambda B}. \quad (13) \]

Neither eligibility criterion \( \bar{a} \) nor the intensity of screening/complexity \( \alpha \) 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, both instruments should screen equally well. It is also
straightforward to show that when  \( \frac{\partial N_l}{\partial \bar{a}} > \frac{\partial N_l}{\partial \alpha} \), \( \bar{a} \) should be increased and/or \( \alpha \) reduced, with the opposite implication when the sign of this inequality is reversed.

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):

\[
\frac{\partial N_a}{\partial \alpha} = \frac{\alpha - a}{\partial \bar{a}} + \frac{\partial P}{\partial \alpha} \frac{\partial N_a}{\partial B}, \text{ where } \frac{\partial P}{\partial \alpha} = -e^{\beta B} - 1 - e^{\beta (\alpha)} f'(\alpha)
\] (14)

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

We will show first that the government will still pursue policies that result in Type Ia errors (along with type Ib errors, of course).

**Proposition 4** When the first-best allocation is not feasible, for any value of \( B \), the optimal choice of \( \bar{a} \) and \( \alpha \) features \( \tilde{P}(\alpha, B) > P(0) \).

**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 \( \frac{\partial N_a}{\partial \alpha} = \frac{\alpha - a}{\partial \bar{a}} \), which implies

\[
\frac{\partial N_l}{\partial \alpha} = \frac{\alpha - a}{\partial \bar{a}} \frac{\partial N_l}{\partial \bar{a}} > \frac{\partial N_l}{\partial \bar{a}} \frac{\partial N_h}{\partial \bar{a}} > \frac{\partial N_h}{\partial \bar{a}}
\]

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

One interesting issue regarding the optimal setting of the 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. Recall however, 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. Note, however, 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. We will show first, however, that such a possibility requires a significant number of high-ability applicants.

**Proposition 5** Suppose that the first-best allocation is not feasible. For small enough budgets, setting \( B = \bar{B} \) is optimal.

**Proof.** See the appendix.
This result is tedious to prove, but the intuition 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. In the next proposition we demonstrate though that benefits are still a particularly good instrument for screening: they bring in low-ability individuals at higher rates than any of the other instruments.

**Proposition 6** Suppose that \( \alpha \) and \( \bar{a} \) are set optimally given \( B \). Then,

\[
\frac{\partial N_L}{\partial B} > \frac{\partial N_H}{\partial B} = \frac{\partial N_L}{\partial \alpha} = \frac{\partial N_L}{\partial \bar{a}} = \frac{\lambda B}{1 - \lambda B}. \tag{15}
\]

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

\[
\frac{\partial N_L}{\partial B} = \frac{\partial N_L}{\partial \bar{a}} \frac{\partial \bar{a}}{\partial \alpha} + \frac{\partial N_L}{\partial \alpha} = \frac{\partial N_L}{\partial \alpha} + \frac{\partial N_L}{\partial \bar{a}} \frac{\partial \bar{a}}{\partial \alpha} = \frac{\partial N_L}{\partial \alpha} + \frac{\partial N_L}{\partial \bar{a}} \frac{\partial \bar{a}}{\partial \alpha}.
\]

The denominator of the first term is equal to \( -\frac{\partial P/\partial a}{\partial P/\partial B} \) times \( \frac{\partial N_L}{\partial B} \) and it is positive because \( \partial \bar{P}/\partial \alpha > 0 \) while \( \partial \bar{P}/\partial B < 0 \). Therefore, the first term of the expression is unambiguously positive. When \( \alpha \) and \( \bar{a} \) are set optimally given \( B \), we have \( \frac{\partial N_L}{\partial \alpha} = \frac{\partial N_L}{\partial \bar{a}} \frac{\partial \bar{a}}{\partial \alpha} = \frac{\lambda B}{1 - \lambda B} \) 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

\[
\frac{\partial N_L}{\partial B} = \frac{\partial N_L}{\partial \alpha} + \frac{\lambda B}{1 - \lambda B} > \frac{\lambda B}{1 - \lambda B} \left( \frac{\partial N_L}{\partial \bar{a}} \right) = \frac{\partial N_L}{\partial \bar{a}}. \tag{16}
\]

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.

We will pursue the question of the optimal choice of \( B \) further. Formally, with \( \bar{a} \) and \( \alpha \) being selected optimally, benefits should be increased from some level \( B \) if

\[
\frac{\partial N_L}{\partial B} - \lambda B \left( \frac{\partial N_L}{\partial B} + \frac{\partial N_H}{\partial B} + \frac{N_L + N_H}{B} \right) > 0
\]

or, equivalently,

\[
\frac{\partial N_L}{\partial B} > \frac{\lambda B}{1 - \lambda B} \frac{\partial N_H}{\partial B} + \frac{\lambda B}{1 - \lambda B} \frac{N_L + N_H}{B}
\]

or

\[
\frac{\partial N_L}{\partial N_H} > \frac{\lambda B}{1 - \lambda B} + \frac{\lambda B}{1 - \lambda B} \frac{N_L + N_H}{B^2N_H}
\]
Recall equation (13): $\lambda_B \frac{1}{1-\lambda_B}$ 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 ($\lambda_B \frac{1}{1-\lambda_B}$ 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.

Using the first equality in formula (16) to substitute for $\frac{\partial N_L}{\partial \bar{a}}$ yields

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

Hence,

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

Recalling that $\frac{\partial N_L}{\partial \bar{a}} = \frac{\lambda B}{1-\lambda B}$ yields the final simplification

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

where the last equality used the budget identity $R = B(N_L + N_H)$. We will want to increase benefits from $\bar{B}$ if this condition holds when evaluated at $\bar{B}$ and the optimal $\alpha$ and $\bar{a}$ (at $\tilde{B}$). 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 $\frac{\partial}{\partial \alpha} \frac{\partial N_H}{\partial \bar{a}}$ is the effect of $\alpha$ 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. This condition suggests that it is possible that benefits should be using for screening purposes. The next propositions adds an extra caveat to this statement:

**Proposition 7** Suppose that the first-best allocation is not feasible. For big enough budgets, setting $B = \bar{B}$ is optimal.

**Proof.** See the appendix. ■

We have now established that it is optimal to set $B = \bar{B}$ when the budget is either small (Proposition 5) or large (Proposition 7). But what about the intermediate cases? We will first identify a sufficient condition for $B > \bar{B}$ and then show that it is satisfied in some interesting cases.

We begin by introducing some additional notation. Let us denote by $N_L^*$ the number of low-ability recipients under the best full separation policy identified by Proposition 2. Recall, that this policy required setting $B > \bar{B}$. Recall also that this policy was an improvement over the best policy that kept $B = \bar{B}$ and $\bar{a} = a_H$ (there was no policy of this kind capable of spending all of the budget), i.e. $\max_{\alpha} N_L(\alpha, a_H, \bar{B}) < N_L^*$. Similarly, when
we pick \( \bar{a} \) that is “close” to \( a_H \) and hold benefits at \( B = \bar{B} \), we will not be able to reach \( N_L^* \) individuals. Let \( \bar{a}^* = \inf \{ \max \{ N_L(\alpha, \bar{a}, \bar{B}) \geq N_L^* \} \} \), and notice that \( \bar{a}^* > a_H \) (also, \( \bar{a}^* < \infty \) because for high enough \( \bar{a} \) we can provide benefits to all low-ability individuals while setting \( B = \bar{B} \).

**Proposition 8** Suppose that the first-best policy is not feasible and that

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

Then, setting \( B > \bar{B} \) is optimal.

**Proof.** 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 \( G_L(\bar{a}^*_L) > N_L^* \) (the number of low-ability applicants which is greater than the number of recipients must be greater than \( N_L^* \)), such that \( \bar{a}^*_L > G_L^{-1}(N_L^*) \). Recall the identity \( \bar{a}^*_L = \bar{a}^*_L - a_H \). This formula is increasing in \( \bar{a} \) and therefore \( \bar{a}^*_L > \bar{a}^*_L - a_H \). Note that this lower limit is strictly positive because \( \bar{a}^* > a_H \).

Now, note that we also have an upper bound for \( \bar{a}^*_L \): We need to have at least \( N_L^* \) low-ability recipients and, with the budget \( R \), we can then have no more than \( \bar{B} - N_L^* \) high-ability recipients. Consequently, \( N_H < \frac{R}{B} - N_L^* \) while \( N_H > P(0) G_H(\bar{a}^*_L) \) (at least a share \( P(0) \) of the high-ability applicants receive benefits, because \( P(0) > P(0) \)). Consequently, \( \bar{a}^*_L < \bar{a}^*_L \frac{1}{P(0)} \left( \frac{R}{B} - N_L^* \right) \).

Putting it together we have \( \frac{\bar{a}^* - a_H}{\bar{a}^* - a_L} G_L^{-1}(N_L^*) \leq \bar{a}^*_L < \bar{a}^*_L \frac{1}{P(0)} \left( \frac{R}{B} - N_L^* \right) \). If the upper bound is lower than the lower bound, there is no \( \bar{a}^*_L \) that satisfies this condition and therefore our original assumption that \( B = \bar{B} \) is optimal must be false.  

Observe that the definitions of \( N_L^* \) and \( \bar{a}^*_L \) are based purely on the 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 — this is because \( \bar{a}^* > a_H \) and \( N_L^* \) delivers fewer low-ability recipients than \( R/\bar{B} \) which the first-best policy would deliver. Therefore, 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.

**Corollary 4** Given the distribution of low-ability applicants and parameters \( a_H, R \) and \( \bar{B} \) such that the first-best policy cannot be implemented, there exist distributions of high-ability individuals for which the optimal policy involves \( a > a_H \) and \( B > \bar{B} \).

---

\[ \textbf{Remarks:} \]

1. There is no inconsistency with \( B = \bar{B} \) for small \( R \) — as we reduce \( R \), \( a^* \rightarrow a_H \) 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 \( \bar{B} \) allows for setting higher \( a \) while holding \( \bar{P} \) constant. With the same \( \bar{P} \) but higher \( a \), 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 \( \bar{B} \) we can guarantee the existence of \( \sigma^H \) that would satisfy the inequality.

24
All that is required is selecting the distribution so that there are more than \( \frac{1}{P(0)} (R/B - N^*_L) \) high-ability individuals with \( \sigma \) smaller than \( a^*_H - a_L G_L^{-1}(N^*_L) \). This is a requirement imposed on \( G_H \) at a particular strictly positive point. The second corollary is an obvious 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 5** Fix parameters of the problem other than the high-ability distribution. Select some distribution of high-ability individuals \( G^0_H(\sigma) \) (with the corresponding number of high-ability individuals \( N^0_H \)) and consider a class of distributions \( G^\eta_H(\sigma) = \eta G^0_H(\sigma) \) (with the corresponding number of high-ability individuals \( \eta N^0_H \)). For high enough \( \eta \), setting \( B > \bar{B} \) is optimal.

**Proof.** For sufficiently high \( \eta \),

\[
G^\eta_H \left( \frac{\bar{a}^* - a_H}{\bar{a}^* - a_L} G_L^{-1}(N^*_L) \right) = \eta G^0_H \left( \frac{\bar{a}^* - a_H}{\bar{a}^* - a_L} G_L^{-1}(N^*_L) \right) \geq \frac{1}{P(0)} \left( \frac{R}{B} - N^* \right),
\]

and this condition is equivalent to the inequality in Proposition (8).

4 Discussion and implications

In this paper, we have stressed the importance of transaction costs and imperfect information for low participation rates in public programs. The only previous work to have modeled welfare program participation is the Moffitt (1983) well-known analysis of welfare stigma. His model distinguishes between a flat component of stigma — a fixed cost associated with program participation — and 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 participation. Moreover, the positive effect of benefits on take up may also be consistent with the presence of imperfect information. 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 policy instrument, while 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.4 If stigma is the

---

3In particular, there is no restriction imposed on the properties of \( G_H(\cdot) \) around zero so that we can pick a distribution that satisfies our technical assumption 3.

4For 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.
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. On the other hand, once we adopt the view that 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 eligibility criteria. It also follows that we should view high complexity and low 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 presented a model to explain the existence of complexity and low take up as an equilibrium. It has been long 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 intended recipients.

More generally, our paper represents a first attempt to incorporate administrative complexity as a choice variable in policy analysis. While it has long been recognized that complexity is a very 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 has an obvious application to the analysis of tax policy and tax evasion. We have already discussed the Earned Income Tax Credit (EITC) which is a quite complex tax credit program with substantial non-participation. But our model can be used to think about the design of tax deductions and exemptions more generally, where the legislation tend to be quite complex. Most likely, this complexity of the tax code reflects exactly 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 welfare policy and take-up behavior in the United States and across different countries. Firstly, the model can explain the occurrence of different take up across “equally” eligible individuals resulting from heterogeneity in the variance of observed skill (or health, nutritional risk, disability, etc.). In the model equilibrium, those with low variances face high acceptance rates in the application process making it optimal for them to apply for the benefit. By contrast, equally eligible individuals with high variances face high risks of being rejected by program administrators and may find it optimal not to apply at all. In other words, these are the individuals 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, and so on, would all contribute to creating more uncertainty in the test. This would seem to explain why, for example, newly arrived 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. Hence, stigma is not a policy instrument per se. Our understanding of the effect of different policies on the preferences for social benefits is still lacking.
immigrants are characterized by a lower take up of social benefits, conditional on being eligible, than immigrants who are more assimilated into society (e.g. Borjas and Hilton, 1996). It may also explain why some ethnic groups (such as Hispanics and Asians) have lower take up than others (like African Americans), e.g. (Currie, 2003).

Secondly, the model can explain the observation of large differences in complexity and participation rates across different programs. In equilibrium, program characteristics and take up depend on the distribution of true skills, the distribution and precision of observed skills, the program budget $R$, and the size of the minimum benefit $\bar{B}$. In other words, different programs have different designs and take up because they serve different populations (say, the poor versus the disabled), because they involve different kinds of eligibility tests (say, income versus disability tests), and because of differences in the size and generosity of programs. Other things being equal, programs with a small budget and a generous benefit, i.e. $R/\bar{B}$ being small, is predicted to have a stringent eligibility rule, high complexity, a high occurrence of type I errors (low take up), 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) described in Section 2. For programs with a larger value of $R/\bar{B}$, the model predicts that eligibility rules are less stringent while complexity is still high, and that both types of classification error occurs (although take up is now higher). This would seem to reflect a program like the Disability Insurance Program (DI), which is a large program involving complex and rigorous testing, and where both type I and type II occur (Benítez-Silva et al., 2004). Finally, our model predicts that programs with a very large $R/\bar{B}$ will be designed as universal benefits with 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.

Thirdly, consistent with the evidence in Section 2.2, we find that complexity and administrative costs can be very high in equilibrium, especially for governments eager to avoid giving money to the “undeserving.” Interestingly, under certain simplifying assumptions — low risk aversion and the noise of observed skill being distributed with zero median — we can show that the costs of complexity on each individual constitute more than 50 percent of the potential benefit. As some applicants are rejected in the process, transaction costs may then be far larger than half of the amount of benefits granted.

Fourthly, our results seem broadly consistent with observed policy design across different countries. In Anglo-Saxon countries, we observe highly targeted, highly complex public programs involving substantial non-participation. In Northern-Continental European countries, and especially in Scandinavia, universal programs are more common and the programs that do involve targeting tend to rely on fewer and simpler eligibility criteria (typically low income and assets). Complexity and administrative hassle seem less important there, and low take up is a non-issue. This is consistent with our prediction for right-wing (“Anglo-Saxon”) versus left-wing (“Scandinavian”) governments.

---

Immervoll et al. (2004) provide a description of transfer policies across the fifteen pre-2004 expansion member countries of the European Union.
Our simple model abstracts from several aspects that we would like to discuss briefly. (i) In practice, there will be revenue costs from the processing of each application due to the paper work involved, the time spent by administrators and specialists for interviewing and testing, and so on. Other things being equal, this favors policies capable of increasing the number of deserving recipients with a relatively small (or no) accompanying increasing in the number of applicants. This would seem to improve the effectiveness of complexity as a policy instrument, since it identifies the truly eligible applicants with more precision while making it more costly for would-be applicants to claim the benefit. Given our results of benefits being a particularly good screening instrument, it could also increase the likelihood that higher benefits are paid in equilibrium. (ii) On the other hand, there might also be a direct revenue cost of increasing the complexity and precision of the application process. A better test costs more money, which ceteris paribus make complexity less effective as a policy instrument. (iii) In the 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. If politicians cared about utility instead of income, as in a model of social welfare maximization, complexity may turn out to be a less effective instrument. Like the Besley and Coate papers, our paper should probably 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.

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 $\sigma$ (precision of observed skill) reflects the degree of irrationality. An individual with a high $\sigma$ face a lot of uncertainty when being tested, even if he is truly eligible, because he tend 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-$\sigma$ 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 $\sigma$). 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
\[
N_a(\bar{a}, \alpha, B) = \int_0^\infty P\left(\frac{\alpha(\bar{a} - a)}{\sigma}\right) dG_a(\sigma) \tag{A.1}
\]
and we want to evaluate $\frac{\partial N_a}{\partial a}$ and $\frac{\partial N_a}{\partial B}$ when $\bar{a} > a$ and $\bar{P}(\alpha, B) = P(0)$. Begin with the first of these. Differentiating explicitly yields
\[
\frac{\partial N_a}{\partial a} = \int_0^\infty p(\cdot) \frac{\bar{a} - a}{\sigma} dG_a(\sigma) + \frac{\partial \bar{\sigma}_a}{\partial a} \bar{P}(\alpha, B) g_a(\bar{\sigma}_a). \tag{A.2}
\]
The first term is unambiguously positive. We will show that the second term is zero when $\bar{a} > a$ and $\bar{P}(\alpha, B) = P(0)$. Evaluating $\frac{\partial \bar{\sigma}_a}{\partial a}$ yields
\[
\frac{\partial \bar{\sigma}_a}{\partial a} = \frac{(\bar{a} - a)^{P-1} - \alpha(\bar{a} - a)}{(P - 1)(\bar{P}(\alpha, B))^2} \frac{\partial \bar{P}}{\partial a} = \frac{\bar{\sigma}_a}{\alpha} - \frac{\bar{\sigma}_a^2}{\alpha(\bar{a} - a)} \left(\frac{\alpha(\bar{a} - a)}{\bar{\sigma}_a}\right)^{-1} \frac{\partial \bar{P}}{\partial a}. \tag{A.3}
\]
Note that $\lim_{P(\alpha, B) \to P(0)} \bar{\sigma}_a = \infty$: 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): $\alpha$, $\bar{a} - a$, $\frac{\partial \bar{P}}{\partial a}$, and $p(\cdot)$ are positive and finite (the argument of $p(\cdot)$ goes to zero as $\bar{\sigma}_a$ goes to infinity and density at 0 is positive). Consequently, as we approach $P(0)$ with $\bar{P}$, expression (A.3) tends to $-\infty$ at the rate of $\bar{\sigma}_a^2$. 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 $\frac{\partial \bar{\sigma}_a}{\partial B} \bar{P}(\alpha, B) g_a(\bar{\sigma}_a)$ and we have $\frac{\partial \bar{\sigma}_a}{\partial B} = -\frac{\alpha(\bar{a} - a)}{(P - 1)(\bar{P}(\alpha, B))^2} \frac{\partial \bar{P}}{\partial B} = -\bar{\sigma}_a^2 \frac{1}{2(\bar{a} - a)} \frac{\partial \bar{P}}{\partial B}$. 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 $\bar{P}(\alpha, B) < P(0)$ 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 $\bar{P}(\alpha, B) < P(0)$, so that it is also continuous at $\bar{P}(\alpha, B) = P(0)$. $\blacksquare$

Proof of Lemma 3. Denote by $(\alpha^*, a_H, B^*)$ 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 $(\alpha^*, a_H, B^*)$: $\frac{\partial N_H(\alpha^*, a_H, B^*)}{\partial a} = 0$. Differentiating (A.1) with respect to $\bar{a}$ yields
\[
\frac{\partial N_a}{\partial a} = \int_0^\infty p\left(\frac{\alpha(\bar{a} - a)}{\sigma}\right) \frac{\alpha g_a(\sigma)}{\sigma} d\sigma + \frac{\partial \bar{\sigma}_a}{\partial a} \bar{P}(\alpha, B) g_a(\bar{\sigma}_a), \tag{A.4}
\]
such that
\[
\frac{\partial N_H(\alpha^*, a_H, B^*)}{\partial a} = \lim_{\bar{a} \to a_H} \int_0^\infty p\left(\frac{\alpha(\bar{a} - a)}{\sigma}\right) \frac{\alpha g_H(\sigma)}{\sigma} d\sigma + \frac{\partial \bar{\sigma}_a}{\partial a} \bar{P}(\alpha^*, B^*) g_{H(\alpha^*, \bar{a}, B^*)}. \tag{A.5}
\]
Note that $\lim_{\bar{a} \to a_H} \bar{\sigma}_a = 0$, and therefore the second term is zero: we have $g(0) = 0$ while the other components $\bar{P}$ and $\frac{\partial \bar{\sigma}_a}{\partial a} = -\frac{1}{2(\bar{a} - a)} \frac{\partial \bar{P}}{\partial a}$ tend to finite limits (the limit of $P^{-1}(\cdot)$ is positive because we are considering a point with $\bar{P} > P(0)$). In the first term, the integrand $p(\cdot) \frac{\alpha g_H(\sigma)}{\sigma}$ 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 $\partial L / \partial \sigma_\alpha > 0$, so that the derivative $\bar{\sigma}_L$ given by A.4 is strictly positive.

Hence, starting at the best full-separation policy $(\alpha^*, a_H, B^*)$, 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 identity 14.** Recall the definition of $N_a$, equation (A.1) and the definition of $\bar{\sigma}_a$ in equation (6). $\alpha$ 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 $\alpha$ 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 $\alpha$ as a combination of the effects with respect to the other two probabilities.

**Proof of Proposition 5.** Let $R^*$ be the maximum budget that allows for implementing the first-best allocation and consider $R = R^* + \varepsilon$. Note that the optimal policy at $\varepsilon = 0$ involves $a = a_H$ (otherwise increasing $a$ would increase the number of low-ability recipients and allow for implementing the first-best at an even greater budget), $B = \bar{B}$ and $P > P(0)$ (at $P = P(0)$ increasing $\alpha$ allows for increasing the number of low-ability recipients by Lemma 2 and therefore implementing the first-best for an even greater budget).

We will show first that $\lim_{\varepsilon \to 0} \bar{\sigma}_H = 0$. To see that note that we can achieve $R^* + \varepsilon > N_L > \frac{R^*}{B}$ by sticking to the full separation policies (we can always not use all of the money and stick to full separation with $B = \bar{B}$, but in fact we have shown earlier that we can improve upon this policy while keeping full separation by increasing $B$ slightly). Because we will have $a_H > a$ we will provide some benefits to high-ability individuals and if we pay them more than $\varepsilon$ in total it will mean that we cannot support more than $\frac{R^*}{B}$ low-ability individuals which cannot be optimal. Therefore, $N_H < \frac{R^*}{B}$.

By definition $N_H = \int_0^{\bar{\sigma}_H} P \left( \frac{\alpha(\bar{a} - a_H)}{\sigma} \right) dG_H(\sigma_H)$. We know that $P \left( \frac{\alpha(\bar{a} - a_H)}{\sigma} \right) > P(0)$ for everyone because $\bar{a} > a_H$. Therefore, $N_H > G_H(\bar{\sigma}_H)P(0)$ and consequently $G_H(\bar{\sigma}_H) < \frac{N_H}{P(0)} < \frac{\varepsilon}{B(0)}$, implying that $\lim_{\varepsilon \to 0} \bar{\sigma}_H = 0$.

Now observe that $\lim_{\varepsilon \to 0} \bar{\sigma}_H = 0$ implies $\bar{a} = a_H$.

Recall that $\bar{\sigma}_H = \frac{\alpha(\bar{a} - a_H)}{P'(\alpha, B)}$ and $\bar{\sigma}_L = \frac{\alpha(\bar{a} - a_L)}{P'(\alpha, B)}$, so that $\bar{\sigma}_H = \frac{\alpha(\bar{a} - a_H)}{P'(\alpha, B)} = \frac{\alpha(\bar{a} - a_L)}{P'(\alpha, B)} \bar{a}$. Therefore, $\bar{a} - a_H = \frac{\alpha(\bar{a} - a_L)}{P'(\alpha, B)} a_H - a_L$. Note also that $G_L(\bar{\sigma}_L) > R^* / B$ (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 $\bar{\sigma}_H \to 0$, $\bar{a} - a_H \to 0$ and therefore $\bar{a} - a_H \to 0$.

Next, we will demonstrate that $\lim_{\varepsilon \to 0} \bar{\sigma}_H(\varepsilon) > P(0)$. Consider what happens when $\varepsilon \to 0$. Denote the optimal policy given $\varepsilon$ by $(\alpha(\varepsilon), \bar{a}(\varepsilon), B(\varepsilon))$ and denote by $(\alpha^*, \bar{a}^*, B^*)$ the optimal values in the full separation case $\varepsilon = 0$. The resulting value of the objective is $N_L(\varepsilon)$. $N_L$ is a continuous function in the relevant region. Suppose that $\lim_{\varepsilon \to 0} \bar{\sigma}_H(\varepsilon) \neq (\alpha^*, \bar{a}^*, B^*)$. 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 $P(\alpha(\varepsilon), B(\varepsilon)) > P(0)$ (by Proposition 4), $\bar{a}(\varepsilon) \to a_H$ 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 $\varepsilon$ we would have been better off using the full separation policy (and not using all of the money). Consequently, $\lim_{\varepsilon \to 0} (\alpha, \bar{a}, B) = (\alpha^*, \bar{a}^*, B^*)$ and in particular $\lim_{\varepsilon \to 0} \bar{\sigma}_H(\varepsilon) = \bar{\sigma}_H(0) = \bar{\sigma}_H(\alpha^*, \bar{a}^*, B^*) > P(0)$.

\footnote{\textit{Note:} $N_a$ is continuous when $a > a_L$ and has a discontinuity at $\bar{a} = a$ when $P = P(0)$. In this case, the discontinuity is at $a_L$, but we are considering $\bar{a} \geq a_H > a_L$.}
Denote by $\bar{\lambda}$ the Lagrange multiplier from the problem of maximizing the objective function with respect to $\alpha$ 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

$$\frac{\partial N_L}{\partial B} > \frac{\bar{\lambda}B}{1 - \bar{\lambda}B} \frac{\partial N_H}{\partial B} + \frac{\bar{\lambda}B}{1 - \bar{\lambda}B} \frac{N_L + N_H}{B} \quad (A.5)$$

The left-hand side is non-negative and we don’t have to worry about it increasing without bounds as $\varepsilon$ changes — it converges to a finite limit of $\frac{\partial N_L}{\partial B}(\alpha^*, \alpha, \bar{B})$. All terms on the right-hand side are non-negative and $\frac{N_L + N_H}{B}$ is finite and bounded away from zero ($N_L > \frac{B}{\bar{B}}$ and $B = \bar{B}$). We will show that $\frac{\bar{\lambda}B}{1 - \bar{\lambda}B} \rightarrow \infty$ as $\varepsilon \rightarrow 0$ and thus this inequality is violated for small enough $\varepsilon$. To see that, recall that $\frac{\bar{\lambda}B}{1 - \bar{\lambda}B} = \frac{\partial N_L / \partial a}{\partial N_H / \partial a}$. We will show that the numerator is finite while the denominator falls to zero. To see that, write explicitly $\frac{\partial N_a}{\partial a}$:

$$\frac{\partial N_a}{\partial a} = \int_0^\infty \exp \left( \frac{\alpha(\bar{a} - a)}{\sigma} \right) \frac{g_a(\sigma)}{\sigma^2} d\sigma + \frac{\alpha}{P^{-1}(\bar{P}(\alpha, B))} \bar{P}(\alpha, B) g_a(\bar{\sigma}_a) \quad (A.6)$$

All terms here are non-negative. The first-term vanishes for the high-types as $\varepsilon \rightarrow 0$, because $\lim_{\varepsilon \rightarrow 0} \bar{\sigma}_H = 0$ 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 $\varepsilon \rightarrow 0$. We have just shown that $\lim_{\varepsilon \rightarrow 0} \bar{P}(\alpha, B) = \bar{P}(\alpha^*, B^*) > P(0)$, so that $P^{-1}(\bar{P}(\alpha, B))$ 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 a limit $\lim_{\varepsilon \rightarrow 0} \frac{\bar{\lambda}B}{1 - \bar{\lambda}B} = \lim_{\varepsilon \rightarrow 0} \frac{\partial N_L / \partial a}{\partial N_H / \partial a} = \infty$, implying that the inequality (A.5) must be violated for sufficiently small $\varepsilon$ and therefore $B = \bar{B}$ is optimal for sufficiently small $\varepsilon$.

**Proof of Proposition 7.** This is an implication of condition (17) evaluated at $\bar{B}$ and the optimal $\alpha$ and $\bar{a}$. To see this, hold $\bar{B}$ constant and increase $R$. The parameter $\alpha$ is bounded by $P(0) < \bar{P}(\bar{\alpha}, \bar{B}) < 1$, so that $\frac{\partial P / \partial a}{\partial P / \partial B} = -\frac{\partial a}{\partial \sigma} \frac{1}{\bar{P}(\alpha, B)} \tilde{f}'(\alpha)$ is bounded away from zero when evaluated at the optimal $\alpha$ and $\bar{B}$.

Moreover, we can show that $\frac{\partial N_a / \partial a}{\bar{\sigma}_a} \rightarrow 0$ 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 $\bar{R} \geq \bar{B}(N_L^* + N_H^*)$ 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 $\bar{a} \rightarrow \infty$ and $\bar{\sigma}_a \rightarrow \infty$.

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

$$\int_0^\infty \left( \frac{\alpha(\bar{a} - a)}{\sigma} \right) g_a(\sigma) d\sigma = \frac{1}{\bar{a} - a} \left\{ - P \left( \frac{\alpha(\bar{a} - a)}{\sigma} \right) \sigma g_a(\sigma) \bigg|_0^{\infty} + \int_0^\infty P \left( \frac{\alpha(\bar{a} - a)}{\sigma} \right) [g_a(\sigma) + \sigma g_a'(\sigma)] d\sigma \right\}$$

$$= \frac{1}{\bar{a} - a} \left\{ 1 - \bar{P}(\alpha, B) \bar{\sigma}_a g_a(\bar{\sigma}_a) + \int_0^\infty P \left( \frac{\alpha(\bar{a} - a)}{\sigma} \right) [g_a(\sigma) + \sigma g_a'(\sigma)] d\sigma \right\}.$$  

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 $\int_0^\infty g_a(\sigma) + \sigma g_a'(\sigma) d\sigma = \sigma g_a(\sigma) \big|_0^\infty = \bar{\sigma}_a g_a(\bar{\sigma}_a)$ and therefore also disappears as $\bar{\sigma}_a$ gets large. Consequently, $\frac{\partial N_a / \partial a}$ 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.
References


