

NBER WORKING PAPER SERIES

DESIGNING OPTIMAL DISABILITY INSURANCE:
A CASE FOR ASSET TESTING

Mikhail Golosov
Aleh Tsyvinski

Working Paper 10792
<http://www.nber.org/papers/w10792>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2004

The views expressed herein are those of the author(s) and not necessarily those of the National Bureau of Economic Research.

©2004 by Mikhail Golosov and Aleh Tsyvinski. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Designing Optimal Disability Insurance: A Case for Asset Testing
Mikhail Golosov and Aleh Tsyvinski
NBER Working Paper No. 10792
September 2004
JEL No. E6, H2, H3

ABSTRACT

The paper analyzes an implementation of an optimal disability insurance system as a competitive equilibrium with taxes. The problem is modeled as a dynamic mechanism design problem in which disability is unobservable. We show that an asset-tested disability system in which a disability transfer is paid only if an agent has assets below a specified maximum implements the optimum. The logic behind the result is as follows: we show that an agent who falsely claims disability has higher savings than a truly disabled agent, and an asset test prevents false claimants from receiving disability. We also evaluate welfare benefits of asset testing. For a calibrated economy, we numerically compare the optimal system to the best system without asset testing. We find that gains of asset testing are significant and equal to about 0.65% of consumption.

Mikhail Golosov
Department of Economics
Massachusetts Institute of Technology
50 Memorial Drive
E52-391
Cambridge, MA 02142
golosov@mit.edu

Aleh Tsyvinski
Department of Economics
UCLA
Box 951477
Los Angeles, CA 90095-1477
and NBER
aleh@econ.ucla.edu

Designing Optimal Disability Insurance: A Case for Asset Testing

Mikhail Golosov and Aleh Tsyvinski*

September 11, 2004

Abstract

The paper analyzes an implementation of an optimal disability insurance system as a competitive equilibrium with taxes. The problem is modeled as a dynamic mechanism design problem in which disability is unobservable. We show that an asset-tested disability system in which a disability transfer is paid only if an agent has assets below a specified maximum implements the optimum. The logic behind the result is as follows: we show that an agent who falsely claims disability has higher savings than a truly disabled agent, and an asset test prevents false claimants from receiving disability. We also evaluate welfare benefits of asset testing. For a calibrated economy, we numerically compare the optimal system to the best system without asset testing. We find that gains of asset testing are significant and equal to about 0.65% of consumption.

1 Introduction

The Social Security Disability Insurance (SSDI) program is one of the largest social insurance programs in the United States. In 2001, the program provided income to more than 6 million individuals, who accounted for 14 percent of Social Security beneficiaries. The program cost \$61 billion and constituted 15 percent of Social Security benefits. The size of the program far surpasses spending on unemployment insurance, food stamps, or any other similar program (SSA 2000).

*Golosov, MIT; Tsyvinski, UCLA, Harvard, and NBER. We thank audiences at Minneapolis Federal Reserve Bank, Chicago GSB, Berkeley, Princeton, Harvard, MIT, Northwestern, University of Pennsylvania, Carnegie-Mellon, Wharton Business School, Columbia Business School, McGill, Rochester, UCLA, SITE 2003, NBER Summer Institute 2003, and SED 2001. We are grateful to Daron Acemoglu, Andy Atkeson, George-Marios Angeletos, Marco Bassetto, Moshe Buchinsky, Hal Cole, Casey Mulligan, Hugo Hopenhayn, Larry Jones, Patrick Kehoe, Robert Lucas, Jr., Lee Ohanian, Chris Phelan, Nancy Stokey, Ivan Werning and especially V.V. Chari and Narayana Kocherlakota for their comments.

Except for SSDI, few other options provide protection against disability risk. For example, only 25 percent of private-sector employees receive long-term disability coverage (SSA 2001).

In this paper we first show how a dynamic optimal social insurance mechanism can be implemented with a simple, realistic set of tax instruments – an asset-tested disability system. Second, we carefully parametrize the model, compute the optimal system, and show that asset testing is quantitatively superior to programs in which asset testing is not allowed.

As in the classic work of Diamond and Mirrlees (1978, 1986), we assume that it is impossible to know whether an individual is truly disabled, and that disability is a permanent state. Given these assumptions, we solve a dynamic mechanism design problem and provide theoretical and numerical characterizations of the social optimum. A key feature of the optimal allocation is that intertemporal decisions of agents must be distorted.

The first goal of the paper is to find a simple tax system that will implement the optimal allocation. By *implementation* we mean finding a tax system such that a solution to a competitive equilibrium problem with taxes coincides with the optimal solution. We first show that a simple system conjectured by Diamond and Mirrlees (1978), consisting of a linear tax that is equal to the intertemporal wedge in the optimal allocation, does not implement the optimum. Then we propose a very simple tax system implementing the optimum: an asset-tested disability program. An asset test is a form of a means test in which a person receives a disability transfer only if his assets are below a specified threshold. The logic behind the result is as follows: we show that an agent who falsely claims disability has higher savings than a truly disabled agent, and an asset test prevents false claimants from receiving disability.

To evaluate advantages of asset testing, we provide estimates of welfare gain obtained by shifting from the best program without asset testing to the optimal program. In a calibrated model economy, we found a significant welfare gain of using asset tests that is equal to 0.65 percent of consumption. The welfare gain is large because asset testing allows for significantly higher disability transfers compared to the program without asset testing.

Recent empirical evidence supports our argument that persons who falsely claim disability have higher savings than disabled persons. A comprehensive study of disability applicants and recipients by Benitez-Silva, Buchinsky, and Rust (2004) finds (see Table 4 in their paper) that non-disabled awardees of disability insurance have significantly higher assets (87,017 dollars) than disabled recipients (73,911 dollars). If the value of houses and cars is excluded from the calculations of the net worth, then the difference between the assets of the truly disabled and "impostors" is

even larger¹.

Asset-tested programs, such as Medicaid, Temporary Assistance to Needy Families (TANF), and many others, are used widely in U.S. social programs. It is recognized that asset testing has important implications for life cycle savings (see Hubbard, Skinner, and Zeldes 1995). Feldstein (1985) argued for incorporating asset testing as a part of Social Security benefits in a model with myopic agents. Our paper is the first paper, to our knowledge, that derives optimality of asset testing in a mechanism design environment. Since asset testing is a particular class of a means test, this paper also extends the results on the optimality of means-tested programs (Besley and Coate 1995) and optimal income programs (Saez 2002) to the dynamic setting.

This paper also contributes to the study of optimal dynamic social insurance programs (Wang and Williamson (1996) and Hopenhayn and Nicolini (1997)). The focus in these papers was on finding optimal allocations rather than on tax systems implementing them. The difficulty in constructing realistic transfer systems that we highlight in this paper is present in other dynamic mechanism design models, such as models of optimal unemployment insurance. The techniques of implementation in the presence of private savings that we develop in this paper can be used in those settings.

Now we briefly discuss the main technical points of the paper. The optimal allocation features an intertemporal wedge as in Golosov, Kocherlakota, and Tsyvinski (2003) who showed that intertemporal distortion of savings is a feature of the optimum for a general class of dynamic taxation problems. The simplicity of the environment here allows us to find a realistic set of taxes implementing the optimum. Albanesi and Sleet (2003) and Kocherlakota (2003) study a more general version of the dynamic taxation problem and propose implementation with wealth and labor taxes.

We show that an asset-tested disability system is simple yet sufficient to implement the optimum. We first show that a simple system conjectured by Diamond and Mirrlees (1978), consisting of a linear tax that is equal to the intertemporal wedge in the optimal allocation, does not implement the optimum. Intuitively, this tax can control only single but not joint deviations. Given that an agent has correct savings, he chooses to work when able. Given that he works when able, he chooses correct savings. We show that an agent, faced with such a linear savings tax, chooses to jointly deviate: choose not to work when able and choose a different amount of savings than that required to implement the optimum. This subtle reasoning illustrates difficulties in implementing

¹We are grateful to Moshe Buchinsky for extensive discussions of assets of disabled and non-disabled recipients and pointing us to this data.

optimal dynamic taxation or social insurance programs and the intuition behind implementation using asset tests.

We parameterize the model by solving a stylized version of the current social insurance system in the United States and numerically evaluate features of the optimal allocations and welfare gains of asset testing. The best program without asset testing is equivalent to the solution of the optimal program with hidden savings. The welfare gain of asset testing is thus the difference between the optimal program with and without hidden savings. We compare the welfare from the optimal system to that obtained under the optimal system with hidden savings. Computation of a disability model with hidden savings is difficult as the dynamic first-order approach pioneered by Werning (2001) and Abraham and Pavoni (2003) is not applicable in our framework. The simplicity of the model that we chose, however, allowed us to compute optimal allocation with hidden savings without relying on the first-order approach.

The key to our analysis is that we assume disability is unobservable and permanent. In practice, determining disability status proves to be very difficult. Multiple medical and vocational factors are taken into account when determining whether an individual is eligible for disability benefits. However, even the determination of medical factors is often subjective. In 2001, a share of awards to applicants with difficult-to-verify criteria, such as mental disorders (mainly mental stress and excluding retardation) and diseases of the musculoskeletal system (typically back pain), constituted around 50 percent of total awards. Disability is a fairly permanent state. For example, only less than 5 percent of those who start receiving disability benefits return to work². For a detailed discussion of difficulties in determining disability and the data on the number of people leaving disability, see Bound and Burkhauser (1998).

The rest of the paper is structured as follows: In the next section, we describe the setup of the model. In Section 3, we provide a theoretical characterization of the optimum. In Section 4, we discuss implementation of the optimum. In Section 5, we provide numerical results. Finally, we conclude in Section 6.

²A low number of disabled returning to work does not necessarily mean that disability is a permanent state. It could indicate, for example, generosity of benefits. However, a very low number of those returning to work gives us confidence that modelling disability as a permanent state is fairly realistic.

2 Setup

In this section we describe the basics of the model. An agent lives for T periods and has preferences defined over lifetime consumption and labor which are represented by a utility function of the form

$$E \sum_{t=1}^T \beta^{t-1} [u(c_t) + v(l_t)],$$

where E denotes expectation operator, $0 < \beta < 1$, and c_t and l_t denote the period t consumption and labor of an agent. We assume that $u' > 0$, $u'' < 0$, and $v' < 0$.

An agent can become disabled in period t , and his skill θ_t is then equal to zero. We assume that disability is an absorbing state and that once disabled, an agent stays disabled for the rest of his life. Skills of able agents evolve deterministically over time.

Because disability is an absorbing state, we need to keep track only of the agent's age and the age at which he became disabled. In period t , a history h^t belongs to a set $\{1, 2, \dots, t, A\}$. We will say that $h^t = A$ if an agent is able or $h^t = j$ if an agent became disabled at age j . We denote the probability, as of period 1, of a particular history h^t by $\mu(h^t)$. The probability of history h^{t+1} conditional of the realization of history h^t is denoted $\mu(h^{t+1}|h^t)$. We denote consumption of an able agent of age t as $c_t(A)$, and his effective labor as $y_t(A)$. Consumption in periods $t+i$ ($0 \leq i \leq T-t$) of an agent who became disabled at age t is denoted by $c_{t+i}(t)$.

The agent's disability status is unobservable³. When an agent with skill θ_t works l_t units of time, he produces $y_t = \theta_t l_t$ units of effective labor. Effective labor y_t is observable, while the skill θ_t and labor l_t are the agent's private information. The informational problem is also dynamic because an agent learns whether he is disabled at the beginning of period t and not before.

We assume that the disutility of labor if an agent is disabled is equal to infinity, and the agent chooses to supply zero units of labor. If an agent provides a positive amount of effective labor, the planner knows that he is able. However, if an agent does not work, the planner does not know whether the agent is truly disabled, or able but pretending to be disabled. If a person is disabled and has a skill $\theta = 0$ we normalize $v(\cdot) = 0$. We also normalize $v(0) = 0$, i.e., if an able agent chooses not to work, then his utility from labor is equal to zero.

We consider a setting in which the net interest rate R and the wage w are constant over time

³In the working paper version of the paper (Golosov and Tsyvinski 2003) we also consider a case of imperfectly observable disability.

and assume that $\beta = \frac{1}{1+R}$. Since consumption is observable we can restrict our attention to an aggregate intertemporal feasibility constraint rather than consider agent-specific constraints. An allocation of consumption and effective labor (c, y) is *feasible* if and only if

$$\sum_{t=1}^{t=T} \sum_{h^t} \frac{1}{(1+R)^{t-1}} \mu(h^t) c_t(h^t) \leq \sum_{t=1}^{t=T} \sum_{h^t} \frac{1}{(1+R)^{t-1}} \mu(h^t) w y_t(h^t). \quad (1)$$

Allocations must respect incentive-compatibility conditions because the age at which an agent becomes disabled is private information. Since the skill of disabled agents is equal to zero, only able agents can misreport their skills. In particular, for an agent who lives T periods, the number of ex post incentive constraints is T because disability is an absorbing state, and once an agent announces that he is disabled, he cannot claim to be able to work in the future.

We define incentive compatibility constraints recursively. We denote by $V^d(t)$ the utility of an agent who becomes disabled at time t :

$$V^d(t) = \sum_{s=t}^T \beta^{s-t} u(c_s(t)).$$

Now we recursively define the expected utility of an agent who is able at time t . At time T the utility of an able agent is equal to

$$V^a(T) = u(c_T(A)) + v\left(\frac{y_T(A)}{\theta_T}\right).$$

At time $t < T$ the expected utility is defined by:

$$\begin{aligned} V^a(t) &= u(c_t(A)) + v\left(\frac{y_t(A)}{\theta_t}\right) + \\ \mu(h^{t+1} = A | h^t = A) V^a(t+1) &+ \mu(h^{t+1} = t+1 | h^t = A) V^d(t+1). \end{aligned}$$

Incentive compatibility constraint then becomes

$$V^a(t) \geq V^d(t), \text{ for } 1 \leq t \leq T. \quad (2)$$

The incentive-compatibility constraints state that in each period the expected utility of working is higher than the utility of claiming disability.

A social planner maximizes the utility of the representative agent and solves the following

programming problem (P):

$$\max_{c, y \geq 0} \mu(h^1 = A)V^a(1) + \mu(h^1 = 1)V^d(1)$$

subject to (1) and (2).

In each period, a social planner chooses allocations of consumption and effective labor which depend on the agent's history, that is, on when an agent became disabled. A social planner chooses these allocations to maximize the lifetime expected utility of a representative agent. The constraint set of the problem is defined by feasibility and incentive-compatibility constraints.

3 Characterizing Pareto Optima

In this section we provide a theoretical characterization of the optimal allocation.

3.1 Full information benchmark

First, we start with a useful benchmark. If disability status is perfectly observable, then a social planner can achieve full insurance. For $\forall h^t$, $c^*(h^t) = \bar{c}$, i.e., consumption is constant over time, and consumption of able and disabled is equalized. The consumption-labor margin is also undistorted:

$$-v' \left(\frac{y_t^*(A)}{\theta_t} \right) \frac{1}{\theta_t} = u'(c_t^*(A))w.$$

3.2 Features of the optimal contract

In this section, we characterize the optimal solution when disability is unobservable. We call an allocation (c, y) to be *interior* if $y_t(A) > 0$ for all t . In the rest of the paper, we assume that the optimum is interior.

Proposition 1 *In the optimal allocation, the following properties are satisfied:*

1. *The feasibility constraint holds with equality;*
2. *Incentive constraints hold with equality in each period;*
3. *In each period, $c_t(A) > c_t(t)$.*

Proof. In the appendix. ■

Now we show that the consumption-labor margin is undistorted for able agents. This result is reminiscent of a result that in a static environment labor decisions of the highest skilled agent are undistorted (Mirrlees 1976).

Proposition 2 (Consumption-labor of able is not distorted) *Suppose (c^*, y^*) solves (P) . Then for each period t :*

$$-v' \left(\frac{y_t^*(A)}{\theta_t} \right) \frac{1}{\theta_t} = u'(c_t^*(A))w.$$

Proof. See the appendix. ■

The intertemporal margin, however, is distorted. An optimal solution has a wedge between a marginal rate of substitution and a marginal rate of transformation for able agents who face a non-zero probability of becoming disabled. We can now show a result that is similar to that of Golosov, Kocherlakota, and Tsyvinski (2003) to provide a characterization of the intertemporal wedge at the optimal solution.

Proposition 3 (Inverse Euler equation) *Suppose (c^*, y^*) solves (P) . Then for each period $t < T$*

$$\frac{1}{u'(c_t^*(A))} = \frac{1}{\beta(1+R)} \left[\frac{1}{u'(c_{t+1}^*(A))} \frac{\mu(h^{t+1}=A)}{\mu(h^t=A)} + \frac{1}{u'(c_{t+1}^*(t+1))} \frac{\mu(h^{t+1}=t+1)}{\mu(h^t=A)} \right].$$

Proof. In the appendix. ■

If future skills are not perfectly predictable given the current history of skills, then due to Jensen's inequality the above first-order conditions are not equivalent, and we can prove the following lemma.

Corollary 1 (Intertemporal decisions of able are distorted) *Suppose (c^*, y^*) solves P . Then if the probability of becoming disabled is interior $0 < \mu(h^{t+1}=t+1|h^t) < 1$,*

$$\begin{aligned} u'(c_t^*(A)) &< \beta(1+R)\{\mu(h^{t+1}=A|h^t=A)u'(c_{t+1}^*(A)) \\ &+ \mu(h^{t+1}=t+1|h^t=A)u'(c_{t+1}^*(t+1))\}. \end{aligned} \quad (3)$$

Proof. Apply Jensen's inequality to the condition in Proposition 2. ■

After an agent becomes disabled, all uncertainty is resolved, and there is no need to distort his intertemporal decision. Since we assumed that $\beta = \frac{1}{1+R}$, the consumption of disabled is constant.

Corollary 2 (Consumption of disabled is constant) *Suppose (c^*, y^*) solves (P) . Then*

$$c_i^*(t) = c_{t+i}^*(t), \text{ for } 1 \leq i \leq T - t.$$

Proof. In the appendix. ■

4 Implementation of the Optimum

In this section we propose a simple tax system that implements the optimal allocation and includes only taxes and transfers similar to those already in the U.S. tax code.

Since the only restrictions on the social planner's problem are incentive compatibility and feasibility, we implicitly allow a very large set of taxes. In fact, we allow any nonlinear taxes, including lump-sum taxes. Because of the generality of taxes, the social planner's allocation can be implemented in multiple ways, the most obvious of which is a direct mechanism. However, the direct mechanism and many other mechanisms are complex and may include taxes that have never been used in practice.

A theoretical challenge is to design a simple, realistic tax system that includes only elements that are already present in existing social programs. We first illustrate a difficulty in constructing simple tax systems with an example of a linear savings tax. This type of implementation is common in Ramsey literature of optimal taxation (see a review in Chari and Kehoe 1999). We show that such a tax does not implement the optimum, since it cannot prevent agents from overaccumulating assets and falsely claiming disability.

Despite these difficulties, we propose a strikingly simple tax/transfer system that implements the optimum: an asset-tested disability system. The proposed tax system consists of two important features. The first feature is that disability transfers have to depend on the length of pre-disability work history. The second feature is designed to control negative incentive effects of savings. Disability transfers should be asset-tested, that is, paid only to agents who have assets below a pre-specified minimum. Even though these features are not currently in SSDI, all of the proposed taxes and transfers are already present in the U.S. tax code and in other social programs. Some welfare programs, such as Supplemental Security Income and Temporary Assistance for Needy Families, are asset-tested.

This section is organized as follows: We start with formally defining a competitive equilibrium with taxes. We then present a simple two-period example that shows that a linear savings tax does

not implement the optimum. We proceed to design an asset-tested tax system that implements the optimum in a two-period example. Finally, we formally prove that an asset-tested disability system is the optimal allocation for a T -period economy.

4.1 Competitive equilibrium

We want to design a system of taxes and disability transfers so that competitive equilibrium allocations in such economy are constrained efficient. First, we formally define competitive equilibrium with taxes.

Definition 1 *Given a tax system $\{\tau_t(y(h^t), k(h^t))\}$, allocations of consumption, labor supply, and savings ($\{\tilde{c}(h^t)\}, \{\tilde{y}(h^t)\}, \{\tilde{k}(h^t)\}$) constitute a competitive equilibrium if they solve the following problem:*

$$\max_{(c,y) \geq 0, k} \sum_{t=1}^T \sum_{h^t} \beta^{t-1} \mu(h^t) \left\{ u(c(h^t)) + v\left(\frac{y(h^t)}{\theta^t}\right) \right\}$$

subject to

$$c(h^t) + k(h^t) \leq wy(h^t) + (1 + R)k(h^{t-1}) + \tau_t(y(h^t), k(h^t));$$

and feasibility is satisfied

$$\sum_{t=1}^{t=T} \sum_{h^t} \frac{1}{(1 + R)^{t-1}} \mu(h^t) c_t(h^t) \leq \sum_{t=1}^{t=T} \sum_{h^t} \frac{1}{(1 + R)^{t-1}} \mu(h^t) wy_t(h^t).$$

We will say that a tax system $\{\tau_t(y(h^t), k(h^t))\}$ implements the optimal allocation (c^*, y^*) if (c^*, y^*) is equal to competitive equilibrium allocations (\tilde{c}, \tilde{y}) defined above.

4.2 Simple example of implementation

In this section, we illustrate implementation of the solution to the social planner's problem with a simple example that will also clarify the role of asset testing. We consider a setup in which agents live for two periods and are able in the first period of their lives. When an agent is able, he has a skill $\theta = 1$. In the second period of his life, an agent will be able with probability μ_a and disabled with probability $(1 - \mu_a)$. Denote consumption of an able agent in period 1 by c_1 , of an able agent in period 2 by c_a , and of a disabled agent in period 2 by c_d . Denote allocations of effective labor of able agents in periods 1 and 2 by y_1 and y_a , respectively.

Then the optimal allocation $(c^*, y^*) = \{(c_1^*, c_a^*, c_d^*), (y_1^*, y_a^*)\}$ is the one that maximizes the expected lifetime utility of an agent subject to feasibility and incentive compatibility.

$$\max_{(c,y)} u(c_1) + v(y_1) + \beta\mu_a[u(c_a) + v(y_a)] + \beta(1 - \mu_a)u(c_d)$$

subject to

$$c_1 + \frac{\mu_a c_a}{1 + R} + \frac{(1 - \mu_a)c_a}{1 + R} \leq wy_1 + \frac{\mu_a wy_a}{1 + R}$$

$$u(c_a) + v(y_a) \geq u(c_d).$$

It is fairly straightforward to solve this social planner's problem. We take first-order conditions and rearrange them to obtain two features of the optimal allocation: intertemporal savings distortion and no distortion in labor decisions.

We can derive an analog of Proposition 3, the Inverse Euler equation (IEE):

$$\frac{1}{u'(c_1^*)} = \frac{\mu_a}{(1 + R)\beta u'(c_a^*)} + \frac{(1 - \mu_a)}{(1 + R)\beta u'(c_d^*)}. \quad (4)$$

The IEE implies that there is an intertemporal wedge, an analog to Corollary 1

$$u'(c_1^*) \leq (1 + R)\beta[\mu_a u'(c_a^*) + (1 - \mu_a)u'(c_d^*)]. \quad (5)$$

We can also derive an analog to Proposition 2 that labor decisions of able agents are not distorted at the margin:

$$wu'(c_1^*) = -v'(y_1^*)$$

$$wu'(c_a^*) = -v'(y_a^*).$$

Competitive equilibrium with uniform linear taxes is not Pareto optimal Implementation may seem to be simple in this example. One can conjecture (as in Diamond and Mirrlees 1978) that a linear savings tax that is equal to the intertemporal wedge in equation (5) and correctly chosen lump-sum taxes implement the optimal allocation. We will show that this conjecture is false, as there exists a profitable deviation for an agent.

Consider a tax system that consists of a savings tax τ , a tax T_a in period 1, and taxes T_a if an agent provides a positive amount of effective labor, and T_d if an agent does not work in period 2.

Formally, let

$$\tau_1(y_1) = T_a, \quad (6)$$

$$\tau_2(y_2, k_2) = \left\{ \begin{array}{l} -\tau Rk_2 + T_a, \text{ if } y_2 > 0 \\ -\tau Rk_2 + T_d, \text{ if } y_2 = 0 \end{array} \right\}. \quad (7)$$

We define taxes τ , T_a , T_d as follows. Savings tax τ is equal to the intertemporal wedge in the optimal solution defined by equation (3):

$$\tau = 1 - \left(\frac{u'(c_1^*)}{\beta[\mu_a u'(c_a^*) + (1 - \mu_a)u'(c_d^*)]} - 1 \right) / R. \quad (8)$$

We define T_a and T_d from the budget constraints of an agent who works in period 2 if he is able. We solve the system of linear equations to find taxes T_a , and T_d and a level of capital k_2^* such that the budget constraints are satisfied:

$$c_1^* + k_2^* = wy_1^* + T_a \quad (9)$$

$$c_a^* = (1 + R(1 - \tau))k_2^* + wy_a^* + T_a \quad (10)$$

$$c_d^* = (1 + R(1 - \tau))k_2^* + T_d. \quad (11)$$

We now show that this system of taxes does not implement the optimal allocation.

Proposition 4 (Competitive equilibrium with linear tax not efficient) *The tax system $(\tau_1(y_1), \tau_2(y_2, k_2), T_a, T_d)$ that is defined in (6)-(11) does not implement $(c_1^*, c_a^*, c_d^*, y_1^*, y_a^*)$.*

Proof. We will show that the agent who faces the tax system defined in (6)-(11) always obtains a higher utility if he chooses $y_2 = 0$ in the second period and receives disability benefits T_d in both states. This implies that (c^*, y^*) cannot be implemented.

The utility that the agent obtains from such a strategy is the solution to the following problem:

Problem 1:

$$\max_{(c, y, k)} u(c_1) + v(y_1) + \beta u(c_2)$$

subject to

$$c_1 + k_2 = wy_1 + T_a$$

$$c_2 = (1 + R(1 - \tau))k_2 + T_d.$$

This utility is higher than the utility of the disabled agent under optimal allocations (c_1^*, c_d^*, y_1^*) . To see this notice that (c_1^*, c_d^*, y_1^*) belongs to the budget constraint of the agent, so it is feasible. The solution to Problem 1, $(\tilde{c}_1, \tilde{c}_2, \tilde{y}_1)$, satisfies the following necessary condition

$$u'(\tilde{c}_1) = (1 + (1 - \tau)R)\beta u'(\tilde{c}_2). \quad (12)$$

(c_1^*, c_d^*) do not satisfy this condition because $c_a^* > c_d^*$ and

$$\begin{aligned} u'(c_1^*) &= (1 + (1 - \tau)R)\beta[\mu_a u'(c_a^*) + (1 - \mu_a)u'(c_d^*)] \\ &< (1 + (1 - \tau)R)\beta u'(c_d^*). \end{aligned}$$

Therefore (c_1^*, c_d^*, y_1^*) is not a solution to Problem 1. We showed that an agent can obtain a higher utility from the above strategy than the utility of the disabled. Since the incentive constraint binds, it is also higher than the utility from (c^*, y^*) because

$$\begin{aligned} u(\tilde{c}_1) + v(y_1^*) + \beta u(\tilde{c}_2) &> \\ u(c_1^*) + v(y_1^*) + \beta u(c_d^*) &= u(c_1^*) + v(y_1^*) + \beta[\mu_a(u(c_a^*) + v(y_2^*)) + (1 - \mu_a)u(c_d^*)]. \end{aligned}$$

This argument shows that the optimum (c^*, y^*) cannot be implemented. ■

Intuitively, a linear savings tax is not sufficient to implement the optimal allocation because it is designed to preclude single deviations. Given that an agent tells the truth, a linear savings tax guarantees that he chooses the correct amount of savings. Given that an agent chooses a correct amount of savings, an agent chooses to tell the truth. Above we have shown that a joint deviation may be profitable in which an agent decides to jointly lie and change the amount of savings.

We illustrate this intuition graphically. In Figure 1, we plot indifference curves for consumption in the first period c_1 and consumption in the second period c_2 . By incentive compatibility constraint, utility of claiming disability in the social planner's problem $u(c_1^*) + u(c_d^*)$ is equal to the utility of telling the truth. Point A represents this choice of (c_1^*, c_d^*) . Note that the slope of the indifference curve at this point is $-u'(c_1^*)/u'(c_d^*)$, and the tangency is represented by a dashed line. In a competitive equilibrium with a linear tax τ , an agent's budget line has a slope of $-(1 + R(1 - \tau))$

which is higher than $-u'(c_1^*)/u'(c_d^*)$ by the definition of the tax. The agent then will choose point B in which he consumes less in the first period (oversaves) and consumes more when disabled. Point B is on a higher indifference curve and gives a higher utility than $u(c_1^*) + u(c_d^*)$, hence, it gives higher utility than telling the truth.

4.2.1 Asset-tested disability system implements the optimal solution

We claim that a simple tax system exists, consisting of an asset-tested disability system that can implement the optimal allocation as a competitive equilibrium. The proof used in this section provides intuition for the proof of the general case in Section 4.3.

Let's first formally define an asset-tested disability system.

Definition 2 *An asset-tested disability tax system is a collection of taxes $\tau_1(y_1)$ and $\tau_2(y_2, k_2)$, and an asset limit k_2^* defined as follows*

$$\tau_1(y_1) = T_1,$$

$$\tau_2(y_2, k_2) = \left\{ \begin{array}{l} T_d, \text{ if } y_2 = 0 \text{ and } k_2 \leq k_2^* \\ T_a, \text{ otherwise.} \end{array} \right\}$$

Budget constraints in the competitive equilibrium problem then become:

$$c_1 + k_2 = wy_1 + T_1 \tag{13}$$

$$c_a = (1 + R)k_2 + \left\{ \begin{array}{l} T_d, \text{ if } y_2 = 0 \text{ and } k_2 \leq k_2^* \\ wy_2 + T_a, \text{ otherwise.} \end{array} \right\} \tag{14}$$

$$c_d = (1 + R)k_2 + T_d. \tag{15}$$

We then can construct taxes and savings (T_1, T_a, T_d, k_2^*) as a solution to the the system of linear equations:

$$c_1^* + k_2^* = wy_1^* + T_1 \tag{16}$$

$$c_a^* = (1 + R)k_2^* + wy_a^* + T_a \tag{17}$$

$$c_d^* = (1 + R)k_2^* + T_d. \quad (18)$$

Using the equations above, we can re-write the feasibility constraint as follows:

$$T_1 + \frac{\mu_a T_a + (1 - \mu_a) T_d}{1 + R} = 0. \quad (19)$$

We can now show that an asset-tested disability system implements the optimal allocation as a competitive equilibrium.⁴

Proposition 5 *Consider an optimal allocation (c^*, y^*) and any (T_1, T_a, T_d, k_2) that satisfy (16) – (18). Then (c^*, y^*) can be implemented in a competitive equilibrium by an asset-tested disability system that pays T_1 in the first period, T_d if an agent does not work ($y_2 = 0$) and has assets that are lower than the asset limit ($k \leq k_2$) in the second period, and T_a otherwise in the second period.*

Proof. We show that there exists no allocation $(\tilde{c}, \tilde{y}, \tilde{k})$ that satisfies budget constraint (13) – (15) and delivers higher utility to an agent than the utility from (c^*, y^*, k^*) .

It is easy to see that $T_a < T_d$, or, equivalently, $c_a^* - wy_a^* \leq c_d^*$. If this were not true then the planner would give an able agent the allocation for disabled $(c_d^*, 0)$ that gives the same ex ante utility by the incentive compatibility but utilizes less resources.

First, we show that there exists no allocation $(\tilde{c}, \tilde{y}, \tilde{k})$ that satisfies budget constraint (13) – (15) with $\tilde{k} > k_2^*$ and delivers higher utility to an agent than the utility from (c^*, y^*, k^*) . Under such allocation $(\tilde{c}, \tilde{y}, \tilde{k})$, an agent pays T_a in the second period when both able and disabled. Consider the maximum utility an agent can obtain in a competitive equilibrium if he receives T_a in the second period regardless of labor effort and call this allocation $(\hat{c}, \hat{y}, \hat{k})$. This allocation $(\hat{c}, \hat{y}, \hat{k})$ is trivially incentive compatible and requires strictly less resources than the planner's allocation from (19),

$$T_1 + \frac{T_a}{1 + R} \leq 0.$$

The planner could have chosen this allocation $(\hat{c}, \hat{y}, \hat{k})$ but did not do that, and, therefore, the utility of this allocation cannot be higher than the utility from the allocations (c^*, y^*, k^*) that solve the social planner's problem. The utility of the allocation $(\tilde{c}, \tilde{y}, \tilde{k})$ is weakly smaller than the utility

⁴Originally, our implementation had asset-tested disability and a savings tax. We are grateful to Ivan Werning for suggesting an elegant version of the proof that did not have a savings tax.

of the allocation $(\hat{c}, \hat{y}, \hat{k})$ and, therefore, smaller than the utility of (c^*, y^*, k^*) . An agent does not choose any allocations with $\tilde{k} > k_2^*$.

Alternatively, consider any allocation $(\tilde{c}, \tilde{y}, \tilde{k})$ that satisfies budget constraint (13) – (15) with $\tilde{k} \leq k_2^*$. We separately consider allocations with $y_2 > 0$ and with $y_2 = 0$ for an agent who is able in period 2:

(i) It is not optimal for an agent who works when able ($y_2 > 0$) to choose $k < k_2^*$.

Such an agent pays taxes T_a when able and T_d when disabled. Agent chooses allocations to maximize his utility subject to (13)–(15) and $k \leq k_2^*$. It is easy to see that for the utility maximizing allocation the consumption-labor margin is undistorted, and the savings constraint binds ($k = k_2^*$). Therefore, the agent chooses (c^*, y^*, k^*) .

(ii) It is not optimal for an agent who does not work when able ($y_2 = 0$) to choose $k < k_2^*$.

Such an agent pays T_d when both able and disabled. The utility maximizing allocation for an agent solves the problem

$$\max_{(c,y,k)} u(c_1) + v(y_1) + \beta u(c_2)$$

subject to

$$c_1 + k_2 = wy_1 + T_a$$

$$c_2 = (1 + R)k_2 + T_d$$

$$k \leq k_2^*.$$

Similarly to (i), $wu'(c_1) = -v'(y_1)$ and $k = k_2^*$. Therefore, the utility maximizing allocation is equal to that of the agent who deviates and falsely claims to be disabled in planner's problem. By incentive compatibility, utility from such allocation does not exceed the utility from (c^*, y^*, k^*) . Therefore, (c^*, y^*, k^*) are the equilibrium allocations. ■

In Figure 2, we illustrate the intuition behind asset testing. If an agent is allowed to consume freely on the budget line with the slope $-(1 + R)$, he could deviate to a higher indifference curve by reasoning identical to the description of Figure 1. Asset testing introduces a non-convexity in the budget set, and for $c_1 > c_1^*$ ($k > k^*$) shifts the budget line down by the amount $T_a - T_d$. If an agent has assets higher than k^* , he receives transfer T_a instead of T_d . The budget line is now represented by a solid line. One can see that an agent chooses to consume (c_1^*, c_d^*) . Our choice of transfer T_a that a disabled agent receives if he has assets higher than the maximum was motivated

by convenience, because it means that an agent with high assets will be treated exactly as an able agent. However, a transfer higher than T_a for those with higher assets can also implement the optimum, any tax such that the budget line with the slope $-(1 + R)$ is below the tangency to the indifference curve and to the left of c_1^* is also sufficient to implement the optimum.

4.3 Implementation: general case

We proceed to formally prove how to construct a tax system that implements the optimal allocation for a T period case. The tax system we consider is an asset-tested disability insurance system. All transfers and taxes depend on the number of years spent working before claiming disability. Transfers also depend on the amount of assets that an agent has. In particular, an agent receives a disability transfer only if he does not work and his assets are less than some pre-specified limit. If his assets are higher than that limit, he has to pay the taxes of an able agent.

Let's formally define what comprises an asset-tested disability insurance program.

Definition 3 *An asset-tested disability tax system is a collection of taxes $\tau_t(y^t, k^t)$ and asset limits \bar{k}_t where:*

- $\tau_t(y^t, k^t) = T_d(i) - wy_t$ if $y_j = 0$ and $k_j \leq \bar{k}_j$ for some $j \leq t$, where $i \leq j$ is the number of years before j when $y > 0$.
- $\tau_t(y^t, k^t) = T_a$, otherwise.

The agent with assets below the asset limit who chooses not to work in any period j never returns to work again and gets taxes $T_d(i)$ in each period for the rest of his life. We say that such an agent *claims disability* in period j . The size of taxes depends on the work history of the agent. If the agent worked in all years he was productive, his work history is equal to his age.

We will refer to this asset tested system defined above as $(T_a, T_d(t), \bar{k}_t)$.

The theorem that follows states the main theoretical result of the paper. We prove that we can construct an asset-tested disability insurance program such that the optimal allocation is implemented as a competitive equilibrium.

Theorem 1 *For any optimal allocation (c^*, y^*) that solves the social planner's problem, there exists such an asset-tested disability insurance program $(T_a, T_d(t), \bar{k}_t)$ that the optimal allocation is a solution to the competitive equilibrium problem.*

Proof. See the appendix. ■

The proof of Theorem 1 includes the same steps as in the example that we discussed earlier.

5 Quantitative Results

The section is organized as follows. We first describe a stylized model of the current U.S. social insurance system that we use to choose the parameters of the model. We proceed to characterize a solution to the social planner’s problem. We then evaluate the welfare gains of switching to an optimal social insurance system.

5.1 Parameterization

In this subsection we use a stylized current social security system as a benchmark to choose the parameters of the model. Since disability insurance is an integral part of the social insurance system, we model the current social insurance system as a joint disability and retirement system. We do not make distinction in this model between claiming disability and retirement. If an agent decides not to work, he receives a social security transfer. An agent can stop working either because he is truly disabled or because social insurance transfers create a disincentive to work. An agent can save at a rate R . When an agent stops working, he receives a disability transfer T_d independent of the age.

We calibrate the probability of becoming disabled using the data from McNeil (1997), published by the U.S. Census Bureau. The data report the number of self-reported disabled people by age groups⁵. We use the data to calculate a conditional probability of becoming disabled and extrapolate the data to one-year intervals by fitting an exponential function (see Figure 3).

Table 1 reports shares of disabled people in our model by various age groups. We assume that 4 percent of the population is disabled at age 25, before entering the labor force. We compare the numbers we calculated to those reported in the Survey of Income and Program Participation (SIPP) and the Current Population Survey (CPS). The SIPP estimates the number of people with severe disabilities, and the CPS reports the number of people with work disabilities. The CPS does not have information about work disabilities of people who are over age 65.

⁵These numbers are self-reported and hence need not reflect the true disability of these individuals. However, a number of papers showed that self-reported health characteristics are a very good approximation of true health. For one of the recent papers, see Benitez-Silva et al. (2000).

Table 1					
Share of disabled population					
Age groups	25-34	35-44	45-54	55-64	65-74
Model	5.2%	8.33%	13.97%	24.54%	43.19%
CPS (Stoddard et al. (1998))	5.5%	9.1%	13.2%	23.1%	n.a.
SIPP (McNeil (1997))	8.1%		13.9%	24.2%	30.7%

We choose the parameters of the model such that a solution to the stylized version of the social insurance system matches selected empirical observations. The utility function is chosen to be $\ln(c) + a\ln(1 - l)$, where a is relative disutility of labor. We assume that a period in the economy is one year and that each agent begins life at age 25 and lives to the age of 75 years⁶.

We choose the value of disutility of work to be equal to $a = 1.5$ implying that average working hours constitute 30 percent of the time endowment. This number is consistent with the microeconomic evidence that households allocate about one-third of their discretionary time to market activities (Ghez and Becker 1975). We choose the interest rate in the economy to match an annual interest rate net of depreciation R to be equal to 4.3 percent. Then we choose the discount factor $\beta = 0.958$ as a solution to $\frac{1}{\beta} = 1 + R$. The aggregate production function is a standard Cobb-Douglas production function with constant returns to scale $F(K, Y) = K^\alpha Y^{1-\alpha}$. In the production function, we set the capital share equal to $\alpha = 0.33$. We use the calibrated interest rate and the capital share to calculate the wage rate $w = 1.2243$.

We use a lifetime skill profile that we obtained by fitting a quadratic function in the data from Rios-Rull (1996). Agents achieve the highest skill level at around age 50. At age 50 they are 45 percent more productive than at age 25. After age 50, skills decline and reach the minimum at age 75, with skills being roughly equal to the skill level at age 25 (Figure 4).

5.2 Optimal system

In this subsection we numerically characterize an optimal disability insurance system. A social planner seeks to provide insurance to disabled agents, while ensuring that able agents work and do

⁶We also solved a version of the model that does not include the retirement period from age 65 to 75 and obtained similar results. For more detailed results on computations see the working paper version (Goloso and Tsyvinski 2003).

not claim disability. The incentive constraint in each period can be written in the following form:

$$\begin{aligned} u(c_t(A)) - v(y_t(A)/\theta_t) + \mu(h^{t+1} = A|h^t = A)V^a(t+1) \\ + \mu(h^{t+1} = t+1|h^t = A)V^d(t+1) &= V^d(t). \end{aligned}$$

The incentive constraint can be satisfied either if the planner gives a relatively higher consumption to the able agent $c_t(A)$ then to the disabled $c_t(t)$ or by promising a relatively higher value $\mu(h^{t+1} = A|h^t = A)V^a(t+1) + \mu(h^{t+1} = t+1|h^t = A)V^d(t+1)$ to the able agent in future. Similar to the standard results in the dynamic contracting literature, it is optimal for the planner to use both of these instruments. Thus the social planner faces two types of trade-offs. The first trade-off is between intratemporal optimal insurance (making consumption of able and disabled agents constant) and the incentive compatibility constraint (allocating more consumption to able agents to induce them to work). The second trade-off is between intertemporal optimal insurance (making consumption constant across time) and the incentive compatibility constraint (allocating more consumption to able agents who have a longer work history). If disability were perfectly observable, a planner would be able to achieve both intertemporal and intratemporal smoothing. In that case consumption of able and disabled agents would be equal and constant over time. However, we will see that the optimal allocation differs significantly from the solution without informational friction.

We start the analysis of the optimal system by calculating a solution to the social planner's problem for the parametrized economy described in the previous section.

5.2.1 Consumption profiles

We report consumption profiles in Figure 5. The solid line represents the consumption of agents who were able all their lives. This consumption is monotonically increasing with the duration of agent's work history. The social planner partially rewards the agent for working in period t by allocating him a higher value $V^a(t+1)$, which leads to higher consumption at future dates. The consumption of able 65 year olds are about 15% higher than the consumption of able 25 year old agents.

The dashed line represents consumption that an agent receives if he becomes disabled. After becoming disabled agent's consumption remains constant for the rest of his life. Two sample horizontal lines represent consumption of agents who became disabled at age 40 and age 60. There is a significant fall in consumption after an agent becomes disabled. This drop is necessary to

ensure that able agents do not deviate and claim disability. Consumption of a 25-year-old agent falls by about 50 percent from 0.6 to 0.3.

There are two effects that determine consumption of the agent who becomes disabled. First, efficiency requires that more skilled agents work more, and, therefore, the consumption drop should be larger for such agents to provide incentives. We can see that agents who become disabled at ages 26 to 40 receive lower consumption than those who became disabled at the age of 25. That is the period when agent's productivity raises.

The second effect comes from the intertemporal provision of the incentives. The planner rewards an agent for working by increasing the value of $V^d(t)$. This effect calls for higher consumption for the agents who become disabled later in life. It dominates the first effect around the time when an agent reaches the age 33. The second effect increases with the work history, so that the consumption of disabled, the dashed line, raises steeper than the solid line for able agents later in the lifecycle.

In the working paper version of the paper we computed the optimal system for the case when the skills of the able are the same for all ages. In that case there is only the second effect, and the consumption of able and disabled monotonically increases with the length of work history since there is no reason to require middle-aged agents to work more.

5.2.2 Labor

In the calibrated economy it is optimal for all able agents to work. Optimal allocations are influenced by two effects in a way similar to the consumption profiles. On the one hand, it is optimal to require more productive agents to work more, and labor supply inherits the hump-shaped form of the skills profile. Agents who are 45 to 55 year old and have the highest skills spend about 65-70 percent of their time working. Younger and older people are not as productive and work less.

On the other hand, intertemporal provision of incentive calls for an increase in future $V^a(t)$, which can be partially achieved by reducing amount of labor. For example, 30 and 70 year old agents have roughly the same skills, however, 30 year old agents work twice as much. In the working version of the paper we numerically show that when skill profile is constant across time, there is only this second effect, so that labor supply monotonically decreases with work history.

The provision of intertemporal incentives leads to endogenous retirement. Older agents tend to work significantly less than young agents with similar skills. Agents who are 75 year old and still productive spend less than 10% of their time working.

5.2.3 Intertemporal distortion

An important feature of the model is the intertemporal distortion. We define distortion $\tau_t(A)$ as

$$\tau_t(A) = 1 - \left(\frac{\mu(h_t = A)u'(c_t^*(A))}{\beta[\mu(h_{t+1} = A)u'(c_{t+1}^*(A)) + \mu(h_{t+1} = t + 1)u'(c_{t+1}^*(t + 1))]} - 1 \right) / R.$$

The intertemporal distortion depends on three factors: probability of becoming disabled, skill profile, and the length of work history. The probability of becoming disabled increases for older agents, thus making their future consumption more unpredictable, which increases the distortion. For higher skilled agents the incentive problems are more severe, and they face a higher intertemporal wedge. There two factors explain the steep rise in τ until age 55. The third factor, the length of work history, decreases the wedge. Agents with longer work history provide less labor and have smaller variance of consumption. Incentive problems are not as severe for them and they face a smaller intertemporal distortion.

The interaction of these three effects is shown on Figure 7. The wedge grows from 3 percent at age 24 to 10.5 percent at age 56, and decreases to 2 percent at age 74. The intertemporal distortion is quantitatively significant, especially given that the probability of becoming disabled is almost an order of magnitude smaller than the size of the distortion.

5.3 Tax system implementing the optimum

We can use the optimal solution to compute the tax system implementing the optimal allocation. From the proof of Theorem 1 we know how to construct the asset limits and transfers. We can calculate and plot asset limits and transfers to disabled in Figures 8 and 9. In Figure 9, a solid line represents optimal disability transfers that an agent receives for the rest of his life. For example, an agent who stops working at age 45 will receive approximately 0.35 units of consumption for the rest of his life. Asset limits are eventually increasing because agents become wealthier as they accumulate more capital. That is also the reason that disability transfers are decreasing as agents receive a larger proportion of their income from savings. One interpretation of this system is that individuals who became disabled early in life receive large transfers, while those who become disabled later are supposed to supplement their lower disability transfers with savings accumulated while able.

5.4 Welfare benefits of asset testing

In this subsection we compare the welfare of the best program without asset testing with that of the optimal insurance system. Theorem 1 and the discussion in Section 4.3 showed that asset-tested disability transfers allow a government to implement the optimum allocations. In the U.S., only a small part of disability insurance system has elements of asset testing. The majority of the workforce is covered by the Social Security Disability Insurance (SSDI) system in which benefits are not tied to individual asset holdings.

First, we numerically characterize the optimal disability insurance system without asset testing. Then we evaluate how much welfare can be increased if the government switches to asset-tested disability transfers.

We consider a system in which the planner does not have access to asset-tested disability transfers and cannot impose any intertemporal savings distortion. The optimal disability system without asset tests is a solution to the social planner's problem with hidden savings, an example of which is Diamond and Mirrlees (1995). The model with hidden savings is also similar to that of Werning (2001), and Abraham and Pavoni (2003). However, the dynamic first-order approach pioneered by Werning and Abraham and Pavoni of imposing the Euler equation on the social planner's problem is invalid in our setup⁷. An agent presented with the solution to such a problem has a profitable joint deviation of claiming disability and hidden savings. The proof of invalidity of the first-order approach for the model of hidden savings for our setup is similar to the proof that a linear tax does not implement the optimum described earlier.

Our computational method for the model of hidden savings is similar to that in Golosov and Tsyvinski (2004). For each lifetime allocation of consumption and labor that a planner offers to an agent, we compute T optimal joint deviations in which an agent claims disability and chooses the optimal level of hidden savings, and an additional single deviation in which an agent tells the truth but chooses a different level of savings than that prescribed by the planner. This method is relatively simple and in our model allows a globally optimal solution to the social planner's problem with hidden savings.

In Figure 10, we compare the solution to the problem with asset testing (thick lines) to that of the problem without asset testing (thin lines). In the latter case, the planner does not have an ability to distort an intertemporal margin, and, therefore, there is less insurance provided to

⁷See also Kocherlakota (2004) for a discussion of the first order approach.

an agent. The consumption profiles of a young able agent for the case of no asset testing starts below consumption of an able young for the case of asset testing and increases more rapidly. This rapid increase is the first source of the welfare loss compared to the optimal system as agents prefer smooth consumption profiles. The decline in consumption for agents who are 30-40 years old is large for the case of no asset testing. This decline is needed to ensure that an agent does not claim disability before becoming most productive at age 45-50. In the absence of intertemporal distortion (no asset testing), the planner has to penalize agents who declare disability early by lower consumption than when asset testing is available, leading to the second source of welfare losses. Another source of welfare loss in the case of no asset-test is that the consumption profile of disabled is less smooth than when asset-test is available. In particular, consumption allocations to those who become disabled rise steeply after age 35.

In Figure 11, we plot labor allocations with and without asset testing (solid and dashed lines, respectively). The labor profiles for both cases are virtually identical until age 40. After age 40, the absence of asset test implies that it is more difficult to provide incentives to work, and there is a smaller amount of labor provided. This leads to another source of welfare loss: a lower amount of labor implies a smaller amount of resources produced in the economy and a smaller amount of resources available for redistribution to disabled.

In Figure 9, we plotted with a dashed line optimal disability transfers for the model where asset testing is not allowed. It is evident from this figure that asset testing allows a significant increase in the level of disability transfers.

We use the following method to measure welfare gains from switching to an optimal system. We find that proportional increase in consumption by 0.65% for each history under the optimal system without asset testing produces the same lifetime utility as the lifetime utility in an optimal system with asset testing. This number indicates a significant gain to asset testing as the amount of consumption of disabled (the size of the program) in the model is equal to 4.7 percent of the total consumption in the optimal solution. The size of the program of 4.7 percent of total consumption is consistent with the current spending on Social Security and healthcare of elderly and disabled in the U.S.⁸

⁸We have performed extensive robustness checks of the results that are available upon request. Among the important tests we considered an economy where agents live for 65 years and did not find a significant change in the results.

6 Final Remarks

The paper achieves three goals. First, we provide an implementation of a dynamic optimal social insurance problem with a simple set of tax instruments in the presence of private savings. Despite the complexity of such dynamic implementation, we have found a simple and realistic transfer system - an asset-tested disability system. Second, the paper provides an incentive theoretic reasoning for the existence and optimality of asset-tested programs which are prevalent in practice of social insurance. Our results also extend previous work on means-tested programs to dynamic settings. Finally, we found that asset testing is quantitatively superior to the programs where asset testing is not allowed.

The methods developed in this paper may have wider applicability to the design of other social welfare programs such as unemployment insurance. Another application is design of the retirement portion of the social security. One view of the Social Security program (Diamond and Mirrlees 1978, 1986 and Mulligan and Sala-i-Martin 1999) is that it provides insurance against the inability to work in later periods of life. One problem with the current Social Security program is that it may encourage retirement earlier than socially optimal. Theoretical results on asset testing directly translate to such a model of social insurance. In the numerical section of the paper, we explored some implications for welfare of incorporating asset testing.

As the main goal of the paper was to explore a parsimonious dynamic model to highlight the impact of savings on implementation of the social program, there are several dimensions in which it can be extended.

It would be interesting to analyze a setup with several levels of unobservable skills, which will allow more precise choosing of the parameters of the model. We conjecture that general methodology of implementation derived here can be translated to this setting, with asset test taking a form of multiple asset tests similar to those used in determination for college financial aid. Another, yet more difficult extension is considering some form of ex-ante heterogeneity such as in Whinston (1983). This extension would be very challenging as the solutions are very complicated even in the static setting.

Second, an interesting extension would be to study an economy in which disability is not permanent. In that case, optimal disability benefits will also have to encourage individuals who recover from disability to leave disability rolls. This is an interesting extension given current interest in the proposed policy initiatives such as “Ticket to work” (See SSA 1998). It would

be also interesting to compare the cost and benefits of the asset testing in a lifecycle model of Hubbard, Skinner, and Zeldes (1995) who found that asset testing is an explanation of low levels of savings. Our model does not have any other motives for savings than lifecycle and incentive reasons. Incorporating other motives for savings will allow for a more realistic evaluation of costs and benefits of asset testing.

We also treated government as the only provider of disability insurance without considering insurance that is provided by private markets. In our theoretical work (Golosov and Tsyvinski 2004), we provided a model that explored endogenous private markets and the role of government taxation in such settings.

While the described extensions are interesting, the magnitude of the welfare gains from switching to the optimal system gives us confidence that the forces we have captured in this paper are significant from both the theoretical and policy-making perspectives.

References

- [1] Abraham, A. and Pavoni, N., 2003, Efficient Allocations with Moral Hazard and Hidden Borrowing and Lending, Working Paper, University College London.
- [2] Albanesi, S. and Sleet, C., 2004, Dynamic Optimal Taxation with Private Information, Discussion Paper 140, Federal Reserve Bank of Minneapolis.
- [3] Autor, D. H. and Duggan, M. G., 2001, The Rise in Disability Recipiency and the Decline in Unemployment, Working Paper No. 8336, National Bureau of Economic Research (forthcoming in *Quarterly Journal of Economics*).
- [4] Benitez-Silva, H., Buchinsky, and Rust, J., 2004, How Large are the Classification Errors in the Social Security Disability Award Process?, Working Paper No. 10219, National Bureau of Economic Research.
- [5] Benitez-Silva, H., Buchinsky, M., Chan, H. M., Cheidvasser, S., and Rust, J., 2000, How Large Is the Bias in Self-Reported Disability? Working Paper No. 7526, National Bureau of Economic Research.
- [6] Besley, T. and Coate, S., 1995, The Design of Income Maintenance Programmes, *Review of Economic Studies*, 62 (2), 187–221.

- [7] Bound, J. and Burkhauser, R., 1999, Economic Analysis of Transfer Programs Targeted on People with Disabilities, *Handbook of Labor Economics*, Vol. 3C, pp. 3417–3528, Amsterdam; New York and Oxford: Elsevier Science, North-Holland.
- [8] Chari, V. and Kehoe, P., 1999, Optimal Fiscal and Monetary Policy, in Taylor, J. and Woodford, M., eds., *Handbook of Macroeconomics*, Vol. 1C, pp. 1671-1745, Amsterdam: North Holland.
- [9] Diamond, P. A., and Mirrlees, J. A., 1978, A Model of Social Insurance with Variable Retirement, *Journal of Public Economics* 10 (3), 295–336.
- [10] Diamond, P. A., and Mirrlees, J. A., 1986, Payroll-Tax Financed Social Insurance with Variable Retirement, *Scandinavian Journal of Economics* 88 (1), 25–50.
- [11] Diamond, P. and Mirrlees, J., 1995, Social Insurance with Variable Retirement and Private Saving. Mimeo. MIT
- [12] Feldstein, M., 1987, Should Social Security be Means Tested?, *Journal of Political Economy*, 95(3), 468-484.
- [13] Ghez, G. R. and Becker, G. S., 1975, *The Allocation of Time and Goods over the Life Cycle*, New York: Columbia University Press.
- [14] Golosov, M. and Tsyvinski, A., 2003, Designing Optimal Disability Insurance, Working Paper No. 628, Federal Reserve Bank of Minneapolis.
- [15] Golosov, M. and Tsyvinski, A., 2004, Optimal Taxation with Endogenous Insurance Markets, mimeo.
- [16] Golosov, M., Kocherlakota, N., and Tsyvinski, A., 2003, Optimal Indirect and Capital Taxation, *Review of Economic Studies*, 70 (3), 569–587.
- [17] Hansen, G. D. and Imrohoroglu, A., 1992, The Role of Unemployment Insurance in an Economy with Liquidity Constraints and Moral Hazard, *Journal of Political Economy* 100 (1), 118–142.
- [18] Hopenhayn, H. A. and Nicolini, J. P., 1997, Optimal Unemployment Insurance, *Journal of Political Economy* 105 (2), 412-438.

- [19] Hubbard, R. G., Skinner, J., and Zeldes, S., 1995, Precautionary Saving and Social Insurance, *Journal of Political Economy* 103 (2), 360–399.
- [20] Kocherlakota, N. R., 2003, Zero Expected Wealth Taxes: A Mirrlees Approach to Dynamic Optimal Taxation, Stanford University, mimeo.
- [21] Kocherlakota, N.R., 2004, Figuring Out the Impact of Hidden Savings on Optimal Unemployment Insurance, *Review of Economic Dynamics*, 7, 541-554.
- [22] McNeil, J., 1997, Americans with Disabilities: 1994–95. Current Population Reports, Household Economic Studies P70-61, U.S. Department of Commerce, Bureau of the Census.
- [23] Mirrlees, J. A., 1976, Optimal Tax Theory: A Synthesis, *Journal of Public Economics*, 6 (4), 327–358.
- [24] Mulligan, C. B. and Sala-i-Martin, X., 1999, Social Security in Theory and Practice (II): Efficiency Theories, Narrative Theories, and Implications for Reform, Working Paper No. 7119, National Bureau of Economic Research.
- [25] Rios-Rull, J.-V., 1996, Life-Cycle Economies and Aggregate Fluctuations, *Review of Economic Studies* 63 (3), 465–489.
- [26] Saez, E., 2002, Optimal Income Transfer Programs: Intensive Versus Extensive Labor Supply Responses, *Quarterly Journal of Economics*, 117(3), 1039–1073.
- [27] Stoddard, S., Jans, L., Ripple, J., and Kraus, L., 1998, *Chartbook on Work and Disability in the United States. An InfoUse Report*. Washington D.C.: U.S. National Institute on Disability and Rehabilitation Research.
- [28] U.S. Department of Health and Human Services, 1988, *Social Security Handbook*, Tenth Edition.
- [29] U.S. Social Security Administration (SSA), 2000, *Social Security Bulletin: Annual Statistical Supplement*.
- [30] U.S. Social Security Administration (SSA), 1998, 2001, *Social Security Advisory Board Reports*.
- [31] Wang, C. and Williamson, S. D., 1996, Unemployment Insurance with Moral Hazard in a Dynamic Economy, *Carnegie-Rochester Conference Series on Public Policy* 44, 1–41.

- [32] Werning, I., 2001, Optimal Unemployment Insurance with Hidden Savings, University of Chicago, mimeo.
- [33] Whinston, M. D., 1983, Moral Hazard, Adverse selection, and the Optimal Provision of Social Insurance, *Journal of Public Economics*, 22 (1), 49–71.

7 Appendix

7.1 Characterizing Pareto Optimum

Here we formally prove Propositions 1, 2, 3, and Corollary 2.

Proposition 1:

Part 1

It is straightforward to show that if a feasibility constraint does not hold with equality for some allocation, it is possible to find another feasible and incentive compatible allocation that gives a higher utility.

Part 2

Suppose an IC does not bind in some period t . Perturb the allocation by decreasing $c_t(A)$ by small $\varepsilon > 0$ and increasing $c_t(t)$ by $\varepsilon\mu(h^t = A)/\mu(h^t = t)$. The perturbed allocation satisfies feasibility. It is also incentive compatible for ε sufficiently small. To see this, note that the IC in period t is relaxed. >From concavity, $V^a(t-1)$ increases while the value of $V^d(t-1)$ is unchanged. By backward induction we can show that all the incentive constraints in period $1, \dots, t$ do not bind.

Part 3

Consider the FOCs for $c_t(A)$ and $c_t(t)$ where λ and η_t are Lagrangian multipliers on the feasibility and time t period incentive constraints:

$$\begin{aligned} u'(c_t(A))[\mu(h^t = A)\beta^t + \eta_t + \eta_{t-1}\mu(h^t = A|h^{t-1} = A)\beta + \dots + \eta_1\mu(h^t = A|h^1 = A)\beta^t] \\ = \lambda \frac{\mu(h^t = A)}{(1+R)^{t-1}}, \end{aligned}$$

$$\begin{aligned} u'(c_t(t))[\mu(h^t = t)\beta^t - \eta_t + \eta_{t-1}\mu(h^t = t|h^{t-1} = A)\beta + \dots + \eta_1\mu(h^t = t|h^1 = A)\beta^t] \\ = \lambda \frac{\mu(h^t = t)}{(1+R)^{t-1}}. \end{aligned}$$

Dividing these equations by $\mu(h^t = A)$ and $\mu(h^t = t)$ and subtracting one from the other we obtain

$$\begin{aligned} & (u'(c_t(t)) - u'(c_t(A))) \left[\beta^t + \frac{\eta_{t-1}\beta}{\mu(h^{t-1} = A)} + \dots + \frac{\eta_1\beta^t}{\mu(h^1 = A)} \right] \\ = & \eta_t \left(\frac{u'(c_t(A))}{\mu(h^t = A)} + \frac{u'(c_t(t))}{\mu(h^t = t)} \right). \end{aligned}$$

The right hand side of this equation is positive as well as the expression in the square brackets. Therefore $u'(c_t(t)) - u'(c_t(A)) > 0$ which implies that $c_t(A) > c_t(t)$.

Propositions 2, 3, and Corollary 2:

Simplify incentive compatibility constraints. There are only T incentive constraints. We rewrite them by substituting constraint at age T , into constraint at age $T - 1$, and so on to period 1:

$$u(c_t(A)) + v\left(\frac{y_t(A)}{\theta_t}\right) + \sum_{i=1}^{T-t} \beta^i u(c_{t+i}(t)) = \sum_{i=0}^{T-t} \beta^i u(c_{t+i}(t))$$

for all $\theta_t > 0$. Taking the first order conditions and simplifying yields Propositions 2 and 3. Corollary 2 follows from the assumption that $\beta = \frac{1}{1+R}$.

7.2 Proof of Theorem 1.

We construct an asset-tested disability system $(T_a, T_d(t), \bar{k}_t)$ that implements (c^*, y^*) as a competitive equilibrium. Transfers T_a are obtained from the following equation:

$$\sum_{t=1}^{t=T} \frac{c_t^*(A)}{(1+R)^{t-1}} = \sum_{t=1}^{t=T} \frac{wy_t^*(A)}{(1+R)^{t-1}} + T_a \sum_{t=1}^{t=T} \frac{1}{(1+R)^{t-1}}.$$

Taxes $T_d(j)$ are determined as follows:

$$\sum_{t=1}^{t=j-1} \frac{c_t^*(A)}{(1+R)^{t-1}} + \sum_{t=j}^{t=T} \frac{c_t^*(j)}{(1+R)^{t-1}} = \sum_{t=1}^{t=j-1} \frac{wy_t^*(A) + T_a}{(1+R)^{t-1}} + T_d(j) \sum_{t=j}^{t=T} \frac{1}{(1+R)^{t-1}}.$$

Finally, the asset limits \bar{k} are defined recursively

$$c_t^*(A) + \bar{k}_{t+1} = wy_t^*(A) + (1+R)\bar{k}_t + T_a$$

with $\bar{k}_1 = 0$.

Using the equations above, we can re-write the feasibility constraint as follows:

$$\sum_{t=1}^{t=T} \left(\left(\sum_{i=1}^{i=t} \frac{\mu(h^i = i)T_d(i)}{(1+R)^{t-1}} \right) + \frac{\mu(h^t = A)T_a}{(1+R)^{t-1}} \right) \leq 0. \quad (20)$$

Now we show that this asset tested system implements (c^*, y^*) as a competitive equilibrium. First, we prove that the expected present value of transfers for the able agent is lower than the present value of disability transfers.

Lemma 1 *If (c^*, y^*) is optimal, then $\{T_a, T_d(t)\}$ satisfy for all $t = 1 \dots T$*

$$\begin{aligned} & T_a + \sum_{i=t+1}^{i=T} \left[\mu(h^i = i | h^t = A) \left(\sum_{s=i}^{s=T} \frac{T_d(i)}{(1+R)^{s-t}} \right) + \frac{\mu(h^i = A | h^t = A) T_a}{(1+R)^{i-t}} \right] \\ & \leq \sum_{i=0}^{i=T-t} \frac{T_d(t)}{(1+R)^i}. \end{aligned} \quad (21)$$

Proof. Equation (21) is equivalent to

$$\begin{aligned} & c_t^*(A) - w y_t^*(A) + \sum_{i=t+1}^{i=T} \sum_{h^i > \{h^t = A\}} \frac{\mu(h^i | h^t = A) [c_i^*(h^i) - w y_i^*(h^i)]}{(1+R)^{i-t}} \\ & \leq \sum_{i=0}^{i=T-t} \frac{c_{t+i}^*(t)}{(1+R)^i}. \end{aligned} \quad (22)$$

Suppose equation (22) does not hold for some t . Then the social planner could give the allocation of the disabled ($\{c_{t+i}^*(t)\}_{i=0}^{i=T-t}, 0$) to the agents who are still able in period t . Since the period t incentive constrain holds with equality from Proposition 1, the utility of the agent does not change. The new allocation is still incentive compatible, but the feasibility constraint is relaxed. The social planner can further improve upon such an allocation, therefore, (c^*, y^*) cannot be an optimum. ■

Let's derive a useful corollary to Lemma 1.

Corollary 3 $T_a \leq T_d(t)$ for any t .

Proof. Lemma 1 implies that $T_a \leq T_d(T)$. Consider period $T - 1$. From Lemma 1:

$$\begin{aligned} & T_a + \frac{1}{1+R} (\mu(h^T = A | h^{T-1} = A) T_a + \mu(h^T = T | h^{T-1} = A) T_d(T)) \\ & \leq T_d(T-1) \left(1 + \frac{1}{1+R} \right). \end{aligned}$$

Since $T_a \leq T_d(T)$ the above equation implies

$$T_a \leq T_d(T-1).$$

By backward induction can show that for all t

$$T_a \leq T_d(t).$$

■

Now we are ready to prove *Theorem 1*.

Proof. Consider the asset tested system constructed as described above. Since the agent maximizes continuous utility function over a compact budget set⁹, there exists a maximum. Pick any allocation (\tilde{c}, \tilde{y}) together with his saving decisions (\tilde{k}) that maximizes agent's utility. We will show that the utility from such allocations cannot be higher than the utility from (c^*, y^*) .

Step 1: There exists a utility maximizing allocation $(\tilde{c}, \tilde{y}, \tilde{k})$ such that an agent never claims disability if he is able.

Suppose that an agent is strictly better off by claiming disability if he is able in some period j . The agent can claim disability in period j only if his assets in that period are $\tilde{k}_j \leq \bar{k}_j$. Suppose $\tilde{k}_j = \bar{k}_j$. By construction the maximum utility the agent can obtain if his assets are \bar{k}_j and his taxes are $T_d(j) - wy_t$ for all the subsequent periods is $u(c_j^*(j)) + \dots + \beta^{T-j}u(c_T^*(j))$, which is the utility that the Social Planner allocates to the agent who becomes disabled in period j . But the agent with assets \bar{k}_j in period j can choose the future path $(\{c^*\}_{t=j}^T, \{y^*\}_{t=j}^T)$ as it is in his budget constraint. By the incentive compatibility of the optimal allocations, this future path gives weakly higher utility than claiming disability in period j .

Alternatively, suppose $\tilde{k}_j < \bar{k}_j$. Agent's utility maximization implies that $\tilde{c}_j(j) < c_j^*(j)$. The allocation is utility maximizing in this case if

$$u'(\tilde{c}_{j-1}(A)) = \beta(1 + R)u'(\tilde{c}_j(j)).$$

If this Euler equation did not hold, an agent could transfer a small amount of resources ε intertemporally. Such a transfer still allows him to claim disability in period j , and gives strictly higher utility. Since $\tilde{c}_j(j) < c_j^*(j)$ this implies together with Corollary 1 that $u'(\tilde{c}_{j-1}(A)) > u'(c_{j-1}^*(A))$. The agent's budget constraint $\tilde{c}_{j-1}(A) + \tilde{k}_j = w\tilde{y}_{j-1}(A) + (1 + R)\tilde{k}_{j-1}$ and intratemporal optimality condition

$$-v' \left(\frac{\tilde{y}_{j-1}(A)}{\theta_{j-1}} \right) \frac{1}{\theta_{j-1}} = u'(\tilde{c}_{j-1}(A)) w,$$

imply that $\tilde{k}_{j-1} < \bar{k}_{j-1}$. We can continue backward to show that $\tilde{k}_t < \bar{k}_t$ for all $t < j$. However, this implies that $\tilde{k}_1 < \bar{k}_1 = 0$ which is impossible. We showed that there exists a utility maximizing allocation in which an agent never claims disability when he is able.

⁹The budget set is closed since $T_a \leq T_d(t)$ for all t .

Step 2: The constructed asset tested system implements the optimum.

We show that if the conditions of Step 1 are satisfied, the utility maximizing allocation must be feasible and incentive compatible. Therefore, it cannot give a higher utility than (c^*, y^*) .

The allocation is incentive compatible since it comes from the agent's maximization problem.

From Step 1, the able agent always receives a transfer T_a . Corollary 3 shows that $T_a \leq T_d(t)$ for all t , so that this is the lowest possible transfer (the highest tax since $T_a \leq 0$) the agent can receive. Note that if an agent saves more than \bar{k}_i in some period $i - 1$ and becomes disabled in period i , he receives transfer T_a until his savings fall below the asset limit, after which he is entitled to $T_d(i)$. The present value of such transfers is lower than the present value of the transfers to the agent who could claim disability in period i , which is equal to $T_d(i) (1 + \dots + 1/(1 + R)^{T-i})$.

Therefore, the ex ante expected value of transfers cannot be higher than

$$\sum_{t=1}^{t=T} \left(\left(\sum_{i=1}^{i=t} \frac{\mu(h^i = i)T_d(i)}{(1 + R)^{t-1}} \right) + \frac{\mu(h^t = A)T_a}{(1 + R)^{t-1}} \right) \leq 0$$

and from (20) the allocation that has such transfers must be feasible. ■

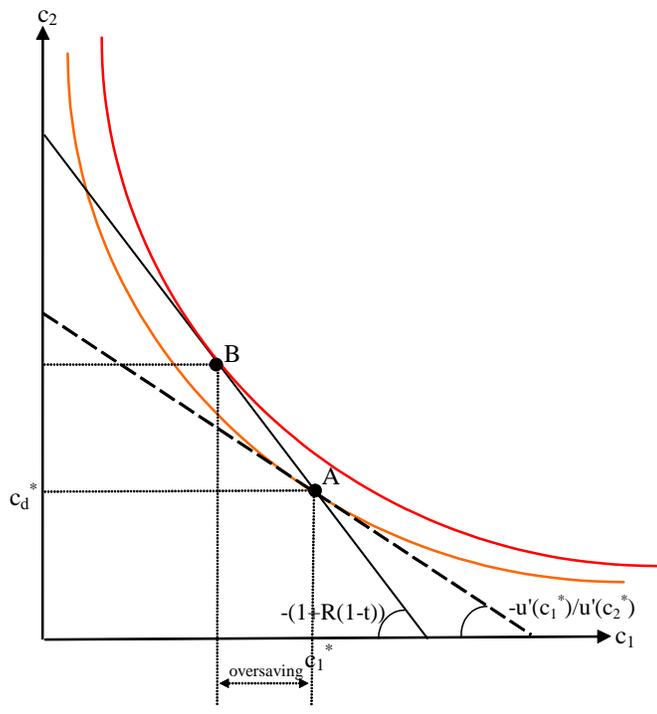


Figure 1: An agent will deviate from optimum in a competitive equilibrium with linear tax

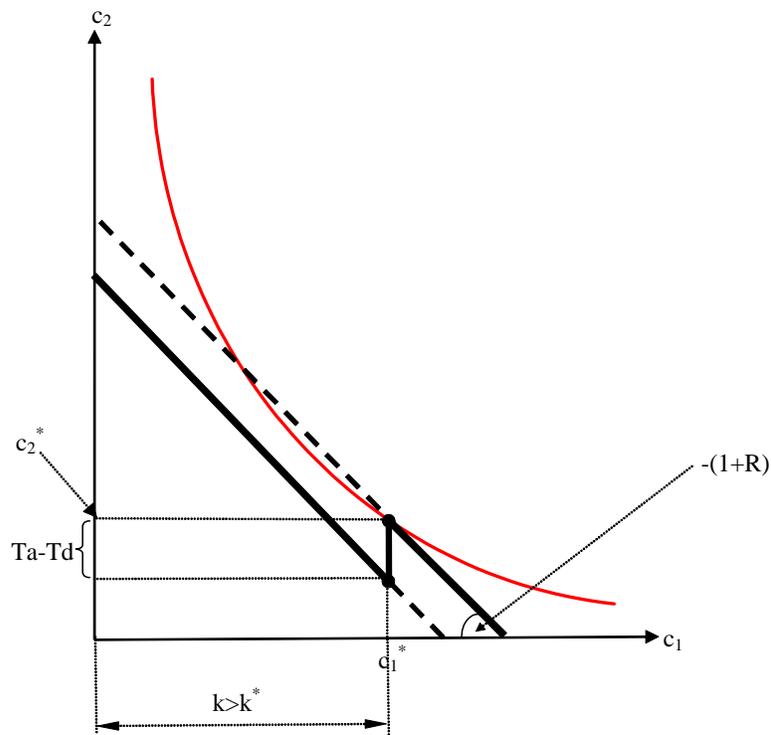


Figure 2: Asset-testing implements the optimum

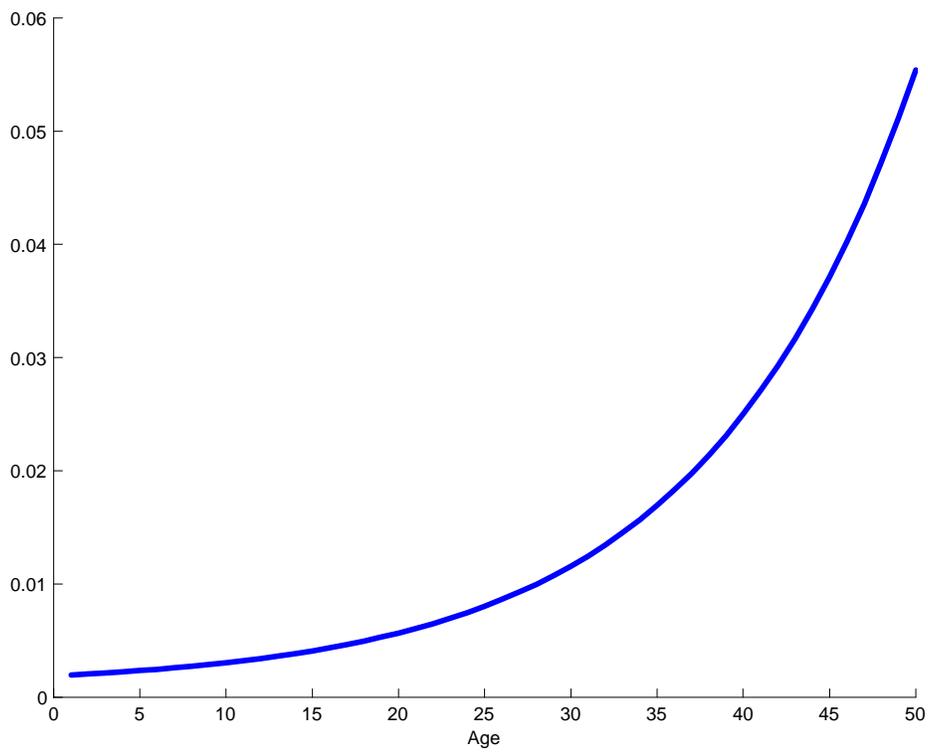


Figure 3: Conditional probability of becoming disabled

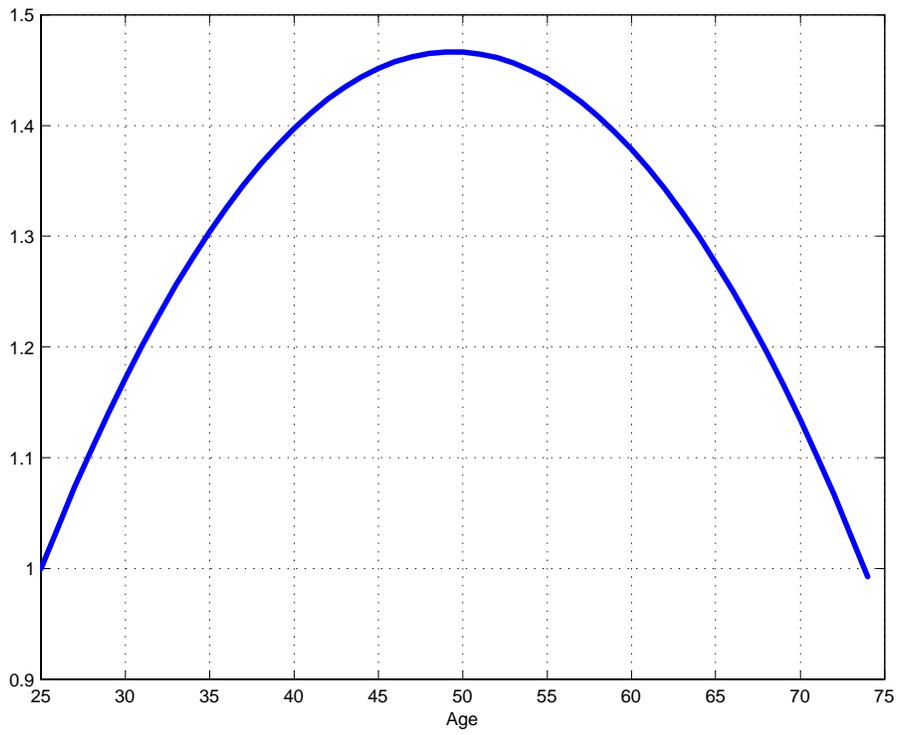


Figure 4: Skill profile

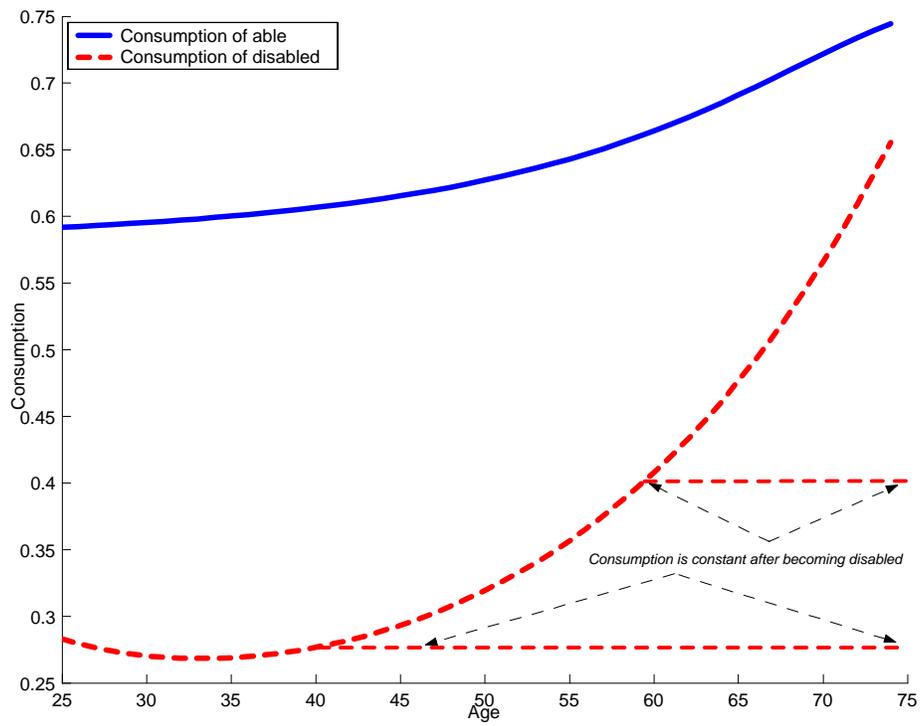


Figure 5: Optimal consumption profiles of agents over lifetime

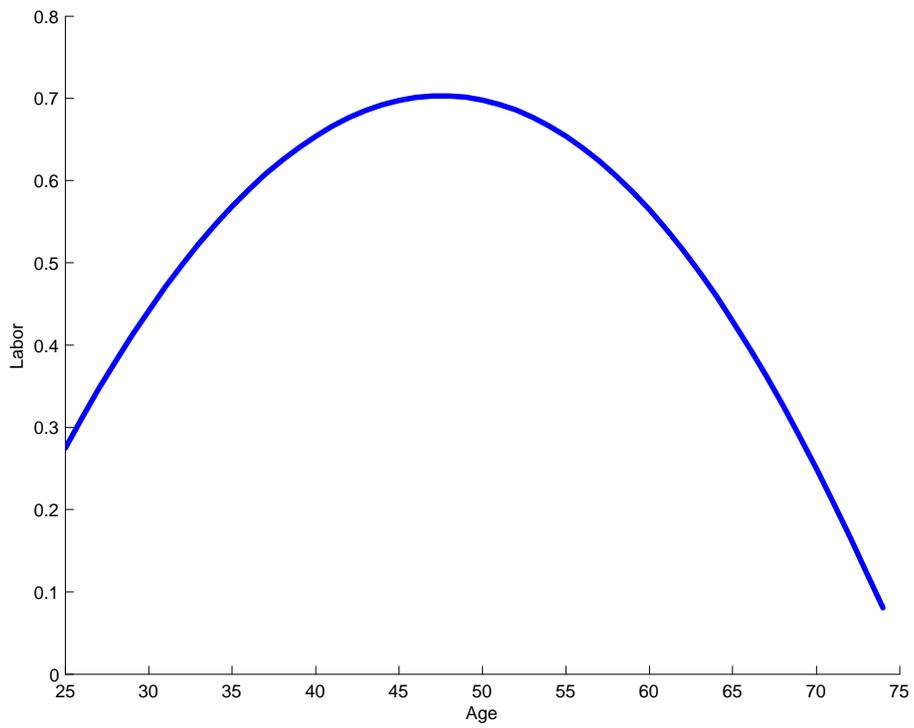


Figure 6: Optimal labor profile

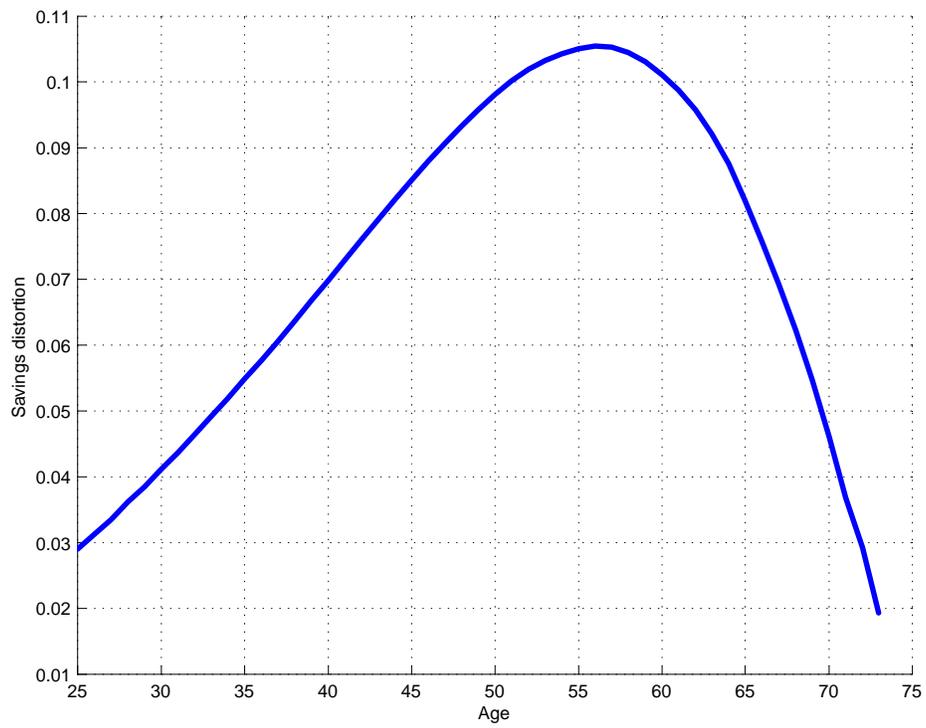


Figure 7: Intertemporal wedge in the optimal system

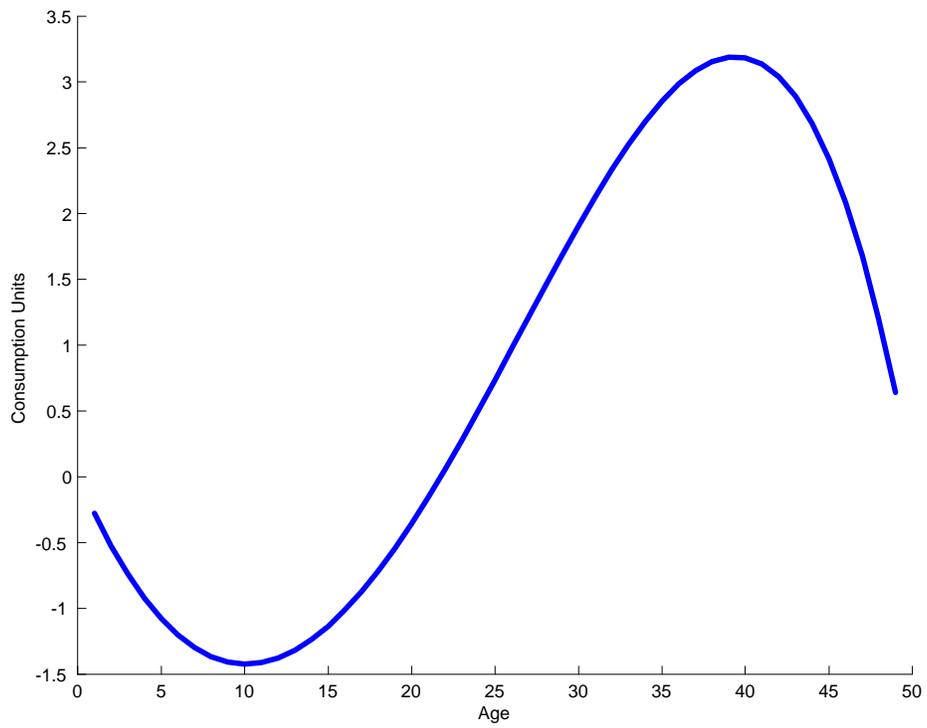


Figure 8: Asset limits implementing optimal system

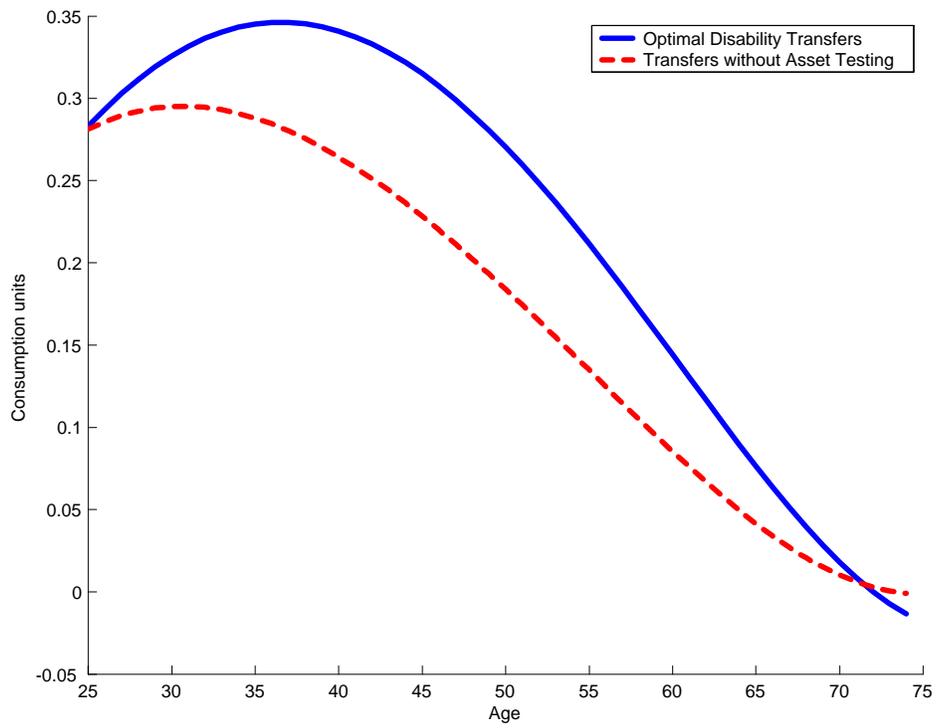


Figure 9: Transfers

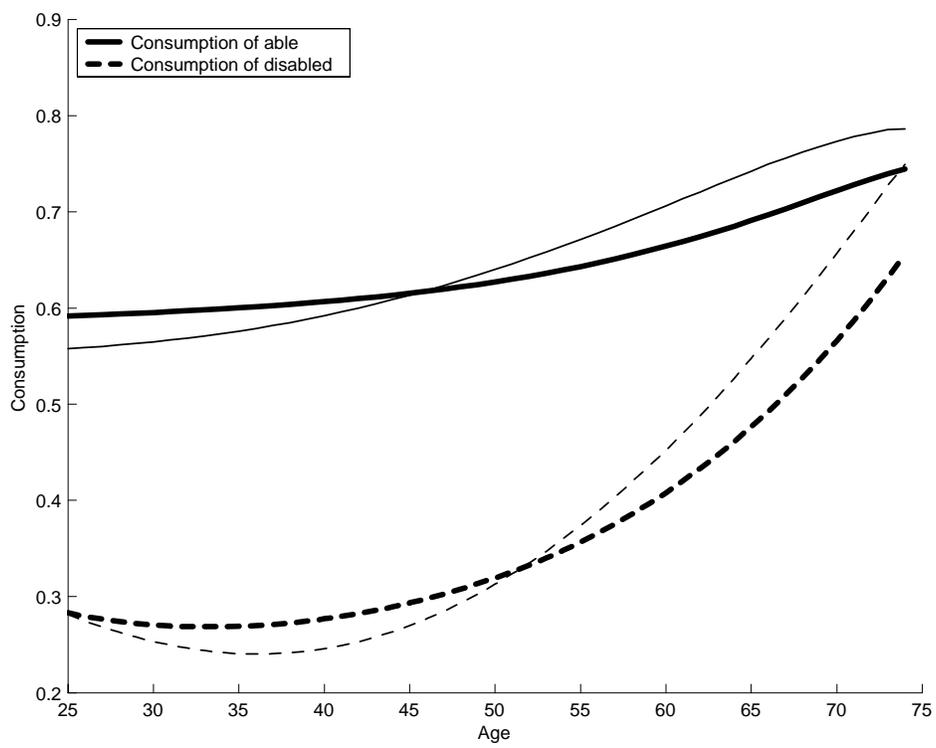


Figure 10: Consumption with asset-test (thick lines) and without asset-test (thin lines).

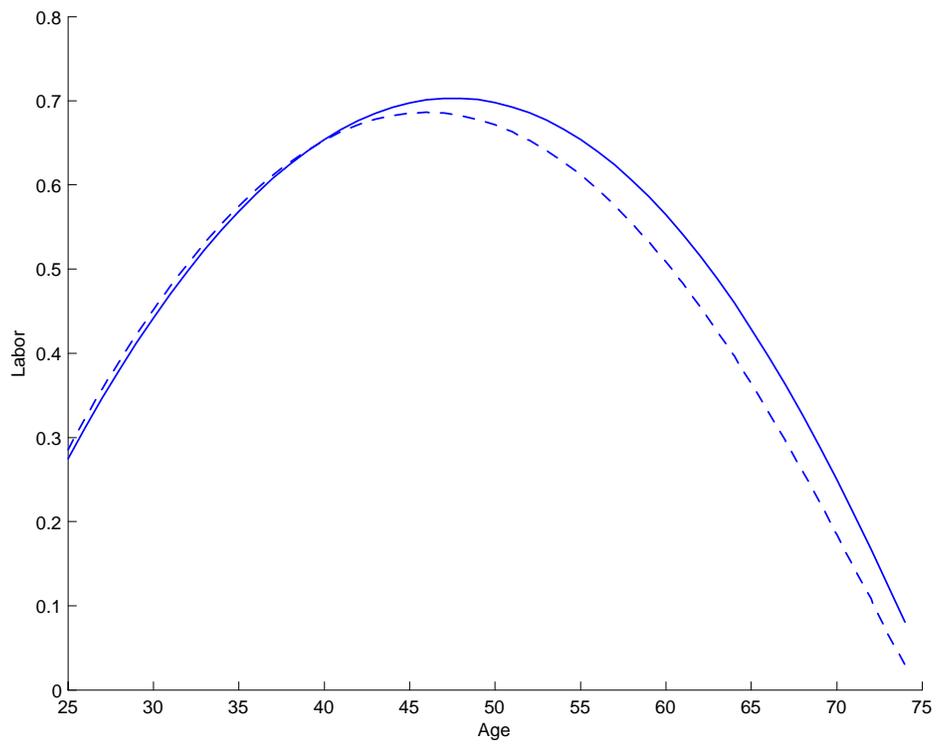


Figure 11: Labor with asset-test (solid) and without asset-test (dashed).