

# Tax Buyouts\*

Marco Del Negro  
Federal Reserve Bank of New York

Fabrizio Perri  
University of Minnesota, Federal Reserve Bank of Minneapolis, CEPR, and NBER

Fabiano Schivardi  
Università di Cagliari, EIEF, and CEPR

March 30, 2010

## Abstract

The paper studies a fiscal policy instrument that can reduce fiscal distortions without affecting revenues, in a politically viable way. The instrument is a *private* contract (tax buyout), offered by the government to each citizen, whereby the citizen can choose to pay a fixed price in exchange for a given reduction in her tax rate for a period of time. We introduce the tax buyout in a dynamic overlapping generations economy, calibrated to match several features of the US income, taxes and wealth distribution. Under simple pricing, the introduction of the buyout is revenue neutral but, by reducing distortions, it benefits a significant fraction of the population and leads to sizable increases in aggregate labor supply, income and consumption.

JEL CODES: E62,H21

KEY WORDS: Taxes, Private Information, Distortions

---

\*The paper previously circulated under the title of “On the Privatization of Public Debt.” We are extremely grateful to our discussant Chris Sleet for excellent comments and suggestions. We thank Da Lin for research assistance, our editor Andy Abel, Lorenzo Forni, Pietro Reichlin, Aleh Tsyvinski and seminar participants at the Fall 2009 Carnegie Rochester Conference on Public Policy, Bank of Italy, NYU, the Bank of Canada, Penn State University, the University of Rochester, the Federal Reserve Bank of Philadelphia, Università di Salerno, the SED, Midwest-Macro and ESSIM meetings for very helpful comments. Perri thanks the NSF (Grant SES-0820519) for financial assistance. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Banks of Minneapolis and New York or the Federal Reserve System.

# 1 Introduction

One of the main legacies of the 2007-8 economic crisis is the huge increase in governments' liabilities, following the expansionary policies enacted to bail out the financial sector and to stimulate economic activity. According to the March 2010 estimates of the CBO, the US federal government will record deficits of 10.3 percent of GDP in 2010 and 8.9 percent in 2011. Against the backdrop of increasing social security and medical assistance liabilities, the pressure to restore fiscal sustainability is already building up: fiscal pressure is bound to increase soon in the US and in many other developed countries. In these circumstances it is fundamental to reconsider the debate on the detrimental role of high taxation on economic activity (see for example Prescott 2004, Mankiw and Weinzierl, 2006). Increased tax rates might not be very effective in improving the fiscal stance if they substantially depress the level of activity.

This paper evaluates an instrument for fiscal policy that is designed to reduce distortions without reducing revenues. The instrument is a private contract between the government and private citizens whereby each agent has the option to pay the government a fixed price in exchange for a given reduction in her marginal tax rate. We refer to this contract as a "tax buyout": the contract makes it possible for the agent to effectively buy out a portion of her distortionary taxes via a lump-sum payment. Participation is entirely voluntary: those individuals that have most to gain from a reduction in the marginal tax rate self-select into the program.

In the first part of the paper we review the working of the tax buyout in a simple model economy with asymmetric information where the price of the contract cannot be made contingent on agents' abilities since the government has only partial information (if any) about these abilities. We show the existence of a simple linear pricing scheme of the tax buyout which is Pareto-improving (abstracting from general equilibrium effects) and at the same time revenue neutral. The intuition for this result is that the inefficiencies generated by distortionary taxation, once removed, create a surplus that the government can share with the agents. Therefore the government can price the contract high enough to make positive revenues, yet low enough to attract agents with high ability and high income prospects. We then discuss the implications of offering a *menu* of contracts such that high ability agents self-select into high tax reduction-high price contracts and discuss the relation between the tax buyout and optimal non linear tax schedules.

Since theory suggests tax buyouts are a simple and effective way to reduce distortions,

in the second part of the paper we ask our main question, i.e., what are the effects of making these contracts available to US residents in a time of high fiscal pressure?

To answer the question we introduce tax buyouts in a dynamic overlapping generation model with heterogeneous agents and incomplete markets as those used by Conesa, Kitao and Krueger (2008) or Heathcote, Storesletten and Violante (2008) for policy analysis. We calibrate the model parameters to match existing micro-studies and several features of the US income and wealth distribution derived from the Current Population Survey and the Survey of Consumer Finance. We start from a steady-state equilibrium with a fixed government spending financed through a progressive tax income schedule which resembles the one observed in the US. We then consider an exogenous increase in government spending (capturing the current fiscal situation) of the order of 20% under two alternative scenarios: with and without a tax buyout scheme. The baseline buyout we consider is a reduction in the marginal tax rate of at most 5% offered at a price of roughly \$4500, which ensures revenue neutrality.

We find that the buyout in each period is purchased by a little less than 10% of the population, nevertheless its effects are macroeconomically relevant as its introduction raises GDP by almost 1%, reducing by approximately by one third the negative consequences of the overall tax increase on the level of economic activity. This is because buyers are generally high ability agents, who contribute most to GDP. We also find that these numbers are fairly stable with respect to comparative statics exercises on a series of key parameters, such as the elasticity of labor supply or the amount of the tax buyout offered. Finally, we show that, in our dynamic setting, even households not buying the contract in the current period might benefit from its introduction in terms of lifetime welfare. For example, since wages increase with age, a young household might not buy the contract today but do so at a future date, when her wage has increased sufficiently. This indicates that the long run benefit of introducing this instrument spreads over a larger share of the population than current buyers.

The idea of the tax buyout is very much related to the idea of offering taxpayers alternative tax schedules and letting them choose among them. Alesina and Weil (1992) and Slemrod et al. (1994) propose such schemes in a static setting. Relative to their work our contribution lies in the quantitative evaluation of the buyout in a dynamic economy that captures several macro moments and key dimensions of heterogeneity of the current US economy. Also we study in more details the issue of non linear pricing and the relation of these schemes with the non linear optimal taxation literature (Mirrlees, 1971). In particular

the contract generates tax schedules that, under some conditions, are reminiscent of those proposed by that literature as it delivers lower marginal rates for high ability individuals. Differently from those mechanisms, from the agents' perspective the exchange is not the result of a unilateral determination on the part of the government, but of free choice. This ensures that the allocation achieved with the contract is a Pareto improvement with respect to that without it, making it a politically viable mechanism of reduction of the distortionary effects of taxation. However, more recent contributions have proposed various instances in which a concave tax schedule might not be optimal. For example, Diamond (1998) and Saez (2001) show that the shape of the optimal tax function depends, among other things, on the distribution of ability in the population. In particular, fat-tail unbounded distributions such as the Pareto can make increasing marginal rates socially desirable. Also, under these conditions a price of the contract that guarantees revenue neutrality may not exist. In these cases, the tax buyout would not be a useful policy tool.

The rest of the paper is organized as follows. In section 2 we review some general results on the tax buyout scheme in a simple static economy. Section 3 introduces the OLG model and discusses the calibration and section 4 uses the model to evaluate the effects of introducing the buyout contract following an large increase in government spending. In the conclusions we discuss additional issues to tackle on the way to making the buyout scheme a concrete policy option, a goal especially important in a period of unprecedented fiscal deficits.

## 2 A simple model with heterogeneous productivity

In this section we consider a small open economy populated by a continuum of agents. Agents' productivity  $A$  is constant over time and distributed according to the CDF  $F_A(A)$ . Agents are endowed with initial amount of wealth  $\tilde{w}$ . The government uses labor taxes to finance a constant flow of government spending  $g$ . We consider the simple case where the tax rate is constant and equal to  $\tau$ . Private agents and the government have access to the world capital market where they can borrow and lend at the risk-free rate  $r$ . We assume that this exogenous risk-free rate equals the inverse of the discount rate:  $\beta(1+r) = 1$ .

Our modification to this otherwise standard set-up is that in each period the government offers each agent a contract, whereby the agent agrees to pay an amount of resources upfront in exchange for a reduction in her tax rate. If the agent does not buy the contract, she faces the pre-contract tax rate  $\tau$ : status quo is always an option. Under perfect information the

contract is designed in such a way that the value of resources extracted from each agent remains the same as in the pre-contract regime, regardless of the amount of tax reduction the agent chooses to purchase. That is, the contract is revenue neutral from the government’s perspective, agent by agent. We show that under these conditions agents will want to pay all their taxes upfront in lump-sum form. The logic of this result is straightforward: The contract gives agents the opportunity to turn distortionary taxation into lump-sum taxation, and revenue neutrality implies that all the benefits from the reduction in distortions accrue to the agent. The economy with perfect information is a useful starting point for our analysis, as it highlights the basic insights of the contract, that will hold also in a more realistic setting. However, under perfect information, a benevolent government could more simply use lump-sum taxation – where the lump-sum taxes are a function of the fixed component of ability  $A$  – in the first place.

We then turn to study an economy with asymmetric information, where the government knows the ability distribution but does not observe individual ability. The question is whether the introduction of the contract can be Pareto-improving in spite of this constraint. We show that this is the case. The intuition for this result is that the inefficiencies generated by distortionary taxation, once removed, create a surplus that the government can share with the agents. Therefore the government can price the contract high enough to make positive revenues, yet low enough to be attractive to the high ability agents.

## 2.1 The economy

Agents’ preferences and technology are given by:

$$\sum_{t=0}^{\infty} \beta^t u(c_t, l_t), \tag{1}$$

and

$$y_t = Al_t, \tag{2}$$

respectively, where  $c_t$  and  $l_t$  are consumption and labor in period  $t$ ,  $y_t$  is labor income,  $\beta > 0$  is the time discount factor and  $u$  is a standard utility function. To simplify notation, define  $w = r\tilde{w}$  as the flow value of wealth. In addition, the assumption  $\beta(1+r) = 1$  implies that the problem for this  $(A, w)$ -agent is stationary and can be simplified as follows:

$$\max_{\{c(A,w,\tau), l(A,w,\tau)\}} u(c(A, w, \tau), l(A, w, \tau)) \tag{3}$$

subject to:

$$w + (1 - \tau)Al(A, w, \tau) - c(A, w, \tau) = 0 \quad (4)$$

where  $c(A, w, \tau) > 0$ ,  $l(A, w, \tau) \in [0, 1]$ , and the initial wealth level is given. For simplicity, we abstract from differences in wealth. All arguments go through when explicitly conditioning on differences in wealth. In the stationary equilibrium the government budget constraint is:

$$g = \tau \int Al(A, w, \tau) dF_A(A). \quad (5)$$

Stationary equilibria for this economy are characterized by policy functions  $c(A, w, \tau)$  and  $l(A, w, \tau)$ , and a Lagrange multiplier  $\lambda(A, w, \tau)$  associated with (4) satisfying:

$$u_c(c(A, w, \tau), l(A, w, \tau)) = \lambda(A, w, \tau) \quad (6)$$

$$A(1 - \tau)u_c(c(A, w, \tau), l(A, w, \tau)) = -u_l(c(A, w, \tau), l(A, w, \tau)) \quad (7)$$

as well as the budget constraint (4), and a tax rate  $\tau$  satisfying (5).

## 2.2 The economy with the contract: perfect information

At the beginning of each period each agent is offered a menu of contracts  $\{\delta, D\}$  from which she can buy a reduction  $\delta$  in its tax rate for the current fiscal year in exchange for the payment of a premium  $D$ . Under perfect information the ability level  $A$  is known to both the government and the agent. We assume the government chooses  $D$  such that the contract is revenue neutral *for each agent*. Therefore  $D$  is a function of both  $A$  and  $\delta$  (as well as  $w$ , non stated explicitly to simplify notation) that satisfies:

$$D(A, \delta) = \tau Al(A, w, \tau) - (\tau - \delta)Al(A, w - D, \tau - \delta) \quad (8)$$

where  $l(A, w - D, \tau - \delta)$  is the policy function for labor under the new regime. Note that the government is exploiting its knowledge of the agents' policy function in designing the contract.

In expression (8) we have used the same policy function for labor supply described in the previous section, with the only difference that the arguments  $w$  and  $\tau$  are replaced by  $w - D$  and  $\tau - \delta$ , respectively. Indeed, one can see that for any given  $\delta$  these policy functions

satisfy the first order conditions as well as the new budget constraint:

$$(w - D) + (1 - \tau + \delta)Al(A, w - D, \tau - \delta) - c(A, w - D, \tau - \delta) = 0. \quad (9)$$

Moreover, since the contract is such that the government's intertemporal budget constraint is unchanged for each agent, it is also unchanged in the aggregate:

$$\begin{aligned} & \int \tau Al(A, w, \tau) dF_A(A) \\ &= \int ((\tau - \delta)Al(A, w - D, \tau - \delta) + D(A, \delta)) dF_A(A) \end{aligned} \quad (10)$$

In the remainder of the paper we will use a short-hand notation  $l_A(\delta, D) = l(A, w - D; \tau - \delta)$ ,  $c_A(\delta, D) = c(A, w - D; \tau - \delta)$ , and  $\lambda_A(\delta, D) = \lambda(A, w - D, \tau - \delta)$  to denote aggregate (ability adjusted) labor supply, consumption, and Lagrange multiplier, respectively, for a  $(A, w)$ -cohort. Thus,  $l_A(0, 0)$  is labor supply without the contract. Next we turn to the question of the optimal choice of  $\delta$ , the amount of tax reduction:

**Proposition 1** *If the government prices the contract according to (8), agents will choose to buy the maximum possible tax reduction. Moreover, labor supply will increase for each agent.*

*Proof:* see the Appendix

The proof uses the fact that the contract can be represented by the combination of a compensated price change and a rightward shift of the budget constrain (defined in terms of consumption and leisure). As a consequence, the pre-contract optimal allocation remains feasible after the purchase of the contract. Hence, purchasing the contract cannot but raise the agent's utility. Moreover, at the new prices the lower tax wedge implies that agents will choose to work more.

### 2.3 The economy with the contract: asymmetric information

This section studies the economy with asymmetric information. We assume that the government has no information whatsoever on  $A$ , and hence needs to offer everyone the same price for the contract  $D(\delta)$ . We want to show that the contract can be Pareto improving even under asymmetric information. Namely, we want to show that there exists a function  $D(\delta)$  such that: (i) a positive mass of agents takes the contract, and (ii) the government budget is still balanced. The second condition implies that those agents who do not take

the contract are made no worse off by its introduction. We will prove (i) and (ii) for a linear pricing function of the form  $D(\delta) = d\delta$ . Of course, more general pricing functions will allow the government to improve on the outcome obtained with the linear pricing function. We will explore this aspect numerically later.

First, we discuss two results on the optimal choice of  $\delta$  for linear pricing functions.

**Lemma 1** *Under a linear pricing function  $D(\delta) = d\delta$ , an agent of ability  $A$  will either not buy into the contract or buy the maximum amount allowed.*

*Proof: see the Appendix*

According to Lemma 1, linear pricing functions are completely characterized by the maximum allowed tax reduction and the per-unit price of such reduction,  $\{\bar{\delta}, d\}$ . For each level of ability we define two (per- $\delta$ -unit) prices,  $\underline{d}_A$  and  $\bar{d}_A$ . The former is the lowest price at which the government is not losing resources from offering the contract to agent  $A$ . The latter is the price for which agent  $A$  is indifferent between taking and not taking the contract.

**Lemma 2** *Under a linear pricing function  $D(\delta) = d\delta$ , for each level of ability  $A$  there exist:*

- i) A per-unit price  $\underline{d}_A$  such that agent  $A$  is willing to enter the contract and the government is neither losing nor gaining resources from the agent.*
- ii) A per-unit price  $\bar{d}_A$  for which agent  $A$  is indifferent between taking and not taking the contract. If agent  $A$  decides to enter the contract, the government is gaining resources from the agent.*

*Proof: see the Appendix*

A consequence of Lemma 2 is:

**Corollary 1** *For  $d \in (\underline{d}_A, \bar{d}_A)$  the government is gaining resources from offering the contract to agent  $A$ , and agent  $A$  is willing to take the contract.*

We have just shown that for each level of ability there is a price that is high enough to make government's revenues positive, yet low enough to be attractive to the agents. The removal of the inefficiency due to distortionary taxation creates a surplus that the government and the agent can share.

Building on this result, we now show that the government can attract into the contract the upper tail of the ability distribution, and still make positive revenues. Once we have shown this, it follows that the contract is Pareto-improving: the government could either rebate these excess revenues to all agents, or lower the price even further to attract more agents into the contract. Here we pursue the latter route, and show that if the price is low enough the government's excess revenues are driven to zero.

**Lemma 3** *Assume the single-crossing condition (Salanié 2003) holds. If agent with ability  $A(d)$  is indifferent between buying and not buying the contract under price  $d\bar{\delta}$ , all agents with ability  $A > A(d)$  will choose to purchase the contract.*

*Proof: see the Appendix*

Define the *marginal ability*  $A(d)$  as the level of ability for which  $\bar{d}_A = d$ . This function is implicitly defined by the equation

$$u(c_A(\bar{\delta}, d\bar{\delta}), l_A(\bar{\delta}, d\bar{\delta})) = u(c_A(0, 0), l_A(0, 0))$$

Such level is well defined for any  $d \leq \bar{d}_A$ , where  $\bar{A}$  is the upper bounds of the support of the ability distribution. The standard regularity conditions on the utility functions imply that the mapping  $A(d)$  is continuous. As a consequence of Lemma 3 agents with ability less than  $A(d)$  will not take the contract, while agents with ability greater than  $A(d)$  will. Therefore government's revenues from the introduction of the contract can be written as:

$$R(d) = \int_{A(d)}^{\bar{A}} \left[ (\tau - \bar{\delta})Al_A(\bar{\delta}, d\bar{\delta}) - \tau Al_A(0, 0) \right] dF_A(A) + d\bar{\delta}(1 - F_A(A(d))) \quad (11)$$

It is easy to show that continuity of  $A(d)$  implies that  $R(d)$  is continuous for  $d < \bar{d}_A$ .

From Lemma 2 we have that the government can attract to the contract the upper tail of the ability distribution, and still extract excess revenues from all of these high ability agents if  $d > \underline{d}_A$ :

**Corollary 2** For  $d \in [\underline{d}_A, \bar{d}_A)$  all agents with ability  $A \in [A(d), \bar{A}]$  will take the contract, where  $A(d)$  is strictly less than  $\bar{A}$ . Moreover,  $R(d) > 0$  in this interval.

Our final assumption is that, when the contract is priced at zero, the government is losing money  $R(0) < 0$ . This is simply assuming that we are not on the declining portion of the Laffer curve, where a tax decrease leads to an increase in revenue, in which case the government should simply lower taxes. The above corollary, the continuity of  $R(d)$ , and the above assumption imply the key result of this section:

**Proposition 2** There exist a pricing function  $D(\delta) = d^* \delta$ , with  $d^* \in (0, \underline{d}_A)$ , such that: (i) a positive mass of agents with  $A \in [A(d^*), \bar{A}]$  enter the contract; (ii) the government's budget is balanced:  $R(d^*) = 0$ .

Since those who enter the contract are better off, and those who do not enter the contract are no worse off given that the government still balances the budget, we have shown that the contract is Pareto improving. Figure 1 below gives an illustration of the proposition, as it plots the government losses as a function of  $d$ , the price of the contract.<sup>1</sup> Starting from the right the figure shows that as the price  $d$  is large losses are close to 0, as only few reach agents elect to purchase the buyout. As the price falls more households participate and initially the revenues from higher participation offset the losses from lower prices, so that government revenues increase (losses fall). As the price is lowered further though the price effect dominates the increased participation, and losses increase. When the price reaches  $d^*$  losses are 0, so the government does not lose from introducing the contract and a positive fraction of the population takes it.

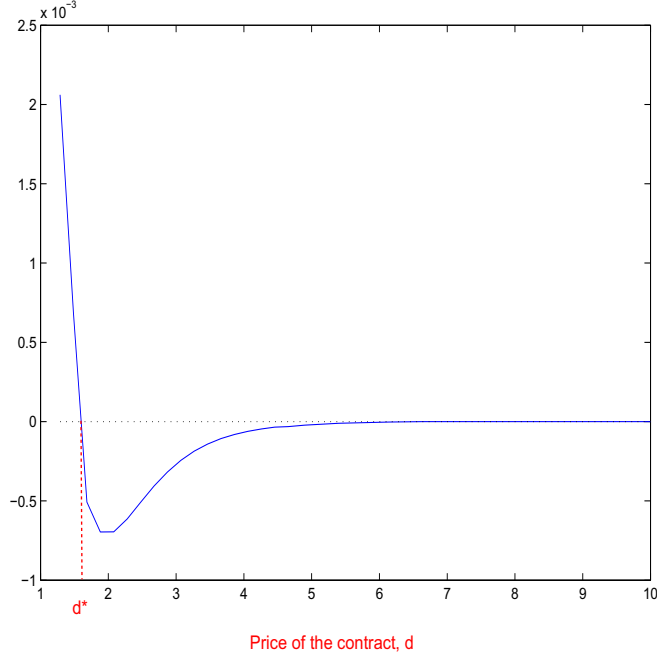
An important caveat needs to be mentioned. We have assumed an upper bound to the ability distribution, and this assumption is quite important in obtaining our result. However, a fat right tail of the ability distribution may imply that the government makes very large revenue losses on high ability individuals when implementing the contract, jeopardizing revenue neutrality. In fact, one can easily show that with a Pareto distribution, under quasi-linear utility, for some parameter values revenue neutrality is never achieved.

## 2.4 Non linear pricing

The theory of static Pareto efficient taxation (Mirrlees 1971, Stiglitz 1982) argues the optimal tax schedule entails different marginal tax rates for individuals with different abilities

<sup>1</sup>The figure is drawn for a log-normal distribution of abilities.

Figure 1: GOVERNMENT LOSSES AND THE PRICE OF THE BUYOUT



so to achieve self-selection. We have excluded this possibility so far. Starting from proportional taxes and a continuum of abilities – the case considered in the previous section – the post-contract equilibrium entails two different marginal tax rates for the entire population, which is obviously far from the continuum suggested by the optimal taxation literature. This feature of the contract can be relaxed. Although, for computational feasibility, in the quantitative assessment of the buyout we will stick to linear schemes, in this section we investigate the effects of allowing for non-linear pricing schemes in the simple model.

We consider a discretized version of our economy, where ability takes up  $N$  values  $A_i$ ,  $i = 1, 2, \dots, N$ , with  $A_i < A_{i+1}$ . We call an agent with ability  $A_i$  agent  $i$ . The government offers  $N$  contracts  $\{\delta_i, D_i\}_{i=1}^N$ . We call this discretization “many contracts” as opposed to the “one contract” of the previous section. In the appendix we describe an algorithm for computing the set of  $N$  contracts that delivers self-selection, where by self-selection we mean that the pricing scheme is such that any two agents  $A_i$  and  $A_j$  will choose different contracts,  $\{\delta_i, D_i\}$  and  $\{\delta_j, D_j\}$  respectively, in case they decide to buy the reduction in the marginal tax rate. This self-selection implies that, consistently with the optimal taxation literature, agents with different abilities will face different marginal

tax rates. The idea of the algorithm is the following. From optimal taxation literature we know that the highest ability individual  $A_N$  should have the lowest marginal tax rate. The government then chooses the couple  $\{\delta_N, D_N\}$  with  $\delta_N$  fixed at the highest allowed amount. The price of the contract  $D_N$  is chosen so that the entire scheme  $\{\delta_i, D_i\}$  is revenue neutral. Then, a second contract  $\{\delta_{N-1}, D_{N-1}\}$  is determined such that agent  $N$  is indifferent between contract  $\{\delta_N, D_N\}$  and  $\{\delta_{N-1}, D_{N-1}\}$  while agent  $N-1$  strictly prefers the latter. This procedure is iterated for all  $i$  until we get to the agent who is indifferent between buying and not buying the marginal rate reduction. Although the algorithm uses only pairwise comparisons of contracts, in the appendix we show that it ensures that agent  $i$  chooses contract  $i$  over any other contract.

Once computed the  $\{\delta_i, D_i\}_{i=1}^N$  schedule that achieves self-selection, we can use it to analyze its properties numerically. The appendix provides details on the calibration of the economy used to obtain the results below. Figure 2 compares the “many contracts” to the “one contract” with  $\bar{\delta} = .13$ . The value of  $\bar{\delta}$  in the linear case and the  $\{\delta_i, D_i\}_{i=1}^N$  schedule in the non-linear case are chosen so that the fraction of agents entering the tax buyout scheme is approximately the same. In all panels the horizontal axis displays the pre-contract income ( $Al_A(0, 0)$ ), which is a monotone function of ability. Panel (a) displays the chosen  $\delta$  as a function of income. The lines stop at the ability level (expressed in terms of pre-contract income) below which agents prefer not to purchase the contract. As shown in the previous section, in the linear case all agents buying the contract choose the maximum level of  $\delta$ . In the non-linear case  $\delta_i$  increases in the pre-contract income. In the example considered here the increase is approximately linear.

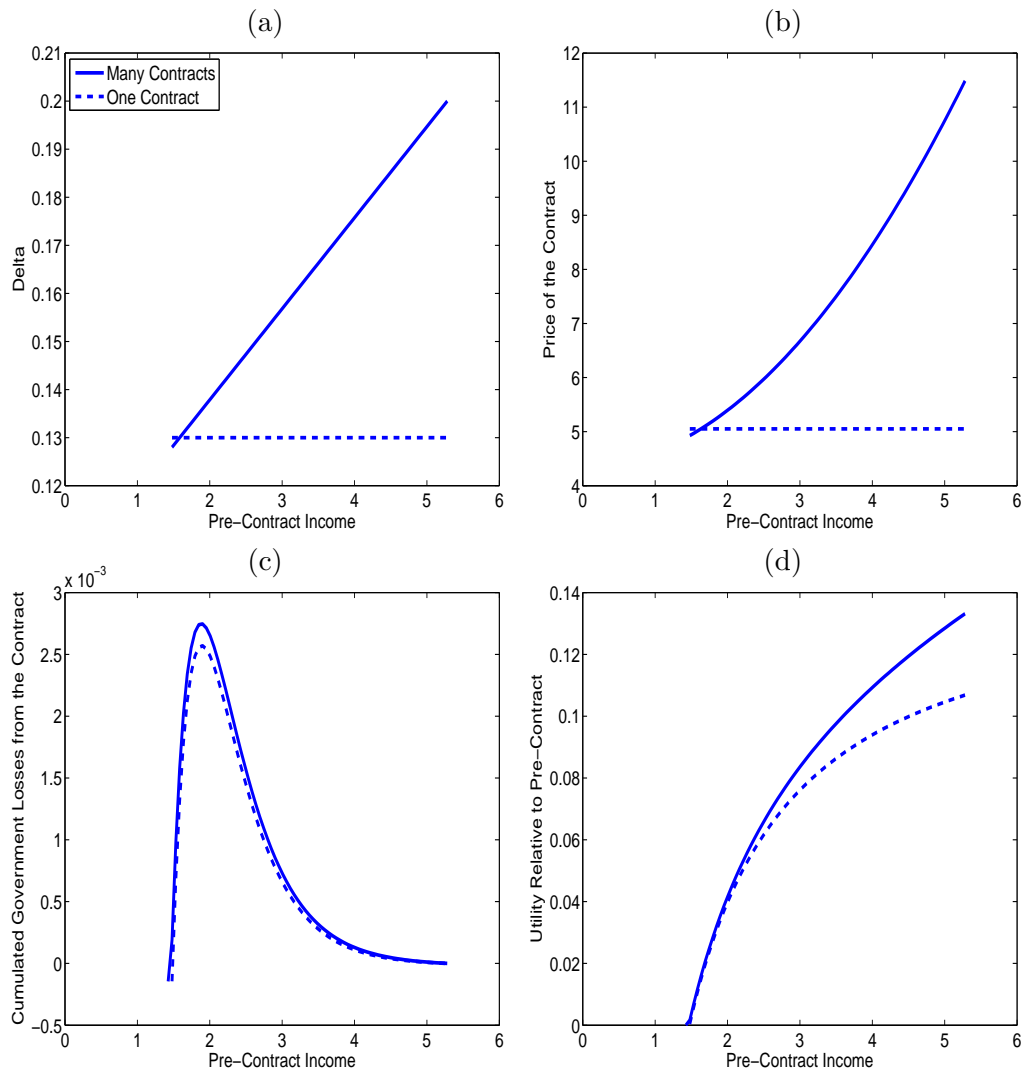
Panel (b) displays the associated price of the contract  $D(\delta)$  as a function of income. In the linear case all agents buying the contract pay the same price. In the non-linear case, the price  $D(\delta)$  is a convex function of  $\delta$ , which is what we would expect from the household’s maximization problem with respect to  $\delta$ . In fact the first order conditions of the agent with respect to  $\delta$  for the model in section 2.1 are:

$$\frac{\partial u}{\partial \delta} = \left( Al_A(\delta, D) - \frac{\partial D}{\partial \delta} \right) \lambda_A(\delta, D), \quad (12)$$

The fact that the first term within parentheses increases in  $A$  implies that  $\frac{\partial D}{\partial \delta}$  also increases in the level of ability.

Panel (c) shows the cumulated government losses from the offering contract, starting from the highest ability individuals (hence, we accumulate losses going from right to left).

Figure 2: ONE VERSUS MANY CONTRACTS



By construction, in both cases the total government losses from the contract are approximately zero (i.e, both lines reach zero for the indifferent agent). The shape of the cumulated government losses is very similar for the linear and the non-linear pricing schemes in this specification. In both cases the government loses revenues on high ability individuals: In the linear pricing scheme  $\delta$  is lower than in the non-linear pricing scheme, but so is the price. In both cases, consistently with the theory, the government makes positive revenues on the agents that are close to being indifferent.

Panel (d) shows utility relative to pre-contract. The figure shows that the non-linear scheme is a Pareto improvement over the linear one: for all agents utility increases. For the close-to-indifferent agents, not surprisingly, utility is very close under the two schemes, but as the level of ability increases, the non-linear scheme delivers progressively higher utility. As figure (c) makes clear, the “many contracts” environment allows the government to deliver higher utility to high ability individuals without having much greater losses. Of course, the lower distortions for high ability individuals under “many contracts” result in higher labor supply and output than in the “one contract” case.

We have also compared the same “one contract” scheme ( $\bar{\delta} = .13$ ) to a “many contracts” environment where the top ability individual is offered a contract with a tax reduction of at most .13 ( $\delta_N = .13$ ). Unlike in the previous example, this set of contracts makes high ability individuals worse off than under the one contract scheme, since the highest ability individual is now paying for the same tax reduction a higher price than under the linear scheme. As a consequence, if for high levels of ability the government is losing money under the “one contract” scheme, the government is losing less money under “many contracts.” The extra resources make it affordable to both charge a lower price to individuals that before were close to being indifferent, and extend the contract to less able people. Both sets of people are better off under the “many contracts” scheme. In our example it also turns out that aggregate output is larger under many contracts, even though higher ability individuals are subject to more distortions.

## 2.5 The contract and the informationally constrained Pareto frontier

In this section we briefly discuss the conditions under which the contract can take the economy to the informationally-constrained Pareto-efficient frontier, that is, can replicate the outcome of Mirrlees’ optimal taxation. In order to discuss issues of optimal taxation we need to consider more general forms of taxation  $T(y_A)$ , where  $y_A = Al_A$ , than the

proportional taxation analyzed so far. Equilibrium in this economy is characterized by policy functions  $c_A$ ,  $l_A$ , and  $\delta_A$  satisfying

$$c_A = Al_A - T(Al_A) + \delta_A Al_A + (w - D(\delta_A)), \quad (13)$$

$$\frac{-\partial u(c, l)/\partial l}{\partial u(c, l)/\partial c} = A(1 - T'(Al_A) + \delta_A), \quad (14)$$

as well as the first order condition with respect to  $\delta$  (12), which is unchanged. Call  $\tilde{T}(\cdot)$  the solution of the constrained social planner's problem and  $\tilde{c}_A$  and  $\tilde{l}_A$  (and  $\tilde{y}_A = A\tilde{l}_A$ ) the corresponding equilibrium allocations, which satisfy (13) and (14) with  $\delta = D(\delta) = 0$ . We want to investigate under which conditions the contract can deliver the Pareto-optimal allocation starting from an arbitrary initial tax schedule  $T(y)$ .

From (14) the amount of tax reduction  $\delta_A$  an agent of ability  $A$  needs to buy is:

$$\tilde{\delta}_A = T'(\tilde{y}_A) - \tilde{T}'(\tilde{y}_A). \quad (15)$$

The tax reduction  $\delta_A$  is positive only if the marginal tax rate at the optimum is lower than the current one. Hence a first, quite intuitive, constraint on what the contract can achieve is that the slope of optimal tax schedule needs to be no higher than the existing one. In other words, the contract can only be used to induce a reduction in marginal tax rates, and not an increase. In addition, participation in the buy-out scheme is voluntary: agents must be made better off by purchasing the contract. This induces additional constraints relative to the informationally constrained planner.

We can reverse-engineer which pricing function  $D(\delta)$  may work from the first order condition with respect to  $\delta$  (12). This condition states that the pricing function that induces an agent of ability  $A$  to buy precisely the quantity  $\delta_A$  of tax reduction needs to satisfy the necessary condition:

$$D'(\tilde{\delta}_A) = A\tilde{l}_A = \tilde{y}_A. \quad (16)$$

Using the second order condition it is immediate to show that the sufficient condition for the optimum is:

$$D''(\delta_A) = \frac{1}{T''(\tilde{y}_A) - \tilde{T}''(\tilde{y}_A)} > A \frac{\partial l_A}{\partial \delta}. \quad (17)$$

Condition (17) says that the extent to which we can implement the optimal allocation with the contract depends on where we want to get to (which point on the Pareto-efficient

frontier) relative to where we are (current tax system). Specifically, the condition tells us that  $D(\delta_A)$  needs to be sufficiently convex and this is the case only if  $\tilde{T}''(\tilde{y}_A) < T''(\tilde{y}_A) - 1/A \frac{\partial l_A}{\partial \delta}$ , that is, only if the optimal marginal tax schedule is decreasing more rapidly with income than the current one. If the current tax system is proportional, the condition translates to  $\tilde{T}''(\tilde{y}_A) < -1/A \frac{\partial l_A}{\partial \delta}$ , i.e., the contract can only get to points on the frontier where the marginal tax rate decreases with income. Indeed, the literature has shown that the optimal tax schedule is not necessarily concave (Diamond, 1998; Saez, 2001). In particular, fat-tailed, unbounded distributions for ability can require convex tax schedules, in which case the contract cannot take the economy on the frontier. Moreover, the larger the gap  $T''(\tilde{y}) - \tilde{T}''(\tilde{y})$  – the reduction in marginal tax rates as a function of income – the harder it is to implement it with the contract. Intuitively, to get a large reduction in marginal tax rates as a function of income one needs  $\delta$  to increase substantially as income grows. That implies that  $D(\delta)$  cannot be too convex. But if  $D(\delta)$  is not too convex, the second order conditions may be violated.

### 3 Quantitative analysis

In this section we present a calibrated model economy which will allow us to evaluate more precisely the effects of making a tax buyout contract available to US households. Our setting is similar to the one used in recent works that use calibrated life-cycle heterogeneous agents economies in order to analyze policy issues (see for example Conesa et al., 2008 or Heathcote et al., 2008 and 2010, Fukushima, 2010). We consider a discrete time small open economy inhabited by overlapping generations of finitely lived households. Production is carried out by a representative firm which uses a constant return to scale technology to produce a single good, used for consumption and investment. A government levies taxes, provides transfers and public spending and might offer tax buyouts. We now describe in more detail households, firms and government, define equilibrium, describe the calibration procedure and finally perform the experiment of introducing a tax buyout.

### 3.1 Households

Each period a new generation of mass  $1/(N + 1)$  of unitary households is born. Preferences of a household at birth are represented by

$$E_0 \left( \sum_{h=1}^N \beta^h u(c_h, l_h) + \beta^{N+1} V_{N+1} \right)$$

where  $c_h$  and  $l_h$  are consumption and labor effort of the household at age  $h$ ,  $u$  is a standard utility function,  $\beta > 0$  is the time discount factor (possibly household specific) and  $V_{N+1}$  represents the value of retirement (to be specified later).<sup>2</sup> Each household is born with a wealth endowment  $a_0$  and is endowed with 1 unit of time in each period of its life. For the first  $N$  years of life households can work in the market and each unit of time spent working yields  $e_h$  efficiency units of work, where  $e_h$  is given by

$$e_h = e^{Az_h \varepsilon_h p_h}$$

where  $A$  is an individual specific fixed effect,  $z_h$  is a deterministic age effect,  $\varepsilon_h$  is an idiosyncratic transitory shock and  $p_h$  is an idiosyncratic persistent shock. All these 4 states are exogenous, in the sense that they do not depend on any household decision. For simplicity we will denote with the letter  $S = (A, z_h, \varepsilon_h, p_h)$  these exogenous states and by  $\mathcal{S}$  the set of all possible values taken by  $S$ .

#### 3.1.1 Retirement

During retirement households are not allowed to work. The only role played by retirement in our setup is to affect the wealth accumulation motive during working life. We capture this effect in this simple fashion

$$V_{N+1} = V_{N+1}(a, S) \equiv \frac{u(\phi a + T^R(S), 0)}{(1 - \beta) + \beta\phi} \quad (18)$$

where  $\phi a$  represents the annuity value of wealth and  $T^R(S)$  represents the annuity value of all social security payments for an agent entering retirement with state  $S$ . The key parameter here is  $\frac{\beta-1}{\beta} < \phi \leq 1$  which captures in a reduced form way the length of retirement. A small  $\phi$  implies a small annuity value (the effect of  $\phi$  in the numerator in equation (18)) but

---

<sup>2</sup>To keep notation light in this section we omit  $t$  subscripts (indicating different calendar dates) and  $i$  subscripts (indicating different households).

a large total utility (the effect of  $\phi$  in the denominator in equation (18)), thus capturing the idea of a long retirement period, which induces agents to accumulate more wealth. Similarly a large  $\phi$  captures the idea of short retirement period, as the limit, with  $\phi = 1$ , retirement lasts exactly one period. In our exercise  $\phi$  is treated as a parameter and is set to match age-wealth accumulation profiles from the SCF data.

### 3.1.2 Working age

In each period households decide how much to work, consume, accumulate wealth in a risk free uncontingent bond and how much tax reduction to buy. Agents are allowed to purchase a tax buyout contract every period, and the purchase affects only tax rates in that period. The contract in its most general form is represented by a set  $\Delta$  of possible tax rate reductions and by a pricing function  $D(\delta_h, S_h)$  which maps quantity of tax reduction purchased  $\delta_h$  and household states into prices. We restrict our attention to pricing functions that are (possibly) dependent only on variables that can be contracted upon, such as age and past level of income. The problem of an agent of age  $h$  with wealth  $a_h$  and state  $S_h$  can be written in a recursive fashion as

$$\begin{aligned}
 V_h(a_h, S_h) &= \max_{\delta_h, a_{h+1}, l_h} u(c_h, l_h) + \beta EV_{h+1}(a_{h+1}, S_{h+1}), h = 1, N \\
 &\quad \text{s.t.} \\
 c_h + D(\delta_h, S_h) + a_{h+1} + T(\omega e_h(S_h)l_h, ra_h, \delta_h) &= (1 + r)a_h + \omega e_h(S_h)l_h \\
 \delta &\in \Delta, a' \geq 0
 \end{aligned}$$

where  $\omega$  is the wage rate per efficiency unit and where, for simplicity, we assume wealth is restricted to be non negative. The function  $T(\omega e_h(S)l_h, ra_h, \delta_h)$  represents the total taxes paid by an agent with labor income  $\omega e_h(S)l_h$ , capital income  $ra_h$  and tax buyout purchase  $\delta_h$ . Let the function  $\bar{T}(\omega e_h(S)l_h, ra_h)$  be the current tax code (without the buyout). Then the function  $T(\omega e_h(S)l_h, ra_h, \delta) = \max(\bar{T}(\omega e_h l_h, ra_h) - \omega l \delta, 0)$ . This implies, in practical terms, that when an agent who has purchased  $\delta$  units of the buyout files taxes, she simply computes the tax bill under the standard tax code and then subtracts from the tax bill a fraction  $\delta$  of labor earnings. The solution to this problem can be represented by age dependent decision rules for labor effort  $l_h(a_h, S_h)$ , consumption  $c_h(a_h, S_h)$ , next period assets  $a_{h+1}(a_h, S_h)$  and tax buyouts  $\delta_h(a_h, S_h)$ . An important consequence of this specification is that although agents face idiosyncratic uncertainty, the decision of whether or not to take

the contract is a period by period one, and it is taken after agents have observed all shocks, so purchasing the buyout does not involve risk; again in practical terms we suggest that agents may make the buyout decision at the time of tax filing. What is crucial for the effectiveness of the contract is for agents to be aware of the option at the time they decide labor supply, as the presence of the contract will, in general, affect returns from working.

### 3.2 Firm

Output is produced by a representative, competitive and profit maximizing firm which hires labor  $L$  at price  $\omega$  and capital  $K$  at price  $r$  and uses a standard constant returns to scale technology so that its profits are given by

$$K^\alpha L^{1-\alpha} - (\kappa + r)K - \omega L$$

where  $0 < \alpha < 1$  is the share of capital in the technology and  $\kappa$  is the rate of depreciation of capital used in production.

### 3.3 Government

In each period the government has to finance a constant flow of unproductive government consumption  $g$ , transfers to retired household and interest on debt  $rb$ ; its receipts are given by taxes, sales of tax buyouts and issue of new debt  $b'$ . Let  $F_h(a, S)$  be a distribution of households of age  $h$  over assets  $a$  and exogenous states  $S$ , then the budget constraint can be written as

$$\begin{aligned} & (1+r)b + g + \int_{S,a} T^R(S) dF_{N+1}(a, S) \\ & \leq \sum_{h=1}^N \int_{a,S} \left( T(\omega e_h(S) l_h(a, S), ar, \delta_h(a, S)) + D(\delta_h(a, S), S) \right) dF_h(a, S) + b' \end{aligned} \quad (19)$$

### 3.4 Equilibrium

A steady state equilibrium is the following set of objects:

- Decision rules of households:  $l_h(a, S), a_{h+1}(a, S), \delta_h(a, S)$  and of firms for  $K(r, \omega)$  and  $L(r, \omega)$

- Prices  $\omega, r$
- Government policies:  $T(\omega e(S)l, ra, \delta), T^R(S), \Delta, D(\delta, S), g, b$
- Distribution of agents over states  $F_h(a, S), h = 1, \dots, N + 1$

such that

- Given prices and government policies the decision rules solve the problem of the agents and of the firms
- Initial distribution of new born agents  $F_1(a, S)$  is given and time invariant. Additional distributions  $F_h(a, S), h \geq 2$  are induced by decision rules and by transition functions for exogenous states
- Labor market clears, i.e.,

$$L = \sum_{h=1}^N \int_{a,S} e_h(S) l_h(a, S) dF_h(a, S)$$

- Government budget balances, i.e., equation (19) holds

### 3.5 Calibration

The spirit of the calibration exercise is to have a steady state equilibrium of the model economy without the tax buyout to replicate some key cross-sectional statistics of a representative sample of US households in 2006. In particular we focus on households with at least one member aged between 25 and 60 and who report positive earnings (i.e., households who participate actively in the labor market) from the 2007 Current Population Survey (CPS) for earnings and tax statistics and on a comparable group of households from the 2007 Survey of Consumer Finance (SCF) for wealth data.<sup>3</sup> We decided to target earnings, tax and wealth statistics and abstract from reported statistics on labor input (i.e., hours worked) as we believe that reported hours worked are an imprecise measure of the true effort exerted by the households and hence not very informative on one key parameter for our study, i.e., the Frisch elasticity of labor supply (for more on this point see Feldstein, 1995). Instead we rely on a series of recent empirical studies to identify a plausible range of values for the Frisch elasticity and then, for each value of the labor elasticity, we set

---

<sup>3</sup>Both datasets refer to the calendar year 2006.

the remaining parameters (i.e., initial distributions, variance of wage shocks, distribution of discount factors) following related micro-studies or to match earnings and wealth data. We now describe in detail our procedure.

**Period length and life span.** A period in our model is a year and since we focus on households with the reference person between age 25 and 60 we set  $N = 36$ .

**Utility function.** Following many of the quantitative studies in the area (for example Conesa et al. 2008) we use the following functional form for period utility

$$U(c, l) = \log(c) - v \frac{l^{1+\gamma}}{1+\gamma}, v > 0, \gamma > 0$$

The crucial parameter for our purposes is  $\gamma$ , which maps one to one on the Frisch elasticity of labor supply (which is equal to  $\frac{1}{\gamma}$ ). Recent quantitative work by Kimball and Shapiro (2008) and by Erosa, Fuster and Kambourov (2009) who take into account variation in the extensive margin, intra-household adjustment and possibly non linear wage schedule suggests a value for  $\gamma$  of 1. Other econometric studies (see for example Blundell, Duncan and Meghir, 1998) favor a value closer to 1/2 so our range for the Frisch elasticity is  $[0.5, 1]$  with our benchmark value being 0.75. We set the parameter  $v$  so that all households which choose consumption exactly equal to their labor income work exactly 1/3 of their time.

**Technology.** Technology parameters are standard and we set capital share  $\alpha = 1/3$  and depreciation rate of capital  $\kappa = 0.1$

**Interest rates and wages.** The interest rate in this economy is exogenously fixed and we set it to 1%. There is abundant empirical research that estimates statistical process for log-wages using panel data in the US and in other countries (for a survey see Katz and Autor, 1999). Most studies model the log wage as a sum of four components: a permanent part (due to fixed characteristics such as ability, education and sex), a deterministic age dependent factor (due to un-modelled accumulation of human capital), a very persistent process (close to a random walk, capturing long run trends in wages due to permanent factors such as career changes), a transitory process (close to i.i.d. capturing temporary shocks such as unemployment). Our choice of the process for individual efficiency units of labor reflects these studies and in table 1 we summarize the full set of parameters characterizing the process.

**Table 1. Wage process**

Component	Distribution	Dynamics	Process	Parameters
$\varepsilon$	Uniform	i.i.d	$\varepsilon_{ih}$	$\sigma^2(\varepsilon_i) = 0.08$
$p$	Normal	AR(1)	$p_{ih} = 0.99p_{ih-1} + \eta_h$ , $p_{i0}$ given	$\sigma^2(\eta_i) = 0.02, \sigma^2(p_{i0}) = 0.1$
$A$	Uniform	Deterministic	$A_{ih} = A_i$ for all $h$	$\sigma^2(A_i) = 0.2$
$z$		Deterministic	$z_{ih} = \chi_1 h - \chi_2 h^2$ for all $i$	$\chi_1 = 0.02, \chi_2 = 0.0013$

Note.  $h$  subscripts indicate age and  $i$  subscripts indicate households. In the computed version of the model the processes for  $\varepsilon, p$  and  $A$  are approximated using 5 states Markov chains

Note that  $\sigma^2(\varepsilon_i)$ , and  $\sigma^2(\eta_i)$  are taken from a recent study (Heathcote, Perri and Violante, 2010) which estimates persistent and temporary components for log wages in the US over the period 1967-2002. The other parameters, such as the variance of the permanent differences  $\sigma^2(A_i)$ , the variance of initial draw of the persistent shock  $\sigma^2(p_{i0})$  and the ones characterizing age/wage profile ( $\chi_1$  and  $\chi_2$ ), are set so that earnings in our model match the age profile for mean and variance of household earnings (per adult equivalent) in the CPS data (see figure 3 below, panels a and b).<sup>4</sup> It is worth noting that the pattern of earnings dispersion over the life cycle (panel b) in the data displays two sizeable increases, the first early in life and the second just before retirement. The pattern of earnings dispersion in the model matches these two increases. The increase in dispersion early in life in the model is captured by picking the dispersion of the initial draw of the persistent shocks below its long run value so households age and the persistent shocks reach their long run distribution earnings dispersion increases. The final increase in dispersion instead is due to the fact that, as retirement nears, persistent shocks become more like temporary shocks. Thus the labor supply response to these shocks is more pronounced, increasing earnings dispersion.

**Initial wealth distribution, discount factor and retirement length.** A well known problem of models like the one considered in this paper is their inability of generating wealth dispersion comparable to the data. This shortcoming is important for our purposes because our specification of utility implies that wealth is an important determinant of labor supply, so if our model fails to match the wealth distribution it will also fail to match the distribution of labor effort and the willingness of agents to take the contract. For this reason we introduce two features to allow our model to match wealth dispersion in US data. The

<sup>4</sup>Adult equivalents are computed using the OECD equivalent scale.

first is that we start agents with a distribution of wealth that matches the coefficient of variation of the distribution of wealth of 25 year old households in the US SCF. Second we divide households in two groups, as for example in Krusell and Smith (1998), one with high discount factor and one with low discount factor and we choose the distribution of discount factors so that the model matches the coefficient of variation of wealth over the life cycle (see figure 3, panel d).<sup>5</sup> Finally we set the parameter  $\phi = 0.04$ , capturing the incentive to save for retirement so that our model matches the average wealth to earnings ratio for the economy (see figure 3, panel c).

### Government policies

**Tax system** A crucial element of our analysis is the current tax system, as this determines households' outside option whenever they are pondering whether to take the contract or not. Applied public finance literature (Gouveia and Strauss, 1994) has proposed the following flexible functional form to capture variation of tax rates over the income distribution

$$\frac{T}{Y} = \tau_0 - \tau_0(\tau_1 Y^{\tau_2} + 1)^{-\frac{1}{\tau_2}} \quad (20)$$

where  $Y$  is pretax income,  $T$  is the tax bill including federal, state and local income taxes plus social security and medicare contribution, and  $\tau_0, \tau_1$  and  $\tau_2$  are parameters determining the levels and the slope of tax rates. This function is flexible enough to capture flat, proportional, and progressive taxes and the parameter  $\tau_0$  has the natural interpretation of the asymptotic marginal tax rate. We estimate the parameters of this function using the sample of households from 2007 CPS data<sup>6</sup> and the results are reported in figure 4 below. Notice how, with the exception of the very low income, the function fits tax rates in 20 quantiles very well. Our estimate of the asymptotic marginal tax rate  $\tau_0$  is 32%, the estimate for the parameter  $\tau_2$ , capturing the progressivity of the tax system, is 2.4 and the (scale dependent) parameter  $\tau_1$ , capturing the overall fiscal pressure, is 11.9. In the CPS data the ratio of total tax bill to total income is about 21% and using the estimated tax function (20) in our model economy yields a comparable ratio.

**Social security** Retired agents receive social security payments  $T^R(S)$ . We specify

<sup>5</sup>In particular we set the discount factors of the 2 groups equal to 1.033 and 0.98 and set the size of the first group to be 1/4 of the size of the second.

<sup>6</sup>We use the same sample to compute earnings statistics to calibrate the wage process described above. The CPS reports state and federal taxes plus FICA due by each household in the sample using a tax simulator model.

Figure 3: EARNINGS AND WEALTH MOMENTS: DATA AND MODEL

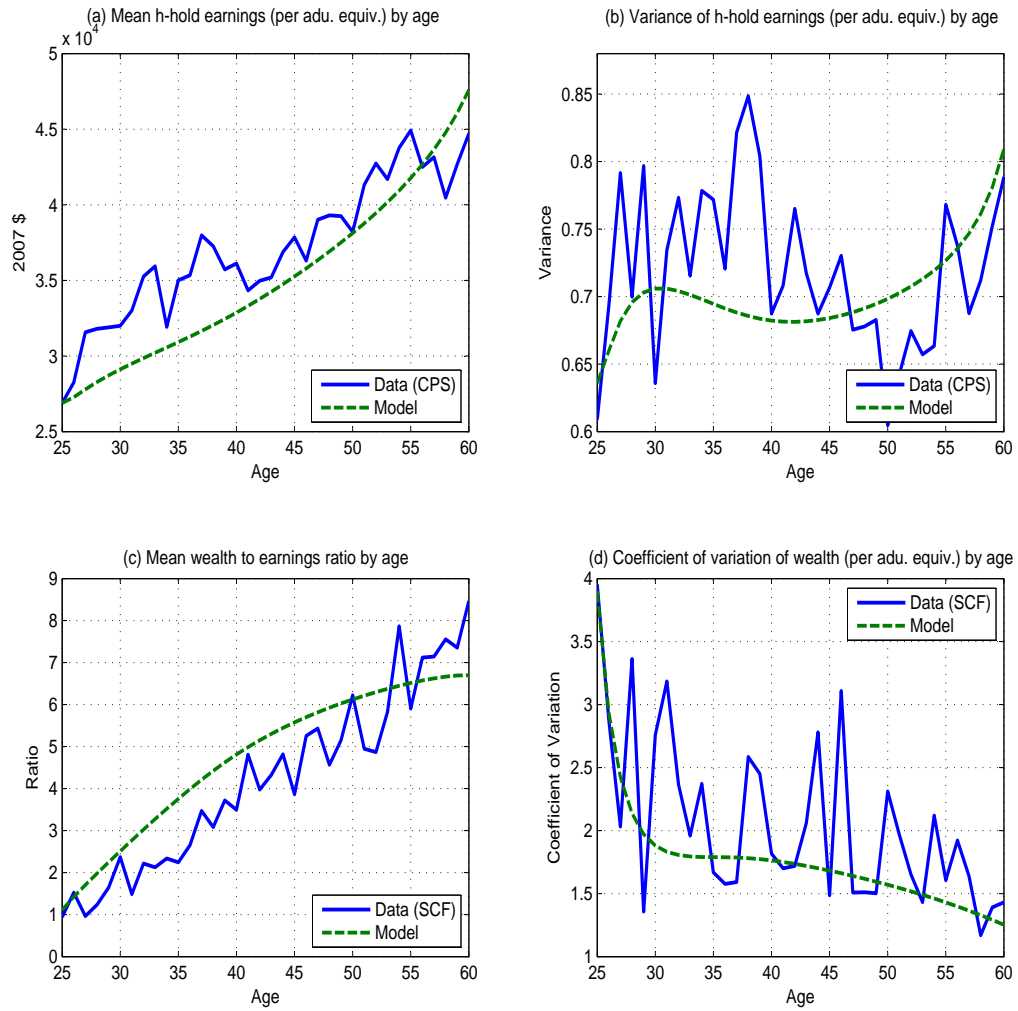
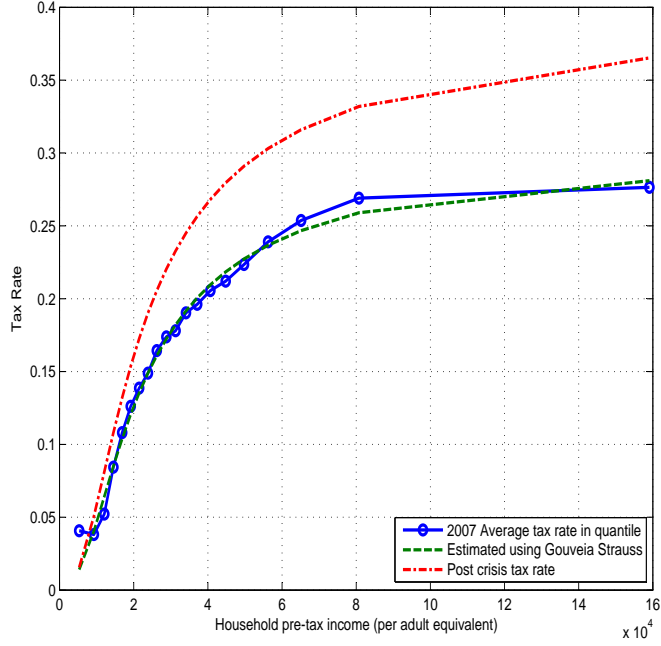


Figure 4: TAX FUNCTIONS



social security payments using a simplified version of the social security payment functions used in Imrohorglu and Kitao (2009), that captures the concavity of social security payments as a function of lifetime income. In particular we assume that social security payments are a function of  $A$ , i.e., the permanent component of the wage and use the following form for the social security payments

$$T^R(S) = \begin{cases} 0.9\bar{l}\omega e^A & \text{for } A \leq A_{20} \\ 0.8\bar{l}\omega e^{A_{20}} + 0.1\bar{l}\omega e^A & \text{for } A_{20} \leq A \leq A_{60} \\ 0.8\bar{l}\omega e^{A_{20}} + 0.1\bar{l}\omega e^{A_{60}} & A > A_{60} \end{cases}$$

where  $A_{20}$  and  $A_{60}$  are the 20<sup>th</sup> and 60<sup>th</sup> percentile of the distribution of  $A$  and  $\bar{l}$  is the average labor effort in the economy.

**Debt and government consumption.** In steady state we set government debt to 0 and so government consumption  $g$  is set residually from the budget constraint (19) as the difference between tax revenues and social security transfers.

## 4 The experiment

In our baseline experiment we start from a steady state equilibrium with low total government expenditures (including  $g$ , transfers and interests on debt) where no tax buyouts are offered. This is the economy we calibrated in the previous section using our sample of US households in 2006. We then assume that, unexpectedly, public expenditures increase a level which is 20% higher than in the initial steady state, capturing with this a structural and permanent change in the economy, such as an aging population resulting in higher total (but not per retired household) transfer expenditures or an increase in debt and interest payments due to the financial sector bailout.<sup>7</sup> The path for government spending is depicted in figure 5 panel (a). We then compute two possible equilibrium paths: one under which the buyout is offered and the other under which the buyout is not offered. We obviously need to make assumptions on how revenue shortfalls are financed and what type of buyout is offered. In our benchmark experiment we assume that in the no-buyout case taxes are increased so that the budget is balanced in the period in which the fiscal spending increases, and then the tax policy is left unchanged. The increase in the tax policy is the one depicted in figure 4, going from the policy represented by the dashed line to one represented by the dashed-dotted line. In particular we increase the parameter governing the asymptotic marginal tax rate,  $\tau_0$  and leave the other parameters unchanged so that additional expenditures are financed leaving the progressivity structure unchanged, which implies that high income households pay a higher fraction of those. In the buyout case we use the same path for fiscal policy of the non buyout case, with the only difference being the offering of the buyout, starting from the date of the fiscal expansion. The particular buyout we consider is a maximum 5% reduction in tax rate offered to everybody at a price of roughly 11.6% of average pretax income per adult equivalent per household (in our CPS 2006 sample mean pretax income per household per capita is about \$39,000 so the cost of the buyout would be about \$4,500). This is the minimum price that guarantees that the government budget is always positive after the introduction of the buyout.

**Aggregate consequences: baseline case.** The aggregate consequences are described

---

<sup>7</sup>In our set-up the composition of increased spending (higher total transfers, higher  $g$  or higher interest payments) does not matter.

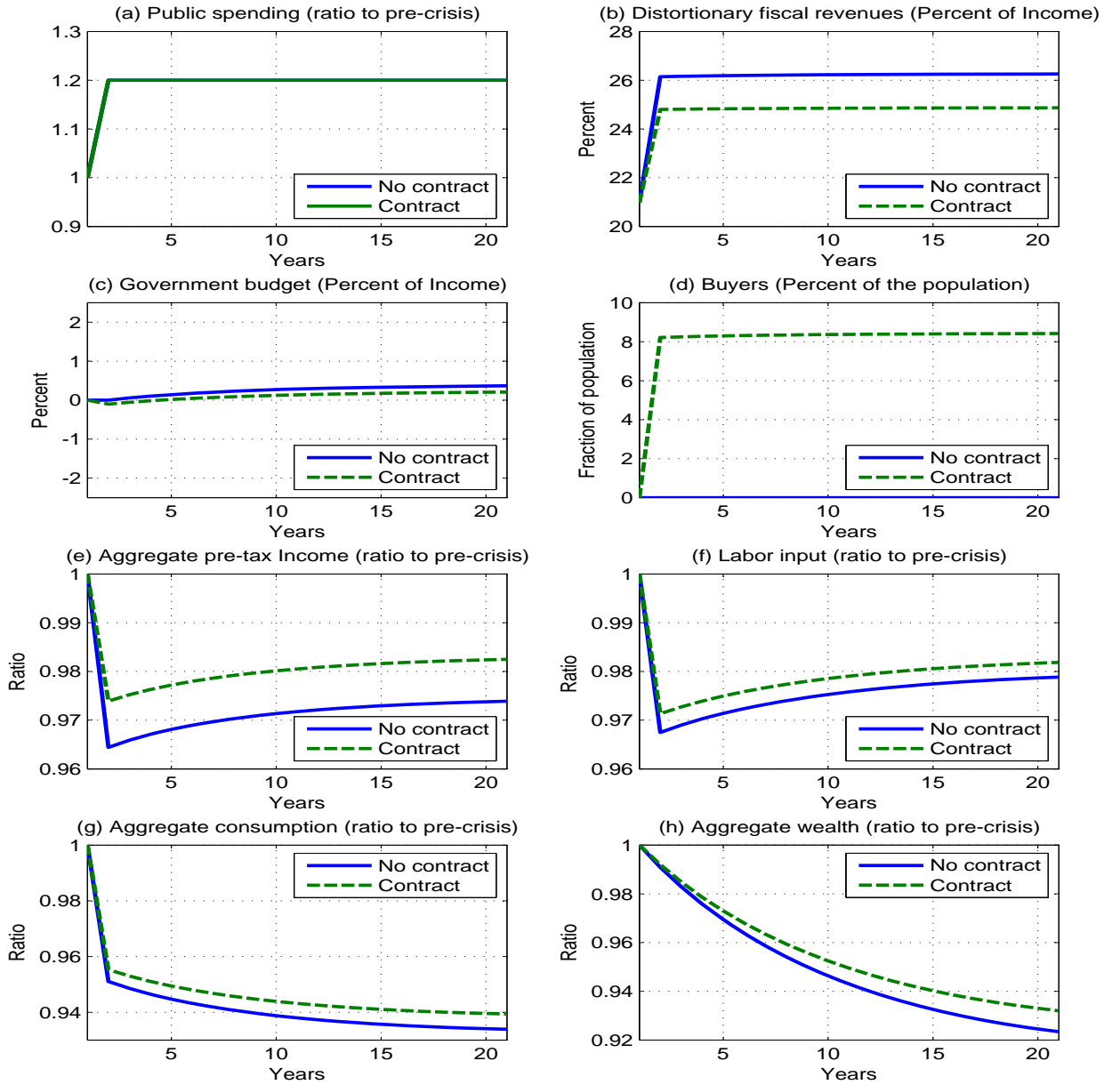
in figure 5. Panel (b) shows the total amount of distortionary taxes,

$$\sum_{h=1}^N \int_{a,S} (T(we(S)l_h(a,S), ra, \delta_h(a,S))) dF_h(a,S)$$

as a fraction of total income that is collected under the two regimes. Notice that under no contract this measure of fiscal pressure is about 26% of income while with the buyout is reduced to around 24.5% of income. Yet panel (c) shows that total government budget is about the same in the two regimes. Since spending is the same in the two regimes, what is making up for the loss of distortionary revenues in the buyout case? The answer is in panel (d) which shows that a significant fraction of the population (over 8%) is purchasing the buyout, thus creating a source of non distortionary revenue that makes up for the loss of distortionary revenue arising from the offering of the contract. This feature, i.e., the transformation of part of government revenues from distortionary into lump sum, done without making anybody worse off and in a revenue neutral fashion, is the essence of the tax buyout contract. Panels (e) through (h) show that the reduction in distortion has actually quantitative important effects. Panel (f) shows that labor supply in the world with buyout is about 1/3% higher than in the world without buyout, precisely because households who elect to purchase the buyout choose to work harder. Notice that because households which purchase the buyout and work harder are agents with high efficiency units (see the discussion below), the buyout increases average labor productivity, which in turn results in a fall in output (panel e) which is over 1% lower than the fall experienced in the world without the buyout. Panels (g) and (h) finally show that as a result of the tax increase, households suffer a negative wealth effect which induces them to choose lower consumption and bring less wealth into retirement. Under the buyout, due to the lower distortions, the wealth effect is less severe and both the reduction in consumption and in wealth are more muted.

**Aggregate consequences: alternative scenarios.** In this subsection we evaluate how the aggregate impact of a tax buyout varies when we change some of the assumptions in the baseline case analyzed above. In particular we evaluate the option of the government of offering smaller or larger buyouts and of using a different pricing scheme. Also we assess the importance of a key parameter: the Frisch elasticity of labor supply. For each experiment we set the price of the buyout to the minimum level that guarantees non negative government budget in every period. In table 2 we report five summary statistics for each experiment:

Figure 5: AGGREGATE CONSEQUENCES OF BUYOUTS



the maximum buyout agents are allowed to sign, the total price of the max buyout (as a percentage of mean income), the percentage of households who buy (in all experiments here, as in the simple model of section 2, either a household does not buy or buys the maximum possible amount), the percentage of government revenues coming from the sales of the contract (hence non distortionary) and the gain in GDP relative to the situation in which the contract is not offered.<sup>8</sup>

Lines 2 and 3 show how increasing the maximum size of the buyouts leads to an increase in price, i.e., larger buyouts are more expensive, even on a per-unit basis. As a consequence there is a smaller number of buyers, but since those buyers buy a larger and more expensive tax reduction, the resulting government revenues from the sale of the buyout increase. Since the buyers have high wages and those are the ones who increase their labor supply even a large buyout has a sizeable impact on GDP. Similarly a smaller buyout is cheaper, it increases the number of buyers, but is less effective in generating revenues and reducing distortions.

Line 4 shows that offering a buyout whose price is increasing in the age of the households can attract more buyers and generate more revenues. Older households, because of the deterministic component of wage that grows with age, benefit more from the contract and are more willing to pay a higher price.

Finally lines 5 and 6 show that changing the Frisch elasticity of labor supply has a limited effect on the outcomes of introducing the buyout. The reason for this result is that when we change the Frisch elasticity we also recalibrate the variance of the wage shocks to match the variance of earnings in the data. When, for example, we consider a higher value for the Frisch elasticity, we increase the effectiveness of the buyout as reduction in distortions are more valuable. But at the same time a higher elasticity requires a lower variance of the wage shocks, including the non permanent ones. The lower variance reduces the extent through which agents adjust labor supply to take advantage of temporary changes in productivity, and hence reduces the effectiveness of the buyout. These two offsetting effects operate when we change elasticity in both directions and explain the relative robustness of the effects with respect to changes in the Frisch elasticity.<sup>9</sup>

---

<sup>8</sup>These last 3 statistics are averaged over the first 5 years in which the buyout is offered.

<sup>9</sup>To verify this claim we experimented with changing the elasticity and not recalibrating the model. In that case we found that changes in the elasticity have a substantially larger impact on the effectiveness of a buyout.

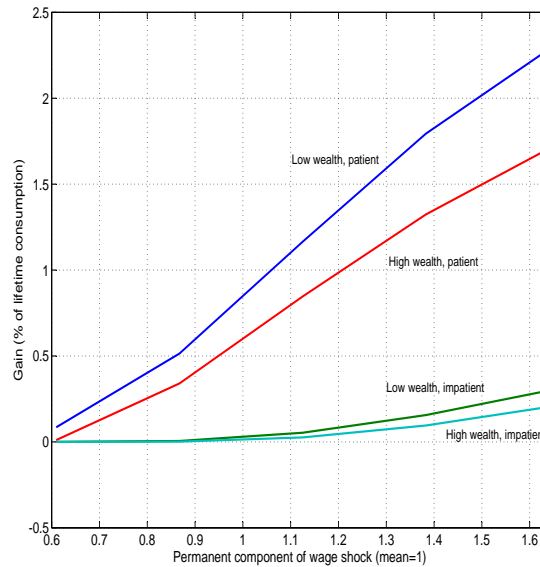
Table 2. Alternative scenarios

	Buyout size	Price (% of mean income)	Buyers	Revenues	GDP Gain
1. Baseline	5%	11.6%	8.2%	4%	0.95%
(a) Varying the size of the buyout					
2. Large buyout	10%	33%	4%	5.6%	0.8%
3. Small buyout	2%	3.3%	14.5%	2%	0.6%
(b) Alternative pricing schemes					
4. Age dependent	5%	$5.5\% + 0.25\%h$	10.1%	4.2%	1.1%
(c) Alternative Frisch Elasticity*					
5. Low (0.5)	5%	12.8%	6%	3.3%	0.55%
6. High (1.0)	5%	12.1%	8%	4%	0.9%

\*In experiments 5 and 6 the variance of wage shocks is recalibrated so the variance of earnings in each case matches the variance of earnings in the data

**Micro Consequences.** An advantage of explicitly considering a dynamic economy is that we can identify households who have the strongest incentive to buy the contract today but also those who will benefit from its introduction as they might take the contract in the future. One can interpret a tax buyout as the possibility for households to buy at a fixed price a percentage subsidy on labor income; obviously households who benefit from it the most are the ones who have (or expect to have) high labor income. This implies that, *ceteris paribus*, high wage (and hence older) households, more patient (because they value more the future chance of getting into a state of high labor income) households, and low wealth households (because lower wealth induces higher labor supply) are more likely to purchase the buyout and hence will benefit more from its introduction. In figure 6 we quantify this benefit by computing the lifetime consumption equivalent that different types of newborn agents are willing to pay to go from an economy without a contract to an economy with a contract. On the  $x$  axis the figure reports the value of permanent wage component  $A$  (as a measure of wage shock) of a given household and on the  $y$  axis lifetime the gain (in terms of lifetime consumption) from the permanent introduction of the contract in the economy. The lines labeled “low wealth” refer to households with 0 wealth (which is the majority of newborn households), while “high wealth” denotes a median wealth level. The labels “patient” and “impatient” refer to households with high and low discount factors (see

Figure 6: BENEFITS FROM THE INTRODUCTION OF THE CONTRACT



section 3.5 above).<sup>10</sup> As discussed before the gains are increasing in wages, decreasing in wealth and increasing in the degree of patience. Perhaps the most interesting feature of the figure is that although none of the households depicted by the lines in the figure is actually purchasing the contract in the current period, most of them enjoy a non negligible benefit from its introduction, as they value the possibility of future use of the contract. This makes the point that in a dynamic world the benefits of the buyout, if evaluated over the long run, spread well beyond the 10% of the population which elect to take the contract at a moment in time.

## 5 Conclusions

A tax buy-out is a simple contract between the government and private agents which can reduce distortions in the economy without adversely affecting government finances. Our work indicates that its effects are quantitatively relevant even in the very basic form we

<sup>10</sup>All agents we considered in the figure have the median level of the temporary and of the persistent wage shock.

have analyzed. This suggests that a tax buyout scheme might actually be useful to reduce distortions in a period in which, due to large fiscal deficits, taxes are bounded to rise.

To implement the scheme in practice more analysis will be needed. There are two main concerns that should be addressed. The first is the presence of general equilibrium effects where the increase in labor supply of the buyers might reduce equilibrium wages, thus hurting the low ability individuals who do not buy the contract. The second is the possibility of a fat (Pareto) tail in the distribution of income/ability. This feature makes the reduction in marginal taxes for high ability individuals, and hence the buyout, socially less desirable (Diamond, 1998; Saez, 2001). A fat right tail of the ability distribution is also concerning as it may imply non existence of a price that guarantees revenue neutrality for the simple buyout considered in this paper. Future research could investigate the possibility of limited buyouts, in which individuals' gains from the tax reduction cannot exceed a multiple of the price paid. These limited buyouts would be less effective in reducing distortions, but at the same time would limit government losses on high ability individuals, enhancing the feasibility of the contract.

On the positive side there are many dimensions along which the contract can be extended to further reduce distortions for a larger fraction of the population. First, as shown in a simplified setting, non linear pricing schemes can extract resources from high ability in favor of low ability individuals. Second, we have performed our quantitative analysis in an informationally constrained setting, where the price of the contract is either fixed or dependent on age. Of course, in reality the government observes many additional individual characteristics, such as education and past labor income, and could condition the price of the contract on them. In this case, however, one would need to consider the dynamic implications of the contract on individual behavior, as they might reintroduce some of the distortions that the contract is supposed to eliminate. For example, conditioning the price on past income will break the lump-sum feature of the tax buyout, as an individual takes into account that labor income today affects the price of the contract tomorrow. Third, in the paper we have assumed the same elasticity of labor supply across individuals but there is evidence of heterogeneity (Aaberge et al. 2002). Implementing the contract in an environment with this feature might lead to self (benign) selection of high elasticity individuals into the contract, allowing lower prices and higher acceptance. Finally, we have excluded from the analysis capital taxation. It will be interesting to analyze simultaneously

the reform of capital and income tax, considering to what extent the buyout can be applied to capital taxation and what types of complementarities can arise between the two schemes.

## References

- [1] Aaberge Rolf, Ugo Colombino and Tom Wennemo, 2002, Heterogeneity in the elasticity of labor supply in Italy and some policy implications, CHILD working paper, 20
- [2] Alesina, Alberto and Philippe Weil, 1992, Menu of linear income tax schedules, NBER Working paper 3968
- [3] Blundell, Richard, Alan Duncan and Costas Meghir, 1998, Estimating Labor Supply Responses Using Tax Reforms, *Econometrica* 66(4), pp. 827-861.
- [4] Conesa, Juan Carlos, Sagiri, Kitao and Dirk Krueger, 2008, Taxing Capital? Not a Bad Idea After All!, *American Economic Review*, 99 (1), pp. 25-48.
- [5] Diamond, Peter A. 1998. Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates, *American Economic Review*, 88 (1), pp. 83-95.
- [6] Erosa, Andres, Luisa Fuster and Gueorgui Kambourov, 2009, The Heterogeneity and Dynamics of Individual Labor Supply over the Life Cycle: Facts and Theory, manuscript, University of Toronto
- [7] Feldstein Martin, 1995, Behavioral Responses to Tax Rates: Evidence from TRA86, *American Economic Review* 85(2), 170-174
- [8] Fukushima, Kenichi, 2010, Quantifying the Welfare Gains From Flexible Dynamic Income Tax Systems, manuscript, University of Minnesota
- [9] Gouveia, Miguel, and Robert P. Strauss. 1994. Effective Federal Individual Income Tax Functions: An Exploratory Empirical Analysis. *National Tax Journal*, 47(2), pp 317-339.
- [10] Heathcote, Jonathan, Fabrizio Perri and Gianluca Violante, 2010, Unequal We Stand: An Empirical Analysis of Economic Inequality in the United States, 1967-2006, *Review of Economic Dynamics*, 13(1), pp. 15-51
- [11] Heathcote, Jonathan, Kjetil Storesletten, and Gianluca Violante, 2008, The Macroeconomic Implications of Rising Wage Inequality in the United States, NBER Working paper 14052

- [12] Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L. Violante. 2010. Redistributive Taxation in a Partial-Insurance Economy, manuscript, Federal Reserve Bank of Minneapolis
- [13] Imrohorglu Selahattin and Sagiri Kitao, 2009, Labor supply elasticity and social security reform, *Journal of Public Economics*, 93(7-8), pp. 867-878
- [14] Katz, Lawrence F., and David H. Autor. 1999. Changes in the Wage Structure and Earnings Inequality. In Handbook of Labor Economics, vol. 3A, edited by Orley Ashenfelter and David Card. Amsterdam: North Holland.
- [15] Kimball, Miles and Matthew Shapiro, 2008, Labor Supply: Are the Income and Substitution Effects Both Large or Both Small?, NBER Working paper 14208
- [16] Krusell Per and Anthony A. Smith, 1998. Income and Wealth Heterogeneity in the Macroeconomy, *Journal of Political Economy*, 106(5), pp. 867-896
- [17] Mankiw, N. Gregory and Matthew Weinzierl. 2006. Dynamic Scoring. A Back-Of-The-Envelope Guide, 90, Issues 8-9, September 2006, Pages 1415-1433
- [18] Mas-Colell, A. and Whinston, M.D. and Green, J.R. 1995 *Microeconomic theory*, Oxford University Press, Oxford.
- [19] Mirrlees, James A. 1971. An Exploration in the Theory of Optimum Income Taxation. *Review of Economic Studies* 38(2), pp. 175-208.
- [20] Prescott Edward C. 2004. Why do Americans work so much more than Europeans? *Federal Reserve Bank of Minneapolis Quarterly Review*, July 28 (1), pp. 2-13.
- [21] Saez, Emmanuel, 2001, Using Elasticities to Derive Optimal Income Tax Rates, *Review of Economic Studies* 68, 205-229.
- [22] Salanié, Bernard. 2003. The Economics of Taxation. MIT Press, Cambridge.
- [23] Slemrod, Joel Shlomo Yitzhaki, Joram Mayshar, Michael Lundhoms, 1994, The optimal two-bracket linear income tax, *Journal of Public Economics* 53, 269-290
- [24] Stiglitz, Joseph E., 1982 "Self-selection and Pareto Efficient Taxation." *Journal of Public Economics* 17, 213-240

## A Appendix

### A.1 Proofs

*Proof of Proposition 1:* Define the budget constraint in terms of consumption and leisure:

$$(1 - \tau)(1 - l(A, w, \tau)) + c(A, w, \tau) = w + 1 - \tau$$

Let's start from the case in which the government prices in such a way that the contract is self-financing at the pre-contract equilibrium labor supply:

$$D = \delta l_A(0, 0)$$

Then, the contract is a compensated price change. In fact, the “endowment” (that is, total income flow at leisure 0) is now:

$$w + (1 - \tau + \delta) - \delta l_A(0, 0)$$

which is exactly the expenditure we obtain for the old allocation at the new prices:

$$(1 - \tau + \delta)(1 - l_A(0, 0)) + c_A(0, 0) = w + (1 - \tau + \delta) - \delta l_A(0, 0)$$

where we have used  $c_A(0, 0) = -(1 - \tau)(1 - l_A(0, 0)) + w + 1 - \tau$ . Then, we only need to assume that preferences over leisure and consumption satisfy the weak axiom of revealed preferences to conclude that labor supply with the contract increases (see prop. 2.F.1 in Mas-Colell et al. 1995). Moreover, given that by construction the old allocation is still in the budget set, the agent is strictly better off.

Now we need to show that the argument carries through when we use our pricing scheme. For given  $\delta$ , define  $\Delta l(D) \equiv l(\delta, D) - l_A(0, 0)$ , where  $D$  is the price of the contract and  $l(\delta, D)$  is labor supply with the contract priced  $D$ . Define the equation

$$D = \delta l_A(0, 0) - (\tau - \delta)\Delta l(D) \equiv g(D). \tag{21}$$

We need to show that there exists a solution  $D^* < \delta l_A(0, 0)$  at which  $\Delta l(D^*) > 0$ . This is our price. We have shown before that  $g(\delta l_A(0, 0)) < \delta l_A(0, 0)$ , because at that value

$\Delta l(\delta l_A(0,0)) > 0$ . Now, under the assumption that  $\Delta l(D)$  is increasing (leisure is normal good) and continuous, by decreasing  $D$  starting from  $\delta l_0$ , the right-hand-side of equation (21) increases, so that the equation must have a solution at a  $D^* < \delta l_0$ . At this price,  $\Delta l > 0$ . Assume not: if  $\Delta l = 0$ , then  $D = \delta l_0$ , at which  $\Delta l > 0$ , which leads to a contradiction. The same reasoning applies for  $\Delta l < 0$ : in that case,  $D > \delta l_0$  and the wealth effect implies that  $\Delta l > 0$ .

*Proof of Lemma 1:* From the first order conditions with respect to  $\delta$  at a saddle point  $\delta^*$  we have that:

$$D = Al_A(\delta^*, d\delta^*). \quad (22)$$

Multiply both sides by  $\delta^*$ , substitute (22) into the agent's budget constraint (9) and obtain:

$$\begin{aligned} c_A(\delta^*, d\delta^*) &= (1 - \tau + \delta^*)Al_A(\delta^*, d\delta^*) + w - d\delta^* \\ &= (1 - \tau)Al_A(\delta^*, d\delta^*) + w. \end{aligned} \quad (23)$$

This shows that  $(c_A(\delta^*, d\delta^*), l_A(\delta^*, d\delta^*))$  is a feasible allocation even in absence of the contract. The fact that it is not chosen – the agent chooses instead to consume and work  $c_A(0,0)$  and  $l_A(0,0)$ , respectively – indicates that the utility at any saddle point is lower than that when the contract is not taken. Hence, any saddle point must be a minimum. Therefore there are only corner solutions to the agent's problem.

*Proof of Lemma 2:* We want to show that there exists a per-unit price  $\underline{d}_A$  such that: i) agent  $A$  is willing to enter the contract, and ii) the government is neither losing nor gaining resources from the agent. Let  $\underline{d}_A$  be such that (8) holds – that is, the second condition is met – when the agent purchases the maximum allowed amount of the contract  $\bar{\delta}$ :

$$\underline{d}_A \bar{\delta} = \tau Al_A(0,0) - (\tau - \bar{\delta})Al_A(\bar{\delta}, d\bar{\delta}).$$

From agent  $A$ 's perspective, this is the same pricing as under perfect information. Therefore Proposition 1 applies: at that price the agent is willing to purchase the contract. Note also that at  $\delta = \bar{\delta}$  the marginal utility of the agent is still increasing in  $\delta$ :

$$Al_A(\bar{\delta}, d\bar{\delta}) - r\underline{d}_A = A \frac{\tau}{\bar{\delta}} (l_A(\bar{\delta}, d\bar{\delta}) - l_A(0,0)),$$

and from Proposition (1) we have that  $l_A(\bar{\delta}, d\bar{\delta}) - l_A(0,0) > 0$ .

Next, we want to show that there exists a per-unit price  $\bar{d}_A$  for which agent  $A$  is indifferent between taking and not taking the contract. Define  $\Delta u_A$  as the difference between the utility from taking the contract (and therefore, from Lemma 1, purchasing the maximum amount  $\delta = \bar{\delta}$ ) and not taking the contract ( $\delta = 0$ ):

$$\Delta u_A = u(c_A(\bar{\delta}, d\bar{\delta}), l_A(\bar{\delta}, d\bar{\delta})) - u(c_A(0, 0), l_A(0, 0)).$$

We are interested in the mapping between the per-unit price  $d$  and  $\Delta u_A$ . Proposition 1 implies that for  $d = \underline{d}_A$  the mapping is strictly positive. As long as  $c_A(\cdot)$  and  $l_A(\cdot)$  are continuous in their arguments, the mapping is also continuous. Let us increase  $d$  to the level  $\bar{d}_A^u$  such that we have a saddle point at  $\delta = \bar{\delta}$ :

$$\bar{d}_A^u - Al_A(\bar{\delta}, d\bar{\delta}) = 0. \quad (24)$$

Since this is a saddle point, we can show that  $\Delta u_A$  is negative at  $d = \bar{d}_A^u$  appealing to Lemma 1:  $(c_A(\bar{\delta}, d\bar{\delta}), l_A(\bar{\delta}, d\bar{\delta}))$  is a feasible allocation even in absence of the contract. The fact that it is not chosen shows that  $\Delta u_A$  is negative at  $d = \bar{d}_A^u$ . Therefore by continuity there exists a  $\bar{d}_A \in (\underline{d}_A, \bar{d}_A^u)$  such that  $\Delta u_A$  is zero and the agent is indifferent. The government's net revenues from agent  $A$ , equal to

$$d\bar{\delta} - \bar{\delta}Al_A(\bar{\delta}, d\bar{\delta})$$

are strictly increasing in  $d$  as long as leisure is a normal good. Hence the government is gaining resources from agent  $A$  for  $d > \underline{d}_A$ .

*Proof of Lemma 3:* We want to show that if agent with ability  $A(d)$  is indifferent between buying and not buying the contract under price  $d\bar{\delta}$ , all agents with ability  $A \geq A(d)$  will choose to purchase the contract. Call  $y_A(\bar{\delta}, d\bar{\delta}) = Al_A(\bar{\delta}, d\bar{\delta})$  the pre-tax income associated with an agent of ability  $A$ . Figure 7 plots the optimal choices with and without contract, namely,  $(c_1, y_1) \equiv (c_{A(d)}(\bar{\delta}, d\bar{\delta}), y_{A(d)}(\bar{\delta}, d\bar{\delta}))$  and  $(c_0, y_0) \equiv (C_{A(d)}(0, 0), Y_{A(d)}(0, 0))$  respectively. By definition,  $(c_1, y_1)$  and  $(c_0, y_0)$  lie on the same indifference curve for agent  $A(d)$  (solid curve), which we denote by:

$$u(c, y, A(d)) = u(c_0, y_0, A(d))$$

If utility is continuously differentiable and  $\bar{\delta}$  is strictly greater than zero, then the two allocations  $(c_1, y_1)$  and  $(c_0, y_0)$  are apart from one another, since the tangency points are  $1 - \tau$  and  $1 - \tau + \delta$ , respectively. Moreover, since  $1 - \tau < 1 - \tau + \delta$  and the indifference curve is convex, we have that  $c_0 < c_1$  and  $y_0 < y_1$ .

Let us consider a marginal increase in the ability level from  $A(d)$  to  $\tilde{A} = A(d) + \varepsilon$ . The single crossing condition postulates that  $\frac{-u_Y(\cdot)}{u_C(\cdot)}$  is decreasing in ability. Graphically, that means that the indifference curve for the agent with ability  $A(d) + \partial A$  is flatter at both  $(c_1, y_1)$  and  $(c_0, y_0)$  than the indifference curve for  $A(d)$  (dashed lines in Figure 7). That means that the indifference point passing through the allocation with the contract  $(c_1, y_1)$  is associated with higher utility than the one passing through the allocation without contract,  $(c_0, y_0)$ . If this were not the case, that is, if the indifference curve associated with  $(c_0, y_0)$  were to pass through (or above)  $(c_1, y_1)$ , then it would be crossing the solid indifference curve again, which contradicts single crossing. Hence the higher ability agent will buy the contract.

Formally, let  $(\tilde{c}_1, \tilde{y}_1)$  and  $(\tilde{c}_0, \tilde{y}_0)$  be the optimal allocations for an agent of ability  $\tilde{A}$  with and without the contract, respectively. Let  $\tilde{u}_1$  and  $\tilde{u}_0$  denote  $u(\tilde{c}_1, \tilde{y}_1, \tilde{A})$  and  $u(\tilde{c}_0, \tilde{y}_0, \tilde{A})$ , respectively. Let

$$u(c, y, \tilde{A}) = \tilde{u}_0$$

be the indifference curve going through  $(\tilde{c}_0, \tilde{y}_0)$ . By the single-crossing condition (Salanié 2003) the indifference curve for agent  $\tilde{A}$  going through  $(c_0, y_0)$  has a slope which is less than  $(1 - \tau)$ . Hence the optimal allocation  $(\tilde{c}_0, \tilde{y}_0)$  is to the right of  $(c_0, y_0)$ , that is,  $\tilde{c}_0 > c_0$  and  $\tilde{y}_0 > y_0$ . Therefore the indifference curve  $u(c, y, \tilde{A}) = \tilde{u}_0$  must be such that  $u(c_0, y_0, \tilde{A}) < \tilde{u}_0$ . Moreover, this indifference curve must cross the indifference curve for the  $A(d)$  ability individual,  $u(c, y, A(d)) = u(c_0, y_0, A(d))$ , at some point  $(\hat{c}_0, \hat{y}_0)$  to the right of  $(c_0, y_0)$ , that is, such that  $\hat{c}_0 > c_0$  and  $\hat{y}_0 > y_0$ . By the same argument, the optimal allocation  $(\tilde{c}_1, \tilde{y}_1)$  is to the right of  $(c_1, y_1)$ , that is,  $\tilde{c}_1 > c_1$  and  $\tilde{y}_1 > y_1$ . Therefore the indifference curve  $u(c, y, \tilde{A}) = \tilde{u}_1$  must be such that  $u(c_1, y_1, \tilde{A}) < \tilde{u}_1$ . Moreover, this indifference curve must cross the indifference curve for the  $A(d)$  ability individual,  $u(c, y, A(d)) = u(c_1, y_1, A(d))$ , at some point  $(\hat{c}_1, \hat{y}_1)$  to the right of  $(c_1, y_1)$ , that is, such that  $\hat{c}_1 > c_1$  and  $\hat{y}_1 > y_1$ . Note that the allocations  $(\hat{c}_0, \hat{y}_0)$  and  $(\hat{c}_1, \hat{y}_1)$  are distinct, since for  $\varepsilon$  small enough we can make them arbitrarily close to  $(c_0, y_0)$  and  $(c_1, y_1)$ , respectively. Moreover  $\hat{c}_0 < \hat{c}_1$  and  $\hat{y}_0 < \hat{y}_1$ .

Hence, we have found two points,  $(\hat{c}_0, \hat{y}_0)$  and  $(\hat{c}_1, \hat{y}_1)$ , at which the indifference curves for the higher ability agent,  $u(c, y, \tilde{A}) = \tilde{u}_0$  and  $u(c, y, \tilde{A}) = \tilde{u}_1$  respectively, cross the indifference curve for the indifferent agent  $u(c, y, A(d)) = u(c_0, y_0, A(d))$ . It cannot therefore be that  $\tilde{u}_0 = \tilde{u}_1$ , since that would imply that the same indifference curve for the higher ability agent crosses the indifference curve for the lower ability agent twice, violating the single crossing condition. If we had  $\tilde{u}_0 > \tilde{u}_1$ , then the indifference curve  $u(c, y, \tilde{A}) = \tilde{u}_0$  would lie above the indifference curve  $u(c, y, \tilde{A}) = \tilde{u}_1$ . But this cannot be, since it would imply that the indifference curve  $u(c, y, \tilde{A}) = \tilde{u}_0$  crosses the curve  $u(c, y, A(d)) = u(c_0, y_0, A(d))$  to the right of  $(\hat{c}_1, \hat{y}_1)$ . This contradicts the fact that  $\hat{c}_0 < \hat{c}_1$ . Hence, there exists an  $\varepsilon > 0$  such that the agents with ability  $\tilde{A} = A(d) + \varepsilon$  strictly prefer to purchase the contract. While the argument given here applies to abilities in the right neighborhood of  $A(d)$  only, we can repeat it to show that for all  $A > A(d)$  agents will choose to purchase the contract.

## A.2 The algorithm used to compute the discretized set of contracts

The algorithm works as follows. Imagine a discretization  $\{A_i\}_{i=1}^N$  of the interval  $[\underline{A}, \bar{A}]$  with  $A_0 = \underline{A}$  and  $A_N = \bar{A}$ . Fix the maximum amount  $\bar{\delta}$  that the government is allowed to offer. From the optimal taxation literature we know that in general the highest ability individual  $A_N$  should have the lowest marginal tax rate, hence it is natural that the maximum tax reduction should be targeted toward agent  $N$ , hence  $\delta_N = \bar{\delta}$ . Fix the price  $D_N$  associated with the maximum tax reduction, and assume that  $D_N$  is such that agent  $N$  is willing to buy the contract (from the analysis in Section 2 we know that if agent  $N$  faces a pre-contract marginal tax rate of  $\tau$ , for  $\delta_N \leq \tau$  she will be willing to buy the contract for any  $D_N \leq \delta A_N l_{A_N}(0)$  where  $l_{A_N}(0)$  is the pre-contract labor supply).

The problem of the government is to choose the remaining contracts  $\{\delta_i, D_i\}_{i=1}^{N-1}$  so that households  $\{A_i\}_{i=1}^N$  will self-select into purchasing only one contract. We tackle the problem in a pairwise fashion: Given the contract tailored for agent  $i + 1$ ,  $\{\delta_{i+1}, D_{i+1}\}$ , we choose  $\{\delta_i, D_i\}$  so that: (a) agent  $i + 1$  is indifferent between contract  $i + 1$  and  $i$ , and (b) agent  $i$  prefers contract  $i$  to  $i + 1$ . We construct  $\{\delta_i, D_i\}$  as follows: We know that, since  $A_{i+1} > A_i$ , in order to have separation we need that  $\delta_{i+1} > \delta_i$  (in this discussion we maintain the assumption that pre-contract marginal tax rates are the same across agents). We therefore

start with a candidate  $\delta_i = \delta_{i+1} - x^\delta$ , with  $x^\delta$  a very small number. If  $D_i = D_{i+1}$ , agent  $i$  will prefer contract  $i + 1$ . Hence we lower  $D_i$  by multiples of  $x^\delta$  – another very small number – until agent  $i$  prefers contract  $i$  over contract  $i + 1$ . For  $D_i < D_{i+1}$  agent  $i + 1$  may also prefer contract  $i$  over contract  $i + 1$ , violating (a). We therefore further lower  $\delta_i$  to  $\delta_i = \delta_{i+1} - 2x^\delta$ , and, if needed, lower  $D_i$  to make it appealing to agent  $i$ . We repeat the two steps until the gap between  $\delta_{i+1} - \delta_i$  makes agent  $i + 1$  indifferent between the two contracts (condition a) while the difference  $D_{i+1} - D_i$  makes condition (b) hold. Once we have chosen the contract for agent  $i$ , we repeat the same steps for agent  $i - 1$ , and so on.

By construction the self-selection works pairwise: agent  $i$  prefers her contract to  $i + 1$  or  $i - 1$ . It turns out that it works for all  $(i, k)$ : agent  $i$  prefers her contract to any other contract  $k$  (we check this condition numerically). The intuition for this result is that indifference curves for the agents in the consumption/income space are steeper for the less able agents (see Stiglitz 1982). Therefore if agent  $i + 1$  is indifferent between her contract and the  $i$  contract, while agent  $i$  is indifferent between her contract and the  $i - 1$  contract, then agent  $i + 1$  will strictly prefer her contract to  $i - 1$ , and vice versa. In fact, in our numerical example we find that the disutility from taking contract  $k$  for agent  $i$  increases with the distance  $|A_k - A_i|$ .

As we go down the ability ladder, at some point an additional constraint comes into consideration: agent  $i$  must prefer contract  $i$  to not taking the contract at all. If this is not the case, we stop: we have identified the highest ability individual not willing to take the contract. All individuals with lower ability will also not buy it.

Finally, we provide details on the numerical results presented in section 2. The economy is the same as the one described in section 2, except that we introduce insurable idiosyncratic productivity to make it more comparable with the calibrated economy of section 3. The utility function and its parameters are the same as the one used in section 3. We use a log-normal distribution for both the permanent and the temporary component of abilities with a variance of log earnings equal to 0.3 and about 1/3 of total variance due to the temporary component of abilities. Finally, the initial tax rate is set to 40% and the real interest rate  $r$  to 0.04. We also assume that for each agent wealth equals five times their pre-contract income.

Figure 7: Optimal choices with and without contract

