

On the Optimality of the Friedman Rule with Heterogeneous Agents and Non-Linear Income Taxation*

Carlos E. da Costa

Getulio Vargas Foundation

Iván Werning

MIT

We study the optimal inflation tax in an economy with heterogeneous agents subject to nonlinear taxation of labor income. We find that the Friedman rule is Pareto efficient when combined with a nondecreasing labor income tax. In addition, the optimum for a Utilitarian social welfare function lies on this region of the Pareto frontier. The welfare costs from inflation are bounded below by the area under the demand curve.

1 Introduction

Friedman (1969) argued that positive nominal interest rates represent a distortive tax on real money balances. To reach the first-best these distortions should be removed, the nominal interest rate should be set to zero. This prescription, known as the Friedman rule, is a cornerstone in monetary economics. Phelps (1973) countered that the second-best world we live in requires tolerating distortions due to government taxation, and that positive taxes are set on most goods. Why should money be treated any differently? What is so special about money? He concluded that money should generally be taxed, that nominal interest rates should be positive.

More recently, many studies have explored the optimal inflationary tax on money in a Ramsey tax setting, assuming proportional taxation and a representative-agent economy.¹ This paper reexamines the optimal inflation tax in a model that explicitly incorporates the distributional concerns that lead to distortionary taxation. Our model builds on a standard dynamic equilibrium framework with money, of the same kind used to examine the optimality of the Friedman rule in Ramsey settings. However, we incorporate agent heterogeneity in productivity and allow nonlinear labor-income taxation. As in Mirrlees (1971), distortionary taxation emerges by assuming that individual productivities are private information.

* We are indebted to the editor and three anonymous referees for their helpful suggestions. We thank the comments received from Stefania Albanesi, Fernando Alvarez, Luis Braido, Francisco Buera, Emmanuel Farhi, Hugo Hopenhayn, Paulo Monteiro, Casey Mulligan, Pedro Teles and various seminar and conference participants at the University of Chicago, Columbia, MIT, Northwestern, Universidad Torcuato di Tella, the Federal Reserve Bank of Minneapolis, the Bank of Portugal and the AER meetings. We are grateful to Dan Cao for his invaluable research assistance and comments.

¹ For example, Chari, Christiano, and Kehoe (1996), Correia and Teles (1996, 1999), Lucas and Stokey (1983), Kimbrough (1976), Guidotti and Vegh (1993) and Mulligan and Sala-I-Martin (1997).

We work with a general money-in-the-utility-function framework. As is well known, this framework nests several specific models of money. An important assumption of our analysis is that money and work effort are complements, so that the demand for money, conditional on the expenditure of goods, weakly increases with the amount of work effort. This assumption is motivated by the notion, stressed by various theories, that money's liquidity services facilitate transactions and save on the time required for purchases. It is satisfied, under standard assumptions, for two common specifications: the shopping-time model (McCallum and Goodfriend, 1987; Lucas, 2000) and the cash-credit model (Lucas and Stokey, 1983; Prescott, 1987; Aiyagari, Braun, and Eckstein, 1998; Erosa and Ventura, 2002).

An important aspect of our analysis is that, instead of adopting a particular social welfare function, such as a Utilitarian criterion, we study Pareto efficient arrangements. Our main result is that the Friedman rule is optimal whenever labor income is positively taxed. That is, an increasing income-tax schedule coupled with a zero inflation tax yields a Pareto efficient allocation. Positive taxation of income identifies the relevant region of the Pareto frontier where redistribution takes place from high- to low-productivity individuals. As we also show, this is the region where the optimum for a Utilitarian planner lies.

To interpret our result, it is important to understand the auxiliary role played by the taxation of money, when nonlinear income taxation is present. When redistribution takes place from high- to low-productivity individuals, a tax on money is useful only if it aids in this redistribution. From a mechanism-design perspective, it must relax the incentive-compatibility constraints which ensure that individuals do not underreport their productivity. For this to be the case, an agent who deviates from truth-telling, by underreporting productivity, must demand more money than the lower productivity agent he claims to be. But when money and work effort are complements, exactly the reverse is true: both individuals share the same before- and after-tax income, but the one underreporting productivity requires less work effort and demands the same or less money.

We also examine the welfare costs of deviating from the Friedman rule when it is optimal. In the absence of labor income taxation, the area under the demand curve measures the welfare costs from inflation (Lucas, 2000). We show that when labor income is positively taxed, the area under the demand curve calculation provides a lower bound on the welfare costs.

There are two important differences of our approach with that of previous contributions within the representative-agent Ramsey literature. First, we consider a richer set of tax instruments. Namely, the tax on labor income is allowed to be nonlinear. Moreover, the set of instruments we consider can be justified by private information regarding productivity. Second, our model incorporates heterogeneity, allowing us to capture potentially important

distributional effects from inflation. In particular, the evidence described in [Mulligan and Sala-i-Martin \(2000\)](#), [Erosa and Ventura \(2002\)](#), [Albanesi \(2007\)](#) and others, paints a rich picture of the cross-sectional holdings of money, suggesting that poorer households hold more money as a fraction of their expenditure.

Two important papers on the inflation tax are [Chari, Christiano, and Kehoe \(1996\)](#) and [Correia and Teles \(1996\)](#). Both derive conditions on preferences or technology for the Friedman rule to be optimal within a Ramsey setting. For a cash-credit model, [Chari, Christiano, and Kehoe](#) show that the Friedman rule is optimal if preferences over goods are separable from work effort, so that utility can be written as $\tilde{U}(h(c^1, c^2), n)$, and provided that the subutility function over goods, $h(c^1, c^2)$, is homogeneous. [Correia and Teles](#) work within a shopping-time framework and show that the Friedman rule is optimal if the transactions technology $s(c, m)$ is homogeneous. In our model, with nonlinear taxation, the optimality of the Friedman rule does not require these homogeneity assumptions.

Our results are related to the public-finance literature on optimal mixed taxation. In particular, [Atkinson and Stiglitz's \(1976\)](#) uniform-tax result shows that, when preferences are weakly separable between work effort and consumption goods, only labor-income taxation is needed to achieve the optimum. Indeed, this is true for all Pareto efficient allocations. However, as we argued above, in our context, separability between money and work effort seems like a poor assumption. When work effort and money are strict complements, we show that, because negative nominal interest rates are not possible, a zero inflation tax is Pareto efficient as a corner solution. This holds on the subset of the frontier where redistribution runs from high- to low-productivity individuals. Our analysis relates this region to tax schedules that are increasing in labor income. In this way, we provide joint restrictions, on the taxation of labor and money, for Pareto efficiency. To the best of our knowledge, this aspect of our approach is novel.

The next section introduces the model and the planner's problem. [Section 3](#) derives our main results on the optimality of the Friedman rule. [Section 4](#) examines the welfare costs of inflation. [Section 5](#) contains our conclusions. Proofs are collected in the appendix.

2 Model Setup

Our model economy is similar to those used in representative-agent Ramsey model such as [Lucas and Stokey \(1983\)](#), [Chari, Christiano, and Kehoe \(1996\)](#) and [Alvarez, Kehoe, and Neumeyer \(2004\)](#). The main difference is that our economy is populated by a continuum of infinitely-lived individuals with fixed differences in productivity. The purpose of this assumptions is to incorporate heterogeneity, in the spirit of [Mirrlees's \(1971\)](#) private information framework, in a simple and tractable way. Distortionary taxation then arises as a

consequence of redistribution. Note that our model can nest the representative-agent case if we consider a degenerate distribution of productivities. However, in this case, the first-best allocation can be achieved by a labor-income tax schedule with zero slope. Since we do not rule out these lump-sum tax schedules, heterogeneity is essential for the emergence of distortionary taxation.

2.1 Preferences and Technology

The economy is populated by a continuum (measure one) of individuals with identical preferences represented by the discounted sum of utility

$$\sum_{t=0}^{\infty} \beta^t u(c_t, n_t, m_t), \quad (1)$$

where $\beta < 1$. Here c_t , n_t and m_t represent consumption, work effort and real money balances, respectively. Real money balances are $m_t \equiv M_t/P_t$ where M_t is nominal money balances, and P_t is the price level. We assume that the utility function $u(c, n, m)$ is continuous, strictly increasing in c , decreasing in n , increasing in m , strictly concave and twice continuously differentiable.

Individuals are indexed by their labor productivity w which is distributed in the population according to the cumulative distribution function $F(w)$ for $w \in W = [\underline{w}, \bar{w}]$. The resource constraints are

$$\int (Y_t(w) - c_t(w)) dF(w) \geq G \quad t = 0, 1, \dots \quad (2)$$

where $Y_t(w) = wn_t(w)$ is the output produced by individuals with productivity w ; G is government consumption. Real money $m_t(w)$ is a free good: it does not appear in the resource constraints.

It will be useful to define the indirect utility function

$$V(y, Y, R, w) \equiv \max_{c, m} u(c, \frac{Y}{w}, m) \quad \text{s.t.} \quad c + Rm \leq y, \quad (3)$$

where Y , y and R represent before-tax income, after-tax income and the nominal interest rate, respectively; let $\gamma(y, Y, R, w)$ and $\mu(y, Y, R, w)$ denote the solutions for c and m , respectively. The indirect utility function $V(y, Y, R, w)$ is concave and strictly quasi-concave in (y, Y) for given (R, w) it is increasing in (y, w) , and decreasing in (Y, R) . Let the expenditure function $e(v, Y, R, w)$ denote the inverse of $V(\cdot, Y, R, w)$.

The most important assumption we make is that money and work effort are complements,

so that the demand for money rises with Y/w , implying that $\mu(y, Y, R, w)$ is increasing in Y and decreasing in w . Finally, we also make the standard assumption that both consumption and money are normal goods, given work effort. These assumptions can be expressed in terms of the marginal rate of substitution between consumption and money.

Assumption 1. *The marginal rate of substitution function $u_m(c, n, m)/u_c(c, n, m)$ is decreasing in m , and is increasing in n and c ; in addition, $u_m(c, n, m)/u_c(c, n, m) \rightarrow \infty$ as $m \rightarrow 0$ for fixed c and n .*

The first, and crucial, part of Assumption 1 captures the idea that money provides liquidity services that economize on the time needed for consumption purchases. This idea is at the center of many theories of money. In particular, it holds in the following two special cases of our money-in-the-utility-function setup:

- (A) In the shopping-time model (McCallum and Goodfriend, 1987; Lucas, 2000), a utility function U is defined over consumption and non-leisure time. Consumption requires shopping time, and money serves to economize on this time. Let $s(c, m)$ denote the shopping time required to obtain consumption c with money balances m . This maps into the money-in-the-utility function as follows:

$$u(c, n, m) = U(c, n + s(c, m)).$$

Assumption 1 then follows from standard normality assumptions. Intuitively, a reduction in work time decreases the need for time-saving money balances.

- (B) In the cash-credit model, introduced by Lucas and Stokey (1983), a utility function \tilde{U} is defined over two consumption goods, c^1 and c^2 , and work effort n . The credit-good, c^1 , can be purchased with credit while the cash-good, c^2 , requires money up-front: $c^2 \leq m$. At an optimum, this latter constraint binds so that defining $c \equiv c^1 + c^2$ and can write

$$u(c, n, m) = \tilde{U}(c - m, m, n).$$

If consumption goods (c^1, c^2) are weakly separable from work effort n in the utility function \tilde{U} , then then (c, m) are weakly separable from n in $u(c, n, m)$. In this case, the demand for money balances $\mu(y, Y, R, w)$ is independent of Y and w . For the cash-credit model, the separable case is a benchmark in the Ramsey literature (Chari, Christiano, and Kehoe, 1996).

We also make the following standard assumptions. Work effort Y is an inferior good (i.e. that leisure is a normal good) and that expenditures y are a normal good. These assumptions can be expressed in terms of the marginal rate of substitution between income and output.

Assumption 2. *The marginal rate of substitution function $-V_Y(y, Y, R, w)/V_y(y, Y, R, w)$ is increasing in y and Y . In addition, for any v, R, w , the slope on the indifference curve satisfies $-V_Y(e(v, Y, R, w), Y, R, w)/V_y(e(v, Y, R, w), Y, R, w) > 1$ for large enough Y .*

The assumed normality of expenditures ensures that abler individuals choose to produce more. It implies the single-crossing condition that $-V_Y(y, Y, R, w)/V_y(y, Y, R, w)$ is decreasing in w , so that the indifference curves over (y, Y) become flatter as productivity w rises. The second condition in Assumption 2 ensures that output choices are bounded when agents are confronted with a nondecreasing tax schedule.

Following [Mirrlees \(1971\)](#), we assume that individual productivities and work effort are private information. This rules out type-specific lump-sum taxation and leads, instead, to nonlinear taxation of labor income. In addition, we assume that individual money balances are not observed by the government. This constrains the taxation of money to be linear and leads us to study a mixed taxation problem, where the taxation of labor income is unrestricted but the taxation of money is linear.²

2.2 Competitive Equilibria with Taxes

Individuals face the sequence of budget constraints

$$P_t c_t + M_t + B_t \leq P_t(Y_t - T_t) + M_{t-1} + (1 + r_{t-1})B_{t-1} \quad t = 0, 1, \dots,$$

where B_t represents nominal bond holdings, r_t is the *nominal* interest rate and $T_t = T_t(Y^t)$ are income taxes that may depend on the history of earned income $Y^t = (Y_0, Y_1, \dots, Y_t)$. We set initial nominal wealth to zero $M_{-1} + (1 + r_{-1})B_{-1} = 0$, to make the initial price level irrelevant and focus, instead, on the determination of inflation and nominal interest rates. We also impose a standard No-Ponzi constraint so that the budget constraints become equivalent to the present-value constraint

$$\sum_{t=0}^{\infty} \psi_t (c_t + R_t m_t - y_t) \leq 0, \tag{4}$$

with after-tax income $y_t \equiv Y_t - T_t$, where $R_t \equiv \frac{r_t}{1+r_t}$ and (using that $1 - R_t = \frac{1}{1+r_t}$)

$$\psi_t \equiv \frac{P_t}{P_0} \prod_{s=0}^{t-1} \frac{1}{1+r_s} = \frac{P_t}{P_0} \prod_{s=0}^{t-1} (1 - R_s)$$

² The formal equivalence between the mechanism-design problem when agents can trade in side markets and the mixed taxation problem is proved in [Hammond \(1987\)](#). The constraint imposed on the mechanism by the presence of side markets finds its counterpart in the tax system by restricting the taxation of goods traded in these markets to be linear.

denotes the price of consumption in period t . The budget constraint (4) shows that the opportunity cost of holding real money balances is equal to a simple transformation of the nominal interest rate $R_t = r_t/(1 + r_t)$. From now, we abuse terminology and call R_t the nominal interest rate. The government also faces a budget constraint, which by a version of Walras law is implied by the individuals' budget constraints (holding with equality) and the resource constraints.

Definition. A competitive equilibrium with taxes $\{T_t(Y^t), R_t\}$ is a sequences of real prices $\{\psi_t\}$, real quantities $\{c_t(w), Y_t(w), m_t(w)\}$ and nominal money and price levels $\{M_t, P_t\}$ with $\psi_t \equiv (P_t/P_0) \prod_{s=0}^{t-1} (1 - R_s)$ such that:

- (i) individuals optimize: $\{c_t(w), Y_t(w)/w, m_t(w)\}$ maximizes utility (1), for each w , subject to the budget constraint (4) taking taxes $\{T_t(Y^t)\}$ and prices $\{\psi_t, R_t\}$ as given;
- (ii) markets clear: the resource constraint (2) holds and $m_t = M_t/P_t$.

The stationary economic environment leads us to focus on stationary equilibria, with constant values $(y(w), Y(w), R)$, with consumption and real money balances given by the policy functions $\gamma(y(w), Y(w), R)$ and $\mu(y(w), Y(w), R)$ from (3), and with nominal balances and nominal prices that grow at the constant rate $\beta(1 + R)$.³ Since everyone faces the same budget constraint it follows that individuals with productivity w cannot prefer their own bundle to that chosen by individuals with productivity $w' \in W$, so that

$$V(y(w), Y(w), R, w) \geq V(y(w'), Y(w'), R, w) \quad \forall w, w' \in W \quad (5)$$

and we say that the triplet $(y(w), Y(w), R)$ is *incentive compatible*. The *resource constraint*

$$\int_W (Y(w) - \gamma(y(w), Y(w), R, w)) dF(w) \geq G \quad (6)$$

must also be satisfied. The converse is also true, inequalities (5) and (6) characterize all stationary equilibria that are attainable for some tax policy $\{T_t(Y^t)\}$. Consequently, we say that a triplet $(y(w), Y(w), R)$ is *feasible* if it is incentive compatible and satisfies the resource constraint. Note that, from the consumer's budget constraint, $Y(w) - \gamma(y(w), Y(w), R, w) = Y(w) - y(w) + R\mu(y(w), Y(w), R, w)$ represents the tax revenue for the government. So the resource constraint (6), can be interpreted as the government's budget constraint.

³ Stationary allocations are without loss in generality when the government is allowed to publicly randomize.

2.3 Pareto Efficiency

We say that $(y(w), Y(w), R)$ is *Pareto dominated* by $(\hat{y}(w), \hat{Y}(w), \hat{R})$ if the latter delivers higher utility for all individuals:

$$V(\hat{y}(w), \hat{Y}(w), \hat{R}, w) \geq V(y(w), Y(w), R, w)$$

and strictly so for a subset of W with positive measure. A Pareto efficient $(y^*(w), Y^*(w), R^*)$ must maximize the tax revenues collected

$$\int_W (Y(w) - \gamma(y(w), Y(w), R, w)) dF(w) \quad (7)$$

subject to $V(y(w), Y(w), R, w) \geq V(y^*(w), Y^*(w), R^*, w)$ and incentive compatibility (5); otherwise, a Pareto improvement is possible by lowering taxes.

2.4 Implementation

There are several tax systems $\{T_t(Y^t)\}$ that can implement any feasible $(y(w), Y(w), R)$. Here we describe a few possibilities. For all of them it is useful to define the *static nonlinear income tax schedule* by

$$T(Y) \equiv \inf \{z : V(Y - z, Y, R, w) \leq V(y(w), Y(w), R, w) \quad \forall w \in W\}, \quad (8)$$

so that $T(Y(w)) = Y(w) - y(w)$ for all $w \in W$. This tax function corresponds to the lowest schedule that implements $(y(w), Y(w), R)$ in a static setting where individuals have preferences given by $V(y(w), Y(w), R)$. As we now discuss, it also plays a key role in implementing stationary equilibria for our dynamic setting.

The most natural candidate tax system $\{T_t(Y^t)\}$ is a history-independent one, where the tax schedule in each period coincides with the static schedule defined above, so that $T_t(Y^t) = T(Y_t)$. Due to its simplicity, this type of policy is of special interest. Suppose the government tax policy is history independent, so that individuals face some interest rate R and some fixed increasing tax schedule $T(Y)$ in all periods $t = 0, 1, 2, \dots$. Suppose this induces a stationary allocation $(y(w), Y(w), R)$. The question is then whether the resulting allocation is Pareto efficient. Our results in the next section provide the answer, stating conditions for the allocation to be efficient if $R = 0$, and inefficient otherwise.

To see how a history-independent policy may implement a stationary $(y(w), Y(w), R)$ as an equilibrium with prices $\psi_t = \beta^t$ and $R_t = R$, note that individuals will find $(y_t(w), Y_t(w)) = (y(w), Y(w))$ optimal, conditional on choosing a path for output $\{Y_t(w)\}$ that is constant

over time. All that remains to ensure that this history-independent tax scheme implements the stationary equilibrium $(y(w), Y(w), R)$, is to make sure that individuals choose a constant path for output. Indeed, the first-order necessary conditions for the individual's optimization problem are satisfied with the constant path $(y(w), Y(w), R)$. If the tax function $T(Y)$ is convex, the individual's problem is convex and the first-order conditions are then sufficient for optimality. Thus, convexity of $T(Y)$ guarantees implementation of the stationary $(y(w), Y(w), R)$. Of course, this is a sufficient, not necessary, condition: since utility $u(c, n, m)$ is concave over work effort n , a constant output path may be optimal for individuals if the tax schedule is not too concave.⁴

If the tax schedule $T(Y)$ is concave enough that a history-independent tax policy fails to implement the allocation $(y(w), Y(w), R)$, then there are several tax systems $\{T_t(Y^t)\}$ that introduce limited forms of history dependence to ensure that individuals choose constant output paths and implement any feasible $(y(w), Y(w), R)$. For example, the government can set taxes so that $T_t(Y^t) = T(Y_t)$ whenever $Y_t = Y_{t-1}$, while setting $T_t(Y^t)$ high enough if $Y_t \neq Y_{t-1}$. This imposes the static tax schedule along the equilibrium path, with constant output, but penalizes individuals that deviate, off the equilibrium, to non-constant output paths. Note that even in this example the static-tax function $T(\cdot)$ plays a critical role. Thus, from now on we summarize the allocation and tax system in terms of $T(\cdot)$ and R .

3 Optimality of the Friedman Rule

We now study monetary policy. We first show that if a policy involves $R > 0$, the government can reduce the interest rate and increase the tax schedule in a way that increases work effort and leaves utility unchanged for each type of individual. This simple result then underlies our main result on the Pareto optimality of the Friedman rule.

3.1 Pareto Efficiency and Positive Income Taxation

We say that an allocation is downward incentive compatibility if

$$V(y(w), Y(w), R, w) \geq V(y(w'), Y(w'), R, w) \quad \forall w' \leq w \text{ and } w, w' \in W.$$

Starting from any incentive-compatible allocation $(y(w), Y(w), R)$, with $R > 0$ we now construct another allocation, with a lower interest rate, that is downward incentive compatible, maintains individuals' utility and saves resources.

⁴ If the net-income schedule $\tilde{Y} - T(\tilde{Y})$ has regions that are too convex then some individuals may prefer working harder in some periods than others. A similar issue can arise in a [Mirrlees \(1971\)](#) static settings if consumption itself is not controlled by the planner and individuals can engage in lotteries: randomizing on the output they produce and pooling their net income allows them to reduce their total tax liability.

For $\hat{R} \leq R$, the new allocation is as follows. Output $Y(w)$ is unchanged. After tax income is set to $\hat{y}(w; \hat{R})$ so that individuals' utility is maintained at the original level

$$V(\hat{y}(w; \hat{R}), Y(w), \hat{R}, w) = V(y(w), Y(w), R, w). \quad (9)$$

As a result, after-tax income $\hat{y}(w; \hat{R})$ and consumption $\gamma(\hat{y}(w; \hat{R}), Y(w), \hat{R}, w)$ are both increasing in \hat{R} .

We now argue that this allocation is downward incentive compatible. For fixed Y , consider how the preferences over (y, R) pairs vary with w . For any $w' < w$

$$\mu(y, Y, R, w') = \frac{-V_R(y, Y, R, w')}{V_y(y, Y, R, w')} \geq \frac{-V_R(y, Y, R, w)}{V_y(y, Y, R, w)} = \mu(y, Y, R, w).$$

Thus, the indifference curve over (y, R) for type w' crosses that of type w at most once, from below. Since $\hat{y}(w'; \hat{R})$ is set to compensate w' (i.e. equation (9) holds at w') it follows that

$$V(\hat{y}(w'; \hat{R}), Y(w'), \hat{R}, w) \leq V(y(w'), Y(w'), R, w) \quad (10)$$

for $w' < w$. Since the original allocation is downward incentive compatible, equation (9) and inequality (10) imply that the new allocation is as well. Note that reducing \hat{R} increases total taxes $\int (Y(w) - \gamma(\hat{y}(w; \hat{R}), Y(w), \hat{R}, w)) dF(w)$. We have proved the following result.

Lemma 1. *Let Assumptions 1 and 2 hold, and let $(y(w), Y(w), R)$ be any incentive compatible allocation with $R > 0$. Then for any $\hat{R} \leq R$ there exists an allocation $(\hat{y}(w; \hat{R}), Y(w), \hat{R})$ that is downward incentive compatible and gives each individual the same utility, so that equation (9) holds. Both $\hat{y}(w; \hat{R})$ and $\gamma(\hat{y}(w; \hat{R}), Y(w), \hat{R}, w)$ are increasing in \hat{R} .*

We say that *income taxation is positive* if the income tax schedule $T(Y)$ is nondecreasing. Intuitively, at an efficient allocation, if income taxation is positive, then redistribution takes place from higher- to lower-type individuals and it is the downward incentive constraints that are relevant; the upward incentive constraints are slack. The lemma then implies that the Friedman rule, setting $R = 0$, is optimal. Proposition 1, the main result of the paper, makes this precise.

Proposition 1. *Let Assumptions 1 and 2 hold, and let $(y^*(w), Y^*(w), R^*)$ be a feasible allocation induced by a nondecreasing tax function $T^*(Y)$ and interest rate R^* . If $R^* = 0$ and the allocation $(y^*(w), Y^*(w), R^*)$ is not Pareto dominated by any feasible allocation $(\hat{y}(w), \hat{Y}(w), \hat{R})$ with $\hat{R} = 0$, then $(y^*(w), Y^*(w), R^*)$ is Pareto efficient. Indeed, $(y^*(w), Y^*(w), R^*)$ Pareto dominates any feasible allocation $(y(w), Y(w), R)$ with $R > 0$.*

Proposition 1 establishes that the Friedman rule is optimal in the sense that it produces a Pareto efficient outcome when combined with positive taxation of income. The result requires the income-tax schedule to be efficient conditional on $R = 0$. What the proposition establishes is that there are no possible Pareto improvements from shifting to $R > 0$ and rearranging the tax schedule. Although we have assumed a bounded set of types, as their proofs reveal, both lemma 1 and proposition 1 hold even when W is unbounded.

Characterizing the set of increasing income-tax schedules that are Pareto efficient is beyond the scope of this paper. However, it is worth remarking that, for example, a linear schedule (i.e. a flat tax) is Pareto efficient for a large set of distributions that are continuous on an unbounded support (Werning, 2007). By implication, the result in proposition 1, that the Friedman rule is optimal whenever income is positively taxed, does not require the optimal tax schedule $T(Y)$ to be nonlinear.

We now provide a converse result, giving conditions that ensure that taxing money $R > 0$ is inefficient. Define the total tax, income taxes plus seignorage, collected from an agent of type w producing Y as

$$\mathcal{T}(Y; w) \equiv T(Y) + R\mu(Y - T(Y), Y, R, w). \quad (11)$$

Our next result concerns situations in which at the original allocation the total marginal tax on income is positive:

$$\frac{\partial}{\partial Y} \mathcal{T}(Y(w); w) \geq 0. \quad (12)$$

Note that Assumption 1 implies that $\mu(y, Y, R, w)$ is increasing in y and Y , so this condition is weaker than $T'(Y) \geq 0$.

Proposition 2. *Let Assumptions 1 and 2 hold, and let $(y(w), Y(w), R)$ be a feasible allocation induced by a continuously differentiable tax function $T(Y)$ satisfying (11)–(12) at all points of differentiability and interest rate $R > 0$. Suppose $Y(w)$ is piecewise continuously differentiable with $Y'(w)$ bounded away from zero. Then there exists a tax function $\hat{T}(Y)$ and interest rate $\hat{R} \leq R$ that induces a feasible allocation $(\hat{y}(w), \hat{Y}(w), \hat{R})$ that Pareto dominates $(y(w), Y(w), R)$.*

The reason proposition 2, unlike proposition 1, requires ruling out bunching is that if two different types were to produce the same output, then the lower types would demand more money and pay more total taxes than the high types; that is, $\mathcal{T}(Y(w), w)$ would be strictly decreasing in w over any region where $Y'(w) = 0$. Thus, redistribution would be taking place in the nonstandard direction: from lower- to higher-type individuals.

This result guarantees that, if income is positively taxed, all individuals prefer to move

towards the Friedman rule. Unlike proposition 1, the availability of a nonlinear income tax is crucial for proposition 2. In particular, even if the original tax schedule $T(Y)$ is linear, the alternative tax schedule $\hat{T}(Y)$ that, along with $\hat{R} < R$, guarantees a Pareto improvement, may be nonlinear. Consequently, if one imposes restrictions on the set of available tax schedules $\hat{T}(Y)$ then Pareto improvements over $(T(Y), R)$ may not be available. For example, Albanesi (2007) studies a cash-credit model with heterogeneity, but imposes proportional labor income taxation. With such a constraint, deviating from the Friedman rule may not be Pareto inefficient.

More generally, what is crucial for proposition 2 is that income-taxation be sufficiently rich relative to the sources of heterogeneity. In our model, as in the canonical Mirrlees (1971), the source for heterogeneity is differences in productivity. Since this leads to differences in output, a nonlinear income-tax schedule is a rich enough instrument to separate individuals. With additional sources of heterogeneity this may no longer be the case (Saez, 2002).

3.2 Discussion

Proposition 1 and 2 are illustrated in Figure 1, which plots the Pareto frontier for the case with two productivity types. The dotted line is the unconstrained Pareto frontier, i.e. the first-best that obtains with type-specific lump-sum taxation. The solid and dashed line represent constrained Pareto frontiers (without type-specific lump-sum taxation). The solid line imposes the Friedman rule, $R = 0$, and optimizes over the income tax schedule; the dashed line imposes some $R > 0$ and optimizes over the income tax schedule. Point A on the figure represents the “autarky” point with no taxation. At this point the solid and dotted lines meet.

Proposition 1 applies whenever income taxation is positive, representing the region to the left of the autarky point A , with redistribution from high- to low-type individuals. The inflation tax interacts unfavorably with positive income taxation because it increases the cost of separating the high- and low-type individuals since, *ceteris paribus*, higher types demand less money. For the same reason, to the right of point A , a positive tax on money may be optimal when redistribution runs the other way.

To gain intuition for these results, and the role played by the complementarity of money and work effort, it is useful to consider a simple example where

$$u(c, n, m) = U(c, \min \langle l(n), m \rangle)$$

for some strictly increasing function $l(n)$. Money demand is then a given function of work effort $\mu(y, Y, R, w) = l(Y/w)$. The important point is that, from the point of view of an individual with productivity w , facing $T(Y)$ and $R > 0$ is equivalent to facing the fictitious,

type-contingent, tax schedule $\mathcal{T}(Y; w) = T(Y) + Rl(Y/w)$ and $\hat{R} = 0$. Thus, when $R > 0$, the tax schedule \mathcal{T} depends negatively on productivity w . If income is positively taxed, this is inefficient since it confronts individuals with a tax that increases with output Y , in an attempt to redistribute from high- towards low-productivity individuals, only to make it decrease with productivity w , redistributing in the opposite direction. Removing the dependence on w , by setting $R = 0$, allows for a reduction in the dependence on Y , which reduces distortive marginal taxes without affecting redistribution.

Finally, we studied the realistic case where money can only be taxed proportionally but labor-income can be taxed nonlinearly. However, the results extend to the case where money can be taxed nonlinearly, so that the government can confront individuals with a tax function that depends on both Y and m .

3.3 Utilitarian Optimum

Finally, we relate the Pareto efficient allocations identified in proposition 1 to the optimum for a Utilitarian social welfare function. The *Utilitarian planning problem* is to maximize

$$\int V(y(w), Y(w), R, w) dF(w)$$

subject to incentive compatibility (5) and the resource constraint (6). The next result relies on showing that only the downward incentive constraints bind, that the solution to a relaxed problem that ignores the upward incentive constraints does not violate them. The result then follows from Lemma 1.

Proposition 3. *Let Assumptions 1 and 2 hold and suppose there exists some feasible allocation satisfying (5) and (6). Then a solution to the Utilitarian planning problem exists and can be implemented by an increasing income tax schedule $T^*(Y)$ with $R^* = 0$.*

A Utilitarian chooses positive taxes on income and a zero tax on money balances. Redistribution runs from high- to low-productivity individuals. That is, the relevant region of the Pareto efficient frontier is precisely that identified by proposition 1, to the left of point A on the figure.

4 Welfare Costs of Inflation

The previous section established the optimality of the Friedman rule. In this section, we examine the welfare losses of deviating from this optimum.

Suppose the hypothesis of proposition 1 hold. Let $(y^*(w), Y^*(w), R^*)$ with $R^* = 0$ denote a Pareto efficient allocation and let $E(R)$ stand for the maximized value of total tax receipts

in problem (7) for given R . Let the aggregate money balances obtained from the solution to this problem be denoted by

$$\mathcal{M}(R) \equiv \int \mu(y(w; R), Y(w; R), R, w) dF(w),$$

where $y(w; R)$ and $Y(w; R)$ solve the problem (7) for given R . Note that this demand schedule incorporates the changes in the income tax $T(Y; R)$ required to compensate individuals so that their welfare does not fall below the baseline given by $V(y^*(w), Y^*(w), R^*)$.

Our measure of welfare losses is $E(0) - E(R)$, which represents the additional resources needed so that no one is made worse off when $R > 0$. For low enough R the constraints that $V(y(w), Y(w), R, w) \geq V(y^*(w), Y^*(w), R^*, w)$ will generally bind. One can then show that

$$E(0) - E(R) = \left(\int_0^R \mathcal{M}(\tilde{R}) d\tilde{R} - R\mathcal{M}(R) \right) - \int_0^R \int \tau(w; \tilde{R}) \cdot \frac{\partial}{\partial R} Y(w; \tilde{R}) dF(w) d\tilde{R}, \quad (13)$$

and that $\partial Y / \partial R \leq 0$, with strict inequality if and only if money and work effort are strict complements, $\mu_w(y, Y, R, w) < 0$.⁵ The term within parenthesis in (13) represents the deadweight-loss triangle computed from the area under the money demand $\mathcal{M}(R)$. The other term captures the effect that inflation has on the income tax revenue. When money and work effort are strict complements, higher inflation reduces work effort which lowers the amount collected from the income tax.

When income taxation is positive, so that $\tau(w; R) \geq 0$, equation (13) reveals that welfare losses are bounded below by an area-under-the-demand-curve calculation. The two coincide only when money and work effort are not complements, so that $\mu_w = 0$, as is the case in the cash-credit model when preferences for goods are separable from work effort.

To illustrate, we compute the welfare losses in a shopping-time model, for a specification that closely follows Lucas (2000), which considers the welfare costs for a representative-agent economy without income taxation. The utility and shopping-time functions are set at

$$U(c, n, m) = \log(c) + \alpha \log(1 - n - s(c, m)) \quad \text{with} \quad s(c, m) = \frac{c}{km}$$

for constants $\alpha, k > 0$. We assume the initial tax schedule is proportional $T(Y) = \bar{\tau}Y$ for some $\bar{\tau} \geq 0$. We set $\alpha = 2$ so that $n = 1/(\alpha + 1) = 1/3$, and set $k = 1200$ which implies the same level of money demand calibrated by Lucas.⁶ For this example, welfare calculations

⁵This expression for $E(R)$ follows by rearranging and integrating the last expression in the proof of proposition 2.

⁶The ratio of money balances to consumption m/c is approximately $\sqrt{\frac{\alpha+1}{kR}}$. Lucas (2000) argues that this provides a good fit for the relation between interest rates and the ratio of monetary aggregate M1 to

turn out to be independent of the skill distribution, so we do not need to specify $F(w)$.⁷

We compute $E(R) - E(0)$ as a fraction of aggregate consumption, evaluated at the original Pareto efficient allocation. In the appendix we show that this measure is given by

$$(1 - S(R))^{-\alpha} + \frac{S(R)}{1 - \bar{\tau}} - 1.$$

where $S(R)$ is the equilibrium shopping time expressed as a function of R (it turns out to be independent of w). When $\tau = 0$ and $\alpha \rightarrow 0$ the welfare cost is simply the shopping time $S(R)$, just as obtained by Lucas (2000) in a version without work effort or income taxation.

Figure 2 displays this cost measure against the nominal interest rate R for three initial tax rates: $\bar{\tau} = 0$ (lower dashed line), $\bar{\tau} = 35\%$ (middle dashed-dotted line) and $\bar{\tau} = 50\%$ (top dotted line). Also plotted is the contribution from the area under the demand schedule $\mathcal{M}(R)$ (solid line), which, normalized by aggregate consumption, turns out to be independent of the initial τ . Behind these calculations, for $R > 0$ the income tax schedule $T(Y; R)$ is adjusted to keep individuals' utility at their original levels. As it turns out, for our simple example, this schedule remains proportional, and the marginal income tax rate that all individuals face decreases with R .

When $\bar{\tau} = 0$ the figure essentially replicates Lucas's findings. The welfare cost of setting $R = 4\%$ is worth 1% of aggregate consumption. Moving from $R = 4\%$, representing a situation with near zero inflation, to $R = 16\%$ entails an additional 1% of consumption cost. The welfare cost is almost indistinguishable from the area-under-demand-curve, the term enclosed in parenthesis in (13).⁸ As a result, as in Lucas (2000), the area under the demand curve provides an excellent approximation to the welfare costs in this case.

However, relative to $\bar{\tau} = 0$, when $\bar{\tau} = 35\%$ welfare costs are approximately 20% higher, and with $\bar{\tau} = 50\%$ this difference becomes 33%. In both cases, welfare costs are strictly larger than the area-under-the-demand-curve term because $\tau(w; R) > 0$. The example illustrates that the last term in the welfare decomposition (13) has the potential to contribute nontrivially.

5 Concluding Remarks

In this paper we explored the optimal taxation of income and money. Distortionary taxation emerges due to agent heterogeneity and the fact that taxation is anonymous. Under the nominal output in the United States from 1900-1994.

⁷ We only require the original allocation generated by a proportional tax to be Pareto efficient, which holds for a large class of continuous and unbounded distributions (see Werning, 2007).

⁸ Actually, the contribution from the last term in (13) is negative because $\tau(w, R) < 0$ for $R > 0$, but it turns out to be minuscule for the range of R plotted here.

assumption that money and work effort are complements, we found that the Friedman rule is optimal whenever labor income is positively taxed, in the sense that such a tax system produces a Pareto efficient outcome. We made several assumptions for our analysis, we now speculate on their role in our main result.

First, our model abstracted from tax evasion. One argument for a tax on money is that an inflation tax can be easily collected. Whether or not this is a relevant consideration for advanced economies is unclear, but it may be more important for less developed ones, which tend to rely more on inflation as a source of revenue. There are many ways of extending our model to incorporate tax evasion. We conjecture that while tax evasion may provide a rationale for an inflation tax, the exact conclusion may depend on the way evasion is introduced, the incidence of inflation and the redistributive goals.

Second, our dynamic environment abstracted from aggregate and idiosyncratic uncertainty. This allowed us to reduce the policy problem to a simple static subproblem, which, in turn, provided a tight connection between the direction of binding incentive constraints and the sign of income taxation. It also made the tax implementation of efficient allocations relatively simple. Incorporating uncertainty complicates the analysis on both dimensions, but the mechanism isolated in our simple stationary model is likely to remain central.

Appendix

Proof of Proposition 1

We proceed by contradiction. Suppose there exists an alternative allocation $(y(w), Y(w), R)$ with $R > 0$ that is incentive compatible, has

$$V(y(w), Y(w), R, w) \geq V(y^*(w), Y^*(w), 0, w) \quad \text{for all } w \in W,$$

and satisfies the resource constraint (6), implying that tax revenues satisfy

$$\int (Y^*(w) - y^*(w))dF(w) \leq \int (Y(w) - \gamma(y(w), Y(w), R, w))dF(w).$$

Note that incentive compatibility implies that $Y(w)$ is nondecreasing.

Lemma 1 then implies that there exists another allocation $(\hat{y}(w), \hat{Y}(w), \hat{R})$ with $\hat{R} = 0$ and $\hat{Y}(w)$ nondecreasing, that is downward incentive compatible, has

$$V(\hat{y}(w), \hat{Y}(w), 0, w) \geq V(y^*(w), Y^*(w), 0, w) \quad \text{for all } w \in W,$$

and collects higher revenue:

$$\int (Y(w) - \gamma(y(w), Y(w), R, w))dF(w) < \int (\hat{Y}(w) - \hat{y}(w))dF(w).$$

We now show that this is not possible by showing that if this were the case there would exist a tax schedule \tilde{T} strictly below T^* that induces an incentive-compatible allocation $(\tilde{y}(w), \tilde{Y}(w), \tilde{R})$, with $\tilde{R} = 0$, that collects still higher revenue.

Define the tax schedule associated with the alternative allocation as

$$\hat{T}(\theta) \equiv \inf \{z : V(\theta - z, \theta, 0, w) \leq V(\hat{y}(w), \hat{Y}(w), 0, w) \quad \forall w \text{ s.t. } \hat{Y}(w) \geq \theta\}. \quad (14)$$

Although, the tax schedule \hat{T} may not be a continuous function of θ , it can only have downward jumps at points of discontinuity.

A Pareto improvement requires taxes to be lower at the alternative allocation:

$$\hat{T}(\theta) \leq T^*(\theta). \quad (15)$$

Otherwise, if $\hat{T}(\theta_0) > T^*(\theta_0)$ for some θ_0 , there is a type $w_0 \in W$ such that $V(\theta_0 - T^*(\theta_0), \theta_0, 0, w) > V(y^*(w_0), Y^*(w_0), 0, w)$.

Now define the tax schedule

$$\tilde{T}(\theta) \equiv \sup_{\hat{\theta} \leq \theta} \hat{T}(\hat{\theta}),$$

which irons out decreasing regions of \hat{T} . The function \tilde{T} is nondecreasing and continuous (since it removes any downward jumps in \hat{T}). Moreover, inequality (15) and the fact that $T^*(\theta) = \sup_{\hat{\theta} \leq \theta} T^*(\hat{\theta})$ (since $T^*(\theta)$ is nondecreasing) imply that:

$$\tilde{T}(\theta) \leq T^*(\theta) \quad (16)$$

We now consider the allocation generated by this tax function. That is, let the associated incentive-compatible allocation (breaking potential indifference in favor of higher output) be

$$\tilde{Y}(w) \equiv \max\{\arg \max_{\theta} V(\theta - \tilde{T}(\theta), \theta, 0, w)\}, \quad (17)$$

and $\tilde{y}(w) \equiv \tilde{Y}(w) - \tilde{T}(\tilde{Y}(w))$. This allocation is well defined because: (i) \tilde{T} is continuous; and (ii) we can restrict the maximization in (17), for each w , to the set of θ such that $V(\theta - \tilde{T}(\theta), \theta, 0, w) \geq V(\hat{Y}(w) - \tilde{T}(\hat{Y}(w)), \hat{Y}(w), 0, w)$, which is nonempty ($\hat{Y}(w)$ belongs to this set) and compact (using Assumption 2 with the fact that \tilde{T} is nondecreasing and

continuous).

First, it follows immediately from (16) that all agents are better off facing $\tilde{T}(w)$ than facing $T^*(w)$. That is, utility is higher at the resulting allocation $(\tilde{y}(w), \tilde{Y}(w))$ than at $(y^*(w), Y^*(w))$:

$$V(\tilde{y}(w), \tilde{Y}(w), R, w) \geq V(y^*(w), Y^*(w), 0, w) \quad \text{for all } w \in W.$$

Second, we argue that all agents decide to pay more taxes at \tilde{T} than they did at the $\hat{Y}(w)$ allocation with the tax schedule \hat{T} :

$$\hat{T}(\hat{Y}(w)) \leq \tilde{T}(\hat{Y}(w)) \leq \tilde{T}(\tilde{Y}(w)).$$

The first inequality follows immediately by construction, i.e. $\tilde{T}(\theta) \geq \hat{T}(\theta)$ for all θ . For the second inequality, there are two cases to consider. In the first case, $\tilde{T}(\hat{Y}(w)) = \hat{T}(\hat{Y}(w))$, so that taxes were not raised at $\hat{Y}(w)$. Since taxes were not lowered for $\theta \leq Y(w)$ it follows that $\hat{Y}(w) \leq \tilde{Y}(w)$. The inequality then follows since \tilde{T} is nondecreasing. In the second case, $\hat{T}(\hat{Y}(w)) < \tilde{T}(\hat{Y}(w))$, so that taxes were raised at $\hat{Y}(w)$, we argue by contradiction. Suppose $\tilde{T}(\tilde{Y}(w)) < \tilde{T}(\hat{Y}(w))$. Then there must exist a $w' < w$ such that: $\tilde{T}(\tilde{Y}(w)) < \tilde{T}(\hat{Y}(w')) < \tilde{T}(\hat{Y}(w))$ and $\tilde{T}(\hat{Y}(w')) = \hat{T}(\hat{Y}(w'))$; as we just showed, the latter condition implies that $\hat{Y}(w') \leq \tilde{Y}(w')$. Incentive compatibility implies that $\tilde{Y}(w)$ is nondecreasing, so that $\hat{Y}(w') \leq \tilde{Y}(w') \leq \tilde{Y}(w)$. Since \tilde{T} is nondecreasing it follows that $\tilde{T}(\tilde{Y}(w')) \leq \tilde{T}(\tilde{Y}(w))$, a contradiction. Hence, $\hat{Y}(w) - \hat{y}(w) \leq \tilde{Y}(w) - \tilde{y}(w)$, so that

$$G \leq \int (Y^*(w) - y^*(w)) dF(w) < \int (\hat{Y}(w) - \hat{y}(w)) dF(w) \leq \int (\tilde{Y}(w) - \tilde{y}(w)) dF(w). \quad (18)$$

This contradicts the Pareto efficiency of $(y^*(w), Y^*(w))$ subject to $R^* = 0$, since Pareto efficient allocations must minimize net resources, as in (7).

Proof of Proposition 2

We use the following standard characterization of the incentive compatibility constraints (e.g. see Fudenberg and Tirole (1991) and Milgrom and Segal (2002)). For any allocation $(y(w), Y(w), R)$ let $v(w) \equiv V(y(w), Y(w), R, w)$ denote the associated utility assignment. An incentive-compatible allocation that is piecewise continuously differentiable must have $Y(w)$ nondecreasing and satisfy the local incentive constraints:

$$v'(w) = V_w(y(w), Y(w), R, w), \quad (19)$$

almost everywhere. Conversely, if an allocation $(y(w), Y(w), R)$ is piecewise continuously differentiable and has $Y(w)$ nondecreasing and satisfies (19) with $v(w) \equiv V(y(w), Y(w), R, w)$, then it is incentive compatible.

The original allocation $(y(w), Y(w), R)$ is incentive compatible and thus satisfies (19) with $v(w) \equiv V(y(w), Y(w), R, w)$. For any $\hat{R} \leq R$, we now construct a new allocation $(\hat{y}(w; \hat{R}), \hat{Y}(w; \hat{R}), \hat{R})$ with $\hat{v}(w; \hat{R}) = V(\hat{y}(w; \hat{R}), \hat{Y}(w; \hat{R}), \hat{R}, w)$ that maintains the same utility profile $\hat{v}(w; \hat{R}) = v(w)$, requiring

$$\hat{y}(w; \hat{R}) = e(v(w), \hat{Y}(w; \hat{R}), \hat{R}, w)$$

where the expenditure function $e(v, Y, R, w)$ represents the inverse of the indirect utility function $V(\cdot, Y, R, w)$. We set $\hat{Y}(w; \hat{R})$ to maintain the local incentive compatibility constraints (19) yielding

$$V_w(y(w), Y(w), R, w) = V_w(\hat{y}(w; \hat{R}), \hat{Y}(w; \hat{R}), \hat{R}, w).$$

Substituting gives

$$V_w(y(w), Y(w), R, w) = V_w(e(v(w), \hat{Y}(w; \hat{R}), \hat{R}, w), \hat{Y}(w; \hat{R}), \hat{R}, w), \quad (20)$$

a single equation in the unknown $\hat{Y}(w; \hat{R})$. By construction, if the resulting allocation has $\frac{\partial \hat{Y}}{\partial w}(w; \hat{R}) > 0$ for all $w \in W$, then it is incentive compatible. Since $\frac{\partial Y}{\partial w}(w; R) \geq \varepsilon$ for all $w \in W$ for some $\varepsilon > 0$ and W is compact, the implicit function theorem guarantees that $\frac{\partial \hat{Y}}{\partial w}(w; \hat{R}) > 0$ for all $w \in W$ for all $R - \hat{R} < \delta$ for some $\delta > 0$.

We now show that the constructed allocation lowers net resources (7), leading to a contradiction. Differentiating (20) with respect to \hat{R} gives

$$\begin{aligned} 0 = & V_{wY}(\hat{y}(w; \hat{R}), \hat{Y}(w; \hat{R}), \hat{R}, w) \frac{\partial \hat{Y}(w; \hat{R})}{\partial R} + V_{wR}(\hat{y}(w; \hat{R}), \hat{Y}(w; \hat{R}), \hat{R}, w) \\ & + V_{wy}(\hat{y}(w; \hat{R}), \hat{Y}(w; \hat{R}), \hat{R}, w) \\ & \times \left(e_R(v(w), \hat{Y}(w; \hat{R}), \hat{R}, w) + e_Y(v(w), \hat{Y}(w; \hat{R}), \hat{R}, w) \frac{\partial \hat{Y}(w; \hat{R})}{\partial R} \right) \quad (21) \end{aligned}$$

To simplify this expression, note that $e_R(v, Y, w, R) = \mu(e(v, Y, R, w), Y, w, R)$ and that (Roy's identity)

$$V_y(y, Y, R, w)\mu(y, Y, R, w) + V_R(y, Y, R, w) = 0,$$

so that differentiating with respect to w gives

$$V_{wR}(y, Y, R, w) + V_{wy}(y, Y, R, w)\mu(y, Y, R, w) + \mu_w(y, Y, R, w) = 0.$$

Also note that $e_Y(v, Y, R, w) = -V_Y(y, Y, R, w)/V_y(y, Y, R, w)$ (evaluated at $y = e(v, Y, R, w)$) so the single-crossing condition in Assumption 2 implies

$$\frac{\partial}{\partial w} \left(\frac{-V_Y(y, Y, R, w)}{V_y(y, Y, R, w)} \right) = \frac{-V_{wY}(y, Y, R, w) - V_{wy}(y, Y, R, w)e_Y(v, Y, R, w)}{V_y(y, Y, R, w)} < 0.$$

So that solving equation (21):

$$\frac{\partial \hat{Y}(w; \hat{R})}{\partial R} = -\frac{\mu_w(\hat{y}(w; \hat{R}), \hat{Y}(w; \hat{R}), \hat{R}, w)}{\frac{\partial}{\partial w} \left(\frac{-V_Y(\hat{y}(w; \hat{R}), \hat{Y}(w; \hat{R}), \hat{R}, w)}{V_y(\hat{y}(w; \hat{R}), \hat{Y}(w; \hat{R}), \hat{R}, w)} \right)} \leq 0.$$

with strict inequality if $\mu_w(\hat{y}(w; \hat{R}), \hat{Y}(w; \hat{R}), \hat{R}, w) < 0$. Now define

$$E(\hat{R}) \equiv \int \left(\hat{Y}(w; \hat{R}) - e(v(w), \hat{Y}(w; \hat{R}), w, \hat{R}) + \hat{R}e_R(v(w), \hat{Y}(w; \hat{R}), w, \hat{R}) \right) dF(w)$$

Differentiating

$$E'(\hat{R}) = \int \left((1 - e_Y(v(w), \hat{Y}(w; \hat{R}), \hat{R}, w) + \hat{R}e_{RY}(v(w), \hat{Y}(w; \hat{R}), \hat{R}, w)) \frac{\partial \hat{Y}(w; \hat{R})}{\partial R} + \hat{R}e_{RR}(v(w), \hat{Y}(w; \hat{R}), \hat{R}, w) \right) dF(w)$$

and evaluating this at $\hat{R} = R$ gives $E'(R) < 0$, using Assumptions 1 and 2 together with the twice continuous differentiability of $U(c, n, m)$ (to conclude that $e_{RR}(v, Y, R, w) < 0$). Thus, a small reduction in R allows for a strict increase in $E(R)$ which contradicts the Pareto optimality of the original allocation.

Proof of Proposition 3

We first prove a version of the result for a finite types economy and then make a passage-to-the-limit argument to cover the general case. To help the reader see the structure of the argument, we organize the proof into subsections.

Finite Types

A finite type problem. Consider an economy with N types, $w^1 \leq w^2 \leq \dots \leq w^N$, with population fractions π^i . Define the (Utilitarian) planning problem for this economy as

maximizing

$$\sum_{i=1}^N V(y(w^i), Y(w^i), R, w^i) \pi^i \quad (22)$$

subject to

$$V(y(w^i), Y(w^i), R, w^i) \geq V(y(w^j), Y(w^j), R, w^i) \quad \forall i, j = 1, 2, \dots, N, \quad (23)$$

the resource constraint

$$\sum_{i=1}^N (Y(w^i) - \gamma(y(w^i), Y(w^i), R, w^i)) \pi^i \geq G,$$

and $0 \leq y(w^i)$, $0 \leq Y(w^i)$. We show that the optimum for this problem exists, has $R = 0$ and $Y(w^i) - y(w^i)$ increasing in w^i and nonnegative marginal tax rates: $1 + V_Y(y(w^i), Y(w^i), 0, w^i)/V_y(y(w^i), Y(w^i), 0, w^i) \geq 0$. This implies that the allocation can be implemented with a nondecreasing tax schedule using (8).

A finite type relaxed problem We proceed by showing that the optimum solves the following *relaxed Utilitarian problem*: defined exactly as the Utilitarian problem except replacing the incentive compatibility condition (5) with local downward incentive constraints

$$V(y(w^{i+1}), Y(w^{i+1}), R, w^{i+1}) \geq V(y(w^i), Y(w^i), R, w^{i+1}) \quad i = 1, 2, \dots, N - 1, \quad (24)$$

and the monotonicity condition that $Y(w^i) \leq Y(w^{i+1})$ (note that $y(w^i) \leq y(w^{i+1})$ is implied by these constraints). We show that a solution to this relaxed problem exists, has $R = 0$ and all the downward incentive compatibility constraints (24) hold with equality. The latter implies that the allocation is incentive compatible, so that it also solves the unrelaxed planning problem.

That any solution to this relaxed problem must have $R = 0$ follows directly from Lemma 1. Since $R = 0$, we now write $c(w^i) \equiv \gamma(y(w^i), Y(w^i), 0, w^i) = y(w^i)$. The following property will be used to establish that the downward incentive constraints (24) hold with equality.

Lemma 2. *Suppose $c' > c$ and $V(c', Y', 0, w') > V(c, Y, 0, w)$ for $w' > w$. Then (i) if $Y'/w' \geq Y/w$:*

$$V_y(c', Y', 0, w') < V_y(c, Y, 0, w); \quad (25)$$

otherwise (ii) if $Y'/w' < Y/w$ either (25) holds or:

$$V_Y(c', Y', 0, w') > V_Y(c, Y, 0, w). \quad (26)$$

Proof. Define $U^*(c, n) \equiv \max_m U(c, n, m) \equiv V(c, nw, 0, w)$. By the envelope condition, $V_y(c', Y', 0, w') < V_y(c, Y, 0, w)$ is equivalent to

$$U_c^*(c', n') < U_c^*(c, n) \quad (27)$$

And

$$U_n^*(c', n') > U_n^*(c, n) \quad (28)$$

implies $V_Y(c', Y', 0, w')w' > V_Y(c, Y, 0, w)w$, which in turn implies that $V_Y(c', Y', 0, w') > V_Y(c, Y, 0, w)$.

The hypothesis imply that

$$U^*(c', n') > U^*(c, n \frac{w}{w'}) \geq U^*(c, n).$$

For case (i) we have $n' = Y'/w' \geq Y/w = n$. Define the consumption compensation function $f(x)$ by $U^*(f(x), x) = U^*(c', n')$. Then $f(n) > c$, so that $U_c^*(c, n) > U_c^*(f(n), n)$ by concavity of $U^*(\cdot, n)$. Next note that

$$\begin{aligned} \frac{\partial}{\partial x}(U_c^*(f(x), x)) &= f'(x)U_{cc}^*(f(x), x) + U_{cn}^*(f(x), x) \\ &= -\frac{U_n^*(f(x), x)}{U_c^*(f(x), x)}U_{cc}^*(f(x), x) + U_{cn}^*(f(x), x) \leq 0, \end{aligned}$$

by Assumption 2. Thus, $U_c^*(f(x), x)$ is decreasing and

$$U_c^*(c, n) > U_c^*(f(n), n) > U_c^*(f(n'), n') = U_c^*(c', n'),$$

which establishes (27).

For case (ii) we have $n' = Y'/w' < Y/w = n$. Define the function $M(z) \equiv U^*(c + z(c' - c), n + z(n' - n))$. This function is strictly concave and differentiable so it follows that:

$$M'(1) - M'(0) = (U_c^*(c', n') - U_c^*(c, n))(c' - c) + (U_n^*(c', n') - U_n^*(c, n))(n' - n) < 0,$$

which implies (27) or (28). \square

Binding Downward Incentive Constraints. Next, suppose that the inequality (24) is strict for some i , so that

$$V(c(w^{i+1}), Y(w^{i+1}), 0, w^{i+1}) > V(c(w^i), Y(w^i), 0, w^{i+1}). \quad (29)$$

Then the Lemma applies with $w = w^i$ and $w' = w^{i+1}$. It is then possible to construct a feasible improvement as follows.

If inequality (25) holds then one can redistribute consumption from w^{j+1} to w^j and increase average welfare. That is, reducing $c(w^{j+1})$ and increasing $c(w^j)$ so that the resource constraint holds is feasible since the incentive constraint is slack: the strict inequality (29) will continue to hold for a small enough variation.

If, instead, inequality (26) holds then one can redistribute output from j to $j + 1$ and increase average welfare. That is, reducing $Y(w^j)$ (together with $Y(w^i)$ of any other individual type i with $Y(w^i) = Y(w^j)$) and increasing $Y(w^{j+1})$ (together with $Y(w^k)$ of any other individual type w^k with $Y(w^k) = Y(w^{j+1})$) so that the resource constraint holds is feasible since the incentive constraint is slack: the strict inequality (29) will continue to hold for a small enough variation.

Existence of a Maximum. This proves that if a maximum exists to the relaxed problem, at an optimum the downward incentive constraints hold with equality. Hence, the allocation is incentive compatible and it is also a solution to the unrelaxed planning problem. We now argue that, as long as the constraint set is nonempty so that there exists some feasible allocation, then a maximum does exist for the relaxed problem.

We have already argued that we can restrict ourselves to $R = 0$. We now argue that we can restrict ourselves to a compact set for $(y(w^i), Y(w^i))$. Both are nonnegative, so we seek upper bounds. We first derive an upper bound for $Y(w^i)$, then use this to derive an upper bound for $y(w^i)$.

Let \bar{U} denote the value for the planner's objective obtained for some feasible allocation. Then, in search of a maximum, we can restrict attention, without loss of generality, to allocations that provide at least this value for the objective. Downward incentive compatibility implies that utility is increasing in w . Thus, agents of type w^N must do better than the average, so that $V(y(w^N), Y(w^N), 0, w^N) \geq \bar{U}$. In addition, without loss of generality, we restrict attention to allocations with no distortion at the top (otherwise, by standard arguments, an improvement is possible): $-V_Y(y(w^N), Y(w^N), 0, w^N)/V_y(y(w^N), Y(w^N), 0, w^N) = 1$. Given Assumption 2, it then follows that there exists a $Y_{\max} < \infty$ such that $Y(w^N) \leq Y_{\max}$. By monotonicity, $Y(w^i) \leq Y_{\max}$.

Turning to the bound for $y(w^i)$. Note that $y(w^i) \geq 0$, so that from the resource constraint

$y(w^N) \leq (Y_{max} - G)/\pi^N$. Since $y(w^i) \leq y(w^N)$, this proves $y(w^i) \leq (Y_{max} - G)/\pi^N \equiv y_{max}$.

It follows that we can restrict $y(w^i)$ and $Y(w^i)$ to a compact set, implying that a maximum exists.

Increasing Taxes. At the optimum marginal tax rates are nonnegative. Otherwise an improvement is possible by decreasing output. Because the downward incentive constraints are binding, this implies that $T(Y)$ defined by (8) is nondecreasing.

Passage to the limit

We now return to the original problem with a continuum of types $w \in W$ distributed according to $F(w)$ and make a passage-to-the-limit argument to the continuum case. Since $F(w)$ is nondecreasing, it has at most countable jumps: $\tilde{w}_1, \tilde{w}_2, \dots$

Approximating with finite types. Take any feasible allocation $(y(w), Y(w), R)$. Without loss in generality, we assume this feasible plan yields some finite value for the utilitarian objective $\int V(y(w), Y(w), R, w) dF(w)$ (otherwise, it is trivial to find an improvement with $R^* = 0$).

Consider a partition of the interval based $\underline{w} = w_{N,0} \leq w_{N,1} \leq w_{N,2} \leq \dots \leq w_{N,2^N+K(N)} = \bar{w}$ composed of: $2^N + 1$ points $\underline{w} + j \frac{\bar{w} - \underline{w}}{2^N}$ for $j = 0, 1, \dots, 2^N$ and $K(N)$ points $\tilde{w}_1, \tilde{w}_2, \dots, \tilde{w}_{K(N)}$, where $K(N) = K$ if $F(w)$ has K jumps and $K(N) = N$ if $F(w)$ has a countably infinite number of jumps.

For any $w \in W$ we define $y_N(w) \equiv y(w_{N,i})$ and $Y_N(w) \equiv Y(w_{N,i})$ if $w \in (w_{N,i-1}, w_{N,i}]$ for $i \in I_N \equiv \{1, \dots, 2^N + K(N)\}$. Equivalently, defining the step function $\omega_N(w) = w_{N,i}$ if $w \in (w_{N,i-1}, w_{N,i}]$ we can write $y_N(w) = y(\omega_N(w))$ and $Y_N(w) = Y(\omega_N(w))$.

This construction guarantees that as $N \rightarrow \infty$ we have $y_N(w) \rightarrow y(w)$, $Y_N(w) \rightarrow Y(w)$ and $\omega_N(w) \rightarrow w$ almost everywhere with respect to the measure implied by $F(w)$. In addition $0 \leq y_N(w) \leq y_N(\bar{w}) = y(\bar{w})$, $0 \leq Y_N(w) \leq Y_N(\bar{w}) = Y(\bar{w})$. Thus, by Lebesgue's Dominated Convergence Theorem applied to the sequence $\{Y_N(w) - \gamma(y_N(w), Y_N(w), R, \omega_N(w))\}_{N=1}^\infty$ (for any N , these functions are step functions, so the integrals can be represented as finite sums)

$$\int_{\underline{w}}^{\bar{w}} (Y(w) - \gamma(y(w), Y(w), R, w)) dF(w) = \lim_{N \rightarrow \infty} \sum_{i \in I_N} (Y_N(w_{N,i}) - \gamma(y_N(w_{N,i}), Y_N(w_{N,i}), R, w_{N,i})) \Pi_N(w_{N,i}), \quad (30)$$

where $\Pi_N(w_{N,i}) \equiv F(w_{N,i}) - F(w_{N,i-1})$, i.e., the measure of the half-open interval $(w_{N,i-1}, w_{N,i}]$ (the set $\{w : \omega_N(w) = w_{N,i}\}$). Also, since $V(y, Y, R, \cdot)$ is an increasing function and due to in-

centive compatibility of the original allocation $V(y(w), Y(w), R, w) \leq V(y(w'), Y(w'), R, w')$ for any $w \leq w'$,

$$V(y(w), Y(w), R, w) \leq V(y_N(w), Y_N(w), R, \omega_N(w)) \leq V(y(\bar{w}), Y(\bar{w}), R, \bar{w}) < \infty,$$

and $V(y_{N+1}(w), Y_{N+1}(w), R, \omega_{N+1}(w)) \leq V(y_N(w), Y_N(w), R, \omega_N(w))$. Thus, by the Monotone Convergence Theorem applied to the sequence $\{-V(y_N(w), Y_N(w), R, \omega_N(w))\}_{N=1}^{\infty}$ (for any N , these functions are step functions, so the integrals can be represented as finite sums)

$$\int_{\underline{w}}^{\bar{w}} V(y(w), Y(w), R, w) dF(w) = \lim_{N \rightarrow \infty} \sum_{i \in I_N} V(y_N(w_{N,i}), Y_N(w_{N,i}), R, w_{N,i}) \Pi_N(w_{N,i}). \quad (31)$$

Improving with Finite Types. Now, for each N , interpret $\{Y_N(w_{N,i}), Y_N(w_{N,i})\}_{i \in I_N}$ and $R > 0$ as an allocation for a finite type economy with types $\{w_{N,i}\}_{i \in I_N}$, population fractions $\{\Pi_N(w_{N,i})\}_{i \in I_N}$ and government expenditures G_N . We can then apply the results for the finite type case to define a new allocation $\{\hat{y}_N(w_{N,i}), \hat{Y}_N(w_{N,i})\}_{i \in I_N}$ with $\hat{R} = 0$ satisfying

$$\begin{aligned} G_N &\equiv \sum_{i \in I_N} (Y_N(w_{N,i}) - \gamma(y_N(w_{N,i}), Y_N(w_{N,i}), R, w_{N,i})) \Pi_N(w_{N,i}) \\ &\leq \sum_{i \in I_N} (\hat{Y}_N(w_{N,i}) - \hat{y}_N(w_{N,i})) \Pi_N(w_{N,i}) \end{aligned} \quad (32)$$

$$\begin{aligned} \bar{U}_N &\equiv \sum_{i \in I_N} V(y_N(w_{N,i}), Y_N(w_{N,i}), R, w_{N,i}) \Pi_N(w_{N,i}) \\ &\leq \sum_{i \in I_N} V(\hat{y}_N(w_{N,i}), \hat{Y}_N(w_{N,i}), 0, w_{N,i}) \Pi_N(w_{N,i}) \end{aligned} \quad (33)$$

and

$$V(\hat{y}_N(w_{N,i}), \hat{Y}_N(w_{N,i}), 0, w_{N,i}) = V(\hat{y}_N(w_{N,i-1}), \hat{Y}_N(w_{N,i-1}), 0, w_{N,i}).$$

Thus, this new allocation improves welfare, total taxes receipts, and has binding downward incentive constraints. Furthermore, taxes $\hat{Y}_N(w_{N,i}) - \hat{y}_N(w_{N,i})$ are nondecreasing.

Converging to a Candidate. Next, we take the limit of this (fictitious) finite type economy to find a new candidate allocation for the (actual) continuum economy.

To ensure a limit exists, we first seek a uniform bound for $\hat{y}_N(w)$ and $\hat{Y}_N(w)$. For each N , define the average utility \bar{U}_N and tax collection G_N obtained by our finite approximation to the original allocation, as in (32) and (33). Then (30) and (31) imply that $\inf_N G_N > -\infty$

and $\inf_N \bar{U}_N > -\infty$. Now, for the upper bound on $\hat{Y}(w)$ we note that $\hat{Y}(w) \leq \hat{Y}(\bar{w})$ and that there is no distortion at the top: $-V_Y(\hat{y}_N(\bar{w}), \hat{Y}_N(\bar{w}), 0, \bar{w})/V_y(\hat{y}_N(\bar{w}), \hat{Y}_N(\bar{w}), 0, \bar{w}) = 1$. Combined with $V(\hat{y}_N(\bar{w}), \hat{Y}_N(\bar{w}), 0, \bar{w}) \geq \inf_N \bar{U}_N$ and Assumption 2 this implies that there exists a Y_{\max} such that $\hat{Y}_N(w) \leq Y_{\max} < \infty$ for all N . Since taxes are nondecreasing, the resource constraint requires that $\hat{Y}_N(\bar{w}) - \hat{y}_N(\bar{w}) \geq G_N$, so that $\hat{y}_N(\bar{w}) \leq Y_{\max} - \inf_N G_N < \infty$.

For any N , define an allocation for any $w \in W$ as: $\hat{y}_N(w) = \hat{y}_N(w_{N,i})$ and $\hat{Y}_N(w) = \hat{Y}_N(w_{N,i})$ if $w \in (w_{N,i-1}, w_{N,i}]$, i.e. $\hat{y}_N(w) = \hat{y}_N(\omega_N(w))$ and $\hat{Y}_N(w) = \hat{Y}_N(\omega_N(w))$. This gives a sequence of nondecreasing functions $\{\hat{y}_N(w), \hat{Y}_N(w)\}_{N=1}^{\infty}$ that is uniformly bounded. Helly's Selection Theorem implies that we can extract a subsequence $\{\hat{y}_{M(N)}(w), \hat{Y}_{M(N)}(w)\}_{N=1}^{\infty}$ that converges (everywhere) pointwise to some nondecreasing limit functions $y^*(w)$ and $Y^*(w)$.

Recall that, for any N , the allocation $(\hat{y}_{M(N)}(w), \hat{Y}_{M(N)}(w), \hat{R})$ is incentive compatible for the finite economy, that is, restricted to the partition points $\{w_{N,i}\}_{i \in I_N}$. Then, in the limit as $N \rightarrow \infty$, since the partition points $\{w_{N,i}\}_{i \in I_N}$ form a dense set for W , the limit allocation $(y^*(w), Y^*(w), R^*)$ with $R^* = 0$ is incentive compatible. Furthermore, the property that taxes are increasing and that marginal tax rates are nonnegative is also preserved in the limit: $Y^*(w) - y^*(w)$ is increasing in w and $-V_Y(y^*(w), Y^*(w), 0, w)/V_y(y^*(w), Y^*(w), 0, w) \leq 1$. Hence, this allocation can be implemented by an increasing tax schedule $T^*(Y)$ with $R^* = 0$. All that remains is to show that this allocation is an improvement over $(y(w), Y(w), R)$.

Because the functions involved are step functions, for any N the integral of $\hat{Y}_N(w) - \hat{y}_N(w)$ can be represented by the finite sum $\sum_{i \in I_N} (\hat{Y}_N(w_{N,i}) - \hat{y}_N(w_{N,i})) \Pi_N(w_{N,i})$. Applying Lebesgue's Dominated Convergence Theorem to the sequence $\{\hat{Y}_{M(N)}(w) - \hat{y}_{M(N)}(w)\}_{N=1}^{\infty}$:

$$\begin{aligned} \lim_{N \rightarrow \infty} \sum_{i \in I_{M(N)}} (\hat{Y}_{M(N)}(w_{M(N),i}) - \hat{y}_{M(N)}(w_{M(N),i})) \Pi_{M(N)}(w_{M(N),i}) \\ = \int_w^{\bar{w}} (Y^*(w) - y^*(w)) dF(w) \quad (34) \end{aligned}$$

Similarly, because $V(\hat{y}_N(w), \hat{Y}_N(w), 0, \omega_N(w))$ is a step function its integral can be represented by the finite sum $\sum_{i \in I_N} V(\hat{y}_N(w_{N,i}), \hat{Y}_N(w_{N,i}), 0, w_{N,i}) \Pi_N(w_{N,i})$. The function is also bounded above since

$$V(\hat{y}_N(w), \hat{Y}_N(w), 0, \omega_N(w)) \leq V(\hat{y}_N(\bar{w}), \hat{Y}_N(\bar{w}), 0, \bar{w}) \leq V(y_{\max}, 0, 0, \bar{w}) < \infty.$$

Thus, by Fatou's Lemma applied, to the sequence $\{-V(\hat{y}_{M(N)}(w), \hat{Y}_{M(N)}(w), 0, \omega_{M(N)}(w))\}_{N=1}^{\infty}$,

we obtain

$$\limsup_{N \rightarrow \infty} \sum_{i \in I_N} V(\hat{y}_{M(N)}(w_{M(N),i}), \hat{Y}_{M(N)}(w_{M(N),i}), 0, w_{M(N),i}) \Pi_N(w_{M(N),i}) \leq \int_{\underline{w}}^{\bar{w}} V(y^*(w), Y^*(w), R, w) dF(w). \quad (35)$$

Combining (30)–(35), gives

$$\int_{\underline{w}}^{\bar{w}} (Y(w) - \gamma(y(w), Y(w), R, w)) dF(w) \leq \int_{\underline{w}}^{\bar{w}} (Y^*(w) - y^*(w)) dF(w)$$

$$\int_{\underline{w}}^{\bar{w}} V(y(w), Y(w), R, w) dF(w) \leq \int_{\underline{w}}^{\bar{w}} V(y^*(w), Y^*(w), 0, w) dF(w).$$

Thus, we have constructed a feasible allocation $(y^*(w), Y^*(w), R^*)$ with $R^* = 0$ that is at least as good as the original allocation $(y(w), Y(w), R)$. Since the latter was arbitrary, it follows that $(y^*(w), Y^*(w), R^*)$ is optimal. This concludes the proof.

Welfare Costs for Shopping-Time Example

Facing $R = 0$ and a proportional tax $T(Y) = \bar{\tau}Y$ agents obtain utility

$$v^*(w) = \log \left(w(1 - \bar{\tau}) \frac{\alpha^\alpha}{(1 + \alpha)^{1+\alpha}} \right)$$

with $n^*(w) = Y^*(w)/w = 1/(\alpha + 1)$, $c^*(w) = (1 - \bar{\tau})w/(\alpha + 1)$ and $s^*(w) = 0$ (with $m^*(w) = \infty$). We now derive $V(y, Y, R, w)$ and $e(v, Y, R, w)$ for this specification. To preserve welfare we set $y(w; R) = e(v^*(w), Y(w; R), R, w)$, and solve

$$v^*(w) = V_w(e(v^*(w), Y(w), R, w), Y(w), R, w) \quad (36)$$

for $Y(w; R)$ in order to preserve incentive compatibility. We later verify that $Y(w; R)$ is increasing in w for all R .

For this specification we obtain

$$V(y, Y, R, w) = \log y + \log \left(\frac{\sigma(Y, R, w)k}{\sigma(Y, R, w)k + R} \right) + \alpha \log \left(1 - \frac{Y}{w} - \sigma(Y, R, w) \right)$$

with

$$m = \mu(y, Y, R, w) = \frac{1}{\sigma(Y, R, w)k + R}y \quad \text{and} \quad c = \gamma(y, Y, R, w) = \frac{\sigma(Y, R, w)k}{\sigma(Y, R, w)k + R}y$$

where

$$s = \sigma(Y, R, w) \equiv \frac{-R(1 + \alpha) + \sqrt{R^2(1 + \alpha)^2 + 4R\alpha k(1 - \frac{Y}{w})}}{2k\alpha} \quad (37)$$

The expenditure function (the inverse of V) is then

$$e(v, Y, R, w) \equiv \exp(v) \frac{\sigma(Y, R, w)k + R}{\sigma(Y, R, w)k} \left(1 - \frac{Y}{w} - \sigma(Y, R, w)\right)^{-\alpha}$$

Using (36) gives

$$\frac{Y(w; R)}{w} = \frac{1 - \sigma(Y(w; R), R, w)}{\alpha + 1}$$

Using this in (37), it follows that $\sigma(Y(w; R), R, w) \equiv S(R)$ is independent of w and is the largest root of the quadratic equation

$$\alpha k S(R)^2 + R(1 + \alpha)S(R) = R \left(1 - \frac{1 - S(R)}{\alpha + 1}\right).$$

Note that $Y(w; R) = w(1 - S(R))/(\alpha + 1)$ is strictly increasing in w .

Finally, using these expressions to compute net resources by

$$E(R) = \int \left(\gamma(e(v^*(w), Y(w; R), w, R), Y(w; R), w, R) - Y(w; R) \right) dF(w)$$

leads to

$$\frac{E(R) - E(0)}{\int c^*(w) dF(w)} = (1 - S(R))^{-\alpha} + \frac{S(R)}{1 - \bar{\tau}} - 1.$$

References

- Aiyagari, R., A. Braun, and Z. Eckstein, 1998. "Inflation, Transaction services and Welfare." *Journal of Political Economy*. 106: 2174–1301.
- Albanesi, Stefania, 2007. "Inflation and Inequality." *forthcoming in Journal of Monetary Economics*.
- Alvarez, Fernando, Patrick J. Kehoe, and Pablo Andrés Neumeyer, 2004. "The Time Consistency of Optimal Monetary and Fiscal Policies." *Econometrica*. 72 (2): 541–567.

- Atkinson, A.B., and J.E. Stiglitz, 1976. "The Design of Tax Structure: Direct vs. Indirect Taxation." *Journal of Public Economics*. 6: 55–75.
- Chari, V.V., Lawrence Christiano, and Patrick Kehoe, 1996. "Optimality of the Friedman Rule in Economies with Distorting Taxes." *Journal of Monetary Economics*. 37: 203–233.
- Correia, Isabel, and Pedro Teles, 1996. "Is the Friedman Rule Optimal when Money is an Intermediate Good?." *Journal of Monetary Economics*. 38: 223–244.
- Correia, Isabel, and Pedro Teles, 1999. "The Optimal Inflation Tax." *Review of Economic Dynamics*. 2: 325–246.
- Erosa, Andrés, and Gustavo Ventura, 2002. "On Inflation as a Regressive Consumption Tax." *Journal of Monetary Economics*. 49: 761–795.
- Friedman, Milton, 1969. *The Optimum Quantity of Money and Other Essays*, Aldine, Chicago, IL.
- Fudenberg, D., and J. Tirole, 1991. *Game Theory*, MIT Press.
- Guidotti, Pablo E., and Carlos A. Vegh, 1993. "The Optimal Inflation Tax when Money Reduces Transaction Costs." *Journal of Monetary Economics*. 31: 189–205.
- Hammond, Peter J, 1987. "Markets as Constraints: Multilateral Incentive Compatibility in Continuum Economies." *Review of Economic Studies*. 54 (3): 399–412.
- Kimbrough, Kent, 1976. "The Optimum Quantity of Money Rule in the Theory of Public Finance." *Journal of Monetary Economics*. 18: 277–284.
- Lucas, Robert E. Jr., 2000. "Inflation and Welfare." *Econometrica*. 68: 247–274.
- Lucas, Robert E. Jr., and Nancy L. Stokey, 1983. "Optimal Fiscal and Monetary Policy in an Economy without Capital." *Journal of Monetary Economics*. 12: 55–93.
- McCallum, Bennett T., and Marvin S. Goodfriend, 1987. *The New Palgrave: A Dictionary of Economics* . chap. Demand for Money: Theoretical Studies, pp. 775–781, London: Macmillan; New York: Stockton Press.
- Milgrom, Paul, and Ilya Segal, 2002. "Envelope Theorems for Arbitrary Choice Set." *Econometrica*. 70 (2): 583–601.
- Mirrlees, James A., 1971. "An Exploration in the Theory of Optimal Income Taxation." *Review of Economic Studies*. 38: 175–208.
- Mulligan, Casey B, and Xavier X Sala-i-Martin, 1997. "The Optimum Quantity of Money: Theory and Evidence." *Journal of Money, Credit and Banking*. 29 (4): 687–715.
- Mulligan, Casey B., and Xavier X. Sala-i-Martin, 2000. "Extensive margins and the demand for money at low interest rates." *Journal of Political Economy*. 108: 961–991.

Phelps, Edmund S., 1973. "Inflation in the Theory of Public Finance." *Swedish Journal of Economics*. 75: 67–82.

Prescott, Edward C., 1987. "A multiple means of payment model." in W.A. Barnett, and K. Singleton (ed.), *New Approaches to Monetary Economics*, Cambridge University Press, Cambridge.

Saez, Emmanuel, 2002. "The desirability of commodity taxation under non-linear income taxation and heterogeneous tastes." *Journal of Public Economics*. 83 (2): 217–230.

Werning, Iván, 2007. "Pareto Efficient Income Taxation." MIT working paper.

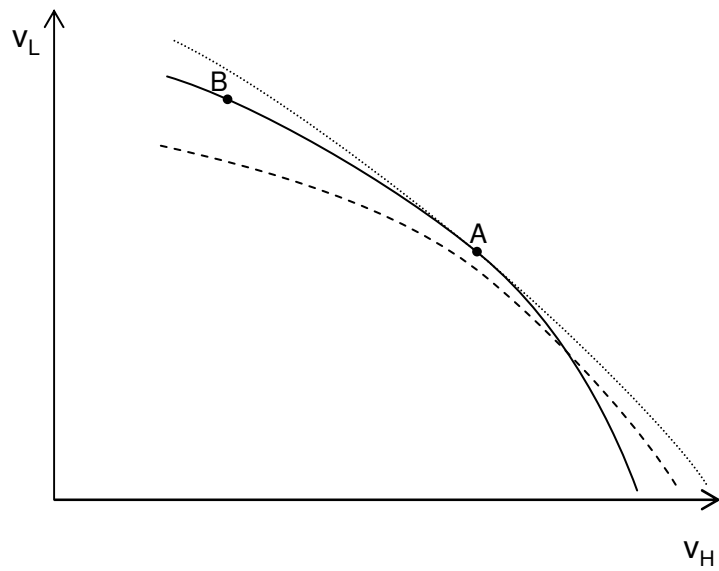


Figure 1: A region of the Pareto frontier for a case with two productivity types (only the region where $v_H > v_L$ is illustrated).

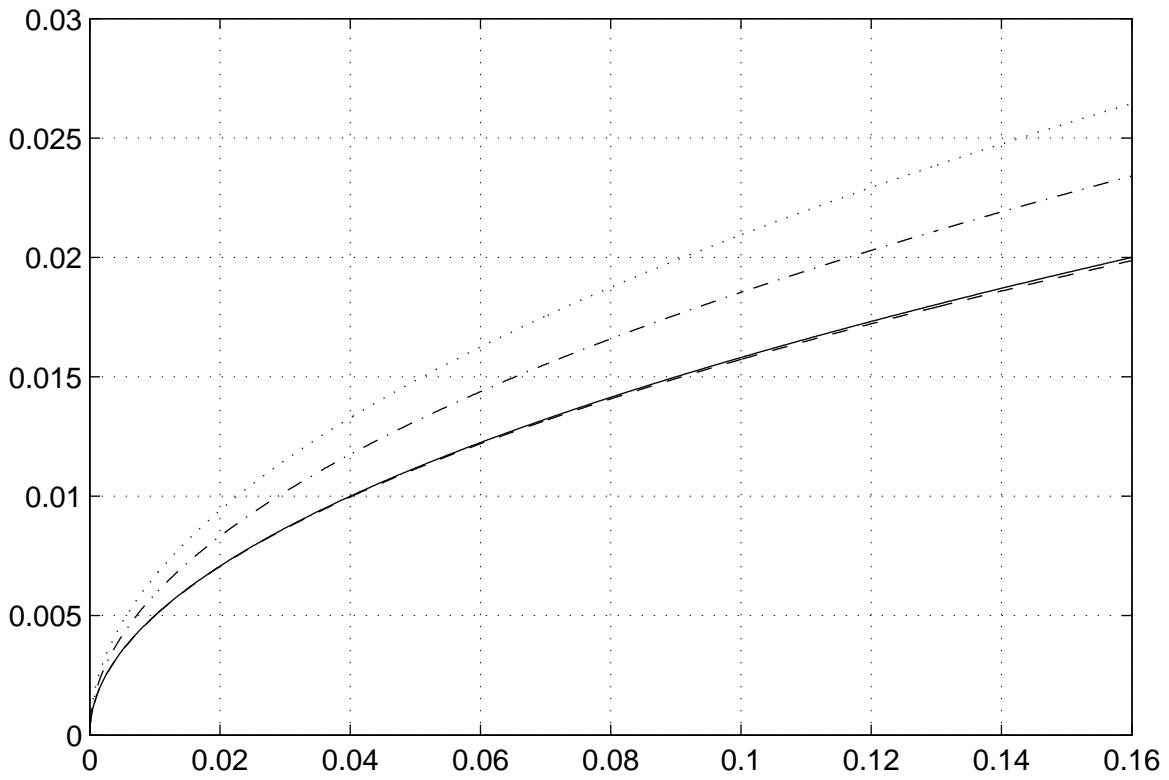


Figure 2: Welfare costs for Lucas (2000) shopping-time specification with $\tau = 0, 35\%, 50\%$ and area under demand curve.