

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

Carlos da Costa, Brazilian National Treasury<sup>†</sup>  
Iván Werning, MIT<sup>‡</sup>

February 11, 2003

## Abstract

We study the optimal “inflation tax” in an environment with heterogeneous agents and non-linear income taxes. We first derive the general conditions needed for the optimality of the Friedman rule in this setup. These general conditions are distinct in nature and more easily interpretable than those obtained in the literature with a representative agent and linear taxation. We then study two standard monetary specifications and derive their implications for the optimality of the Friedman rule. For the shopping-time model the Friedman rule is optimal with essentially no restrictions on preferences or transaction technologies. For the cash-credit model the Friedman rule is optimal if preferences are separable between the consumption goods and leisure, or if leisure shifts consumption towards the credit good. We also study a generalized model which nests both models as special cases.

**JEL classification:** E52, H21

**Keywords:** Friedman rule, optimal taxation, stationarity

---

\*We thank Fernando Alvarez, Luis Braido, Francisco Buera, Hugo Hopenhayn, Casey Mulligan and Pedro Teles for helpful comments and discussions.

<sup>†</sup>email: carlos.dacosta@fazenda.gov.br

<sup>‡</sup>e-mail: iwerning@MIT.edu

# 1 Introduction

Friedman (1969) clarified the point that positive nominal interest rates represent a tax on real money balances. In a first-best world prices should reflect the social costs of goods. Therefore, if the social costs of money are negligible, he argued, the nominal interest rate should be set to zero – the Friedman rule – with deflation generally required to implement this. Perhaps it is because these prescriptions are so far from current and past monetary practice that the Friedman rule has generated such enormous controversy and debate.

Phelps (1973) argued that in a second-best world many distortions are tolerated and taxes levied on most goods. What is so special about money? Why should money be treated differently? He concluded that money should generally be taxed, that nominal interest rates should be positive.

Recently, many authors have explored the special conditions under which the Friedman rule remains optimal [Chari, Christiano and Kehoe (1995), Correia and Teles (1995, 1999), Lucas and Stokey (1983), Kimbrough (1986), Guidotti and Vegh (1993) and Mulligan and Sala-i-Martin (1997)]. Restrictions on preferences and transaction technologies – which amount to assuming that money *is* special – can imply the optimality of Friedman’s rule even in a second-best world.

All these exercises study the optimal “inflation tax” as the solution to a Ramsey (1927) tax problem in a representative agent economy. The main advantage of both assumptions is the tractability they provide. However, the exercises thus ignore distributional effects and restrict the available instruments to linear taxes. As is well known in the public finance literature, tax prescriptions may be sensitive to both assumptions.

More importantly, however, this approach suffers from an inherent contradiction. The arbitrary restrictions on available tax instruments – that rule out lump sum taxation, for example – presumably capture, in a loose way, the undesirability of tax systems that are “too regressive”. Yet, the solutions to most tax problems posed in this way yield tax prescriptions that attempt to emulate – however imperfectly – the missing lump-sum tax. As a result, the prescribed tax system may also be “too regressive”. Thus, there is an important contradiction in this approach: lump-sum taxes are ruled out as *undesirable* only to derive supposedly *optimal* tax prescriptions that imitate lump-sum taxation. Why should the resulting “regressive” tax system based on commodity taxes be any better than one generated by a lump-sum tax?

In this paper we re-examine the optimal “inflation tax” modelling agent

heterogeneity, in the form of productivity differences, explicitly and allowing for non-linear income taxation, following Mirrlees' (1971, 1976) framework. This approach does not restrict the set of tax instruments arbitrarily. Instead, it models the potential conflict between redistribution and efficiency directly with the government restricted only by the informational asymmetry of the economic environment.

We first derive the general conditions for the optimality of the Friedman rule in such a setup. We then examine two popular specifications: the shopping-time model [McCallum and Goodfriend (1987)] and the cash-credit model [Lucas and Stokey (1983)]. Somewhat surprisingly, for the shopping-time model the Friedman rule is optimal with essentially no restrictions on preferences or transaction technologies. For the cash-credit model the Friedman rule is optimal if preferences are separable between the consumption goods and leisure, or if leisure shifts consumption towards the credit good. We also study a generalized model which nests both models as special cases.

Understanding these results requires understanding the accessory role played by the inflation tax when a non-linear income tax is present. A positive inflation tax is useful only when it relaxes the incentive constraints imposed by the asymmetric information. This happens if the demand for money conditional on expenditures and production is greater for higher productivity agents. Thus, consider a high productivity agent that shirks by choosing the bundle of expenditure and production intended for a lower productivity agent. Taxing money in such a case relaxes the incentive constraints because it has a higher incidence on the shirking agent than on the lower productivity agent.

In the shopping-time model it is optimal to subsidize money because the model implies exactly the reverse: higher productivity agents have a lower demand from money once we condition on expenditures and production. This is so because a higher productivity agent produces any given amount with less time, and so has more time left over for shopping.

In the cash-credit model if preferences over cash and credit goods are separable, the demand for money does not depend on the level of productivity, once we condition on expenditure and output. Thus it is optimal to neither tax nor subsidize money in this case.

The rest of the paper is organized as follows. Section 2 lays out the economic environment and states the taxation problem we examine. Section 3 re-expresses the government's problem in a more tractable way and derives the optimality condition for the nominal interest rate. Using this condition

section 4 examines the optimality of the Friedman rule. A proposition characterizing the general conditions for the Friedman rule in terms of demand elasticities is presented. Section 5 uses this proposition to study two popular specifications. Section 6 considers two extensions: multiple consumption goods and consumption taxes paid with money. Section 7 concludes. The appendix contain proofs and an extension to allow random taxation.

## 2 The Model

In this section we present the environment and describe the tax instruments to give a preliminary statement of the tax problem.

### 2.1 Environment

We state and solve the full-commitment taxation problem in a general setup with money directly in the utility function. That is, utility is defined over consumption,  $c$ , real money holding,  $m$ , and non-work time,  $l$  :  $u(c, m, l)$ . As a reduced form this setup nests many specific models of money used in the literature. After studying this general setup we turn to some popular specifications of  $u(c, m, l)$ .

We study the canonical optimal-taxation setup where agents differ only in their productivity levels. Consequently, we index agents by their productivity level,  $w$ , and assume its distribution is represented by the density  $f(w)$  for  $w \in [0, \bar{w}]$ . The productivity of each agent is private information; only the distribution of productivities is known by the government.

Production is linear: a unit of  $c$  requires one efficiency-unit of labor, while  $\bar{R}$  units of labor are required to produce a unit of  $m$ . That is  $h_c = c$  and  $h_m = \bar{R}m$  where  $h_i$  are the efficiency units of labor used by sector  $i$ .

Government revenue is provided by two sources: the non-linear income tax and seigniorage – the tax on real money balances; we introduce consumption taxes in section 6.2.

Most of the Ramsey literature has focused attention on the case where  $\bar{R} = 0$ . We allow  $\bar{R}$  to be positive because we wish to investigate whether the optimality of the Friedman rule in our setup depends critically on money being a free good. Under certain conditions in a Ramsey setting Correia and Teles (1996, 1999) have shown that  $\bar{R} = 0$  plays a role in evaluating the optimality of the Friedman rule.

## 2.2 Agents

Preferences are defined over sequences of  $(c, m, l)$  by the utility function,

$$U[\{(c_t, m_t, l_t)\}_{t=0}^{\infty}] = \sum_{t=0}^{\infty} \beta^t u(c_t, m_t, l_t), \quad (1)$$

where  $m_t = M_t/p_t$  represents real balances. The budget constraint in nominal (dollar) units in period  $t$  is,

$$p_t c_t + M_{t+1} + q_{t+1} B_{t+1} \leq p_t y_t + M_t + B_t.$$

Here  $B_t$  represents the amount of one-period nominal bonds purchased in period  $t-1$  at price  $q_t$ , and  $y_t$  denotes after-tax income in period  $t$ . In period  $t = -1$  we think of agents as selecting their initial holdings of bonds and money with no consumption taking place:<sup>1</sup>

$$M_0 + q_0 B_0 \leq 0$$

Provided the standard no-Ponzi condition is imposed, we can recursively solve and substitute for  $B_t$  to obtain the present value version,

$$\sum_{t=0}^{\infty} \psi_t [c_t + R_t m_t - y_t] \leq 0 \quad (2)$$

$$R_t \equiv \frac{1 - q_t}{q_t}$$

where  $\psi_t \equiv \prod_{s=0}^t q_s p_t / p_0$  denotes the real price of consumption in period  $t$ . Equation (2) shows that the real cost of holding real balances  $m_t$  relative to  $c_t$  is the nominal interest rate,  $R_t$ . From this expression  $R_t \geq 0$  is a necessary condition for equilibrium.

## 2.3 Tax System

To introduce the non-linear income tax we let  $y_t = y_t(w(1 - l_t))$  in all the budget constraints above, where  $y_t(\cdot)$  represents the net of tax income as a

---

<sup>1</sup>Because agents are starting their dynastic life we take the initial nominal wealth to be zero. This assumption also avoids adding another dimension of agent heterogeneity.

function of gross pre-tax income. That is,  $y_t(Y) \equiv Y - T_t(Y)$ , where  $Y$  is pre-tax income and the function  $T_t(Y)$  represents the non-linear income tax schedule.

Consider for a moment stationary policies and allocations. That is, assume  $y_t(\cdot) = y(\cdot)$ ,  $R_t = R$  and  $\psi_t = \beta^t$ . Although the agent's problem is dynamic, we can characterize the solution in this case by studying an associated static problem. To see this, consider the agent's interior first order conditions for the maximization (1) subject to (2):

$$\begin{aligned} u_c(c_t, m_t, l_t) &= \lambda \\ u_m(c_t, m_t, l_t) &= R\lambda \\ u_l(c_t, m_t, l_t) &= wy'(w(1-l_t))\lambda \end{aligned}$$

These equations do not depend directly on  $t$ . Hence, provided they have a unique solution we have  $(c_t, m_t, l_t) = (c, m, l)$  for all  $t$ , for some  $(c, m, l)$ .<sup>2</sup> Furthermore, we can find  $(c, m, l)$  by solving:

$$\max_{c, m, l} u(c, m, l) \tag{3}$$

$$c + Rm \leq y(w(1-l)). \tag{4}$$

The solution to this problem yields the following functions:  $c^*(w, R; y)$ ,  $m^*(w, R; y)$  and  $l^*(w, R; y)$ . To summarize, if the government adopts a stationary policy we can characterize agent behavior from a related static problem.

In the case of stationary policies, the government's budget constraint requires tax revenues to meet expenditures,  $G$ ,

$$\begin{aligned} \int [w(1-l^*(w, R; y)) - y(w(1-l^*(w, R; y)))] \\ +(R - \bar{R})m^*(w, R; y)]f(w)dw = G. \end{aligned} \tag{5}$$

---

<sup>2</sup>Because of non-linear taxation the agent's problem need not be convex, even if  $u(c, m, l)$  is assumed strictly concave. Consequently, stationary solutions to the agent's problem cannot be guaranteed. In this discussion, for simplicity, we assume that the agent's problem does yield stationary solutions in this case. In fact, implementing devices other than the per-period income tax we describe can ensure the agent's solution to be stationary (e.g. an income tax based on life-time earnings).

Substituting the agent's budget constraint (4) into (5) one obtains the economy's resource constraint,

$$\int [w(1 - l^*(w, R; y)) - c^*(w, R; y) - \bar{R}m^*(w, R; y)] f(w)dw = G, \quad (6)$$

which is more convenient than (5) for our purposes.

The government sets  $y(w)$  and  $R$  to solve:

$$\max_{y(w), R \geq 0} \int u(c^*(w, R; y), m^*(w, R; y), l^*(w, R; y)) f(w)dw$$

subject to the resource constraint (6).

While simple to state, this direct approach is not very tractable to solve. The difficulty lies in the fact that the agent's maximization is introduced by using the agent's demand functions. Because of the non-linear income tax, these demand functions are complicated objects that depend on the entire  $y(\cdot)$  function. Also, this analysis does not make clear whether the assumption of a stationary policy was without loss in generality.

Following Mirrlees (1976), we reformulate the problem in a way that makes the incorporation of agent's maximization more manageable and allows government policy to be constrained only by the informational structure of the environment.

### 3 Optimal Taxation

The novelty of Mirrlees' approach was to consider an environment where the first best is not attainable because of informational asymmetries, and to consider the optimal mechanisms that attains the constrained pareto frontier. Once these mechanisms are characterized, they can be mapped into the tax system that implements the allocation.

Recall that the environment we consider is stationary. In the main text we characterize the mechanism by restricting it to be stationary which reduces the notational burden significantly. In appendix *B* we show that the optimal unrestricted mechanism is stationary, provided the problem is sufficiently convex<sup>3</sup>.

---

<sup>3</sup>Non-stationary allocations may play the role of mimicking random allocations. In fact, if lotteries are allowed the optimal mechanism is, without loss in generality, station-

We first need to introduce some notation. Define the indirect utility function as,

$$V(y, Y, R, w) \equiv \max_{c, m} u(c, m, 1 - Y/w) \quad (7)$$

$$c + Rm = y,$$

where  $Y$  and  $y$  represent pre-tax and after-tax income, respectively. Denote the uncompensated demands by  $c(y, Y, R, w)$  and  $m(y, Y, R, w)$ . Likewise, define the expenditure function as,

$$e(v, Y, R, w) \equiv \min_{c, m} [c + Rm] \quad (8)$$

$$v = u(c, m, 1 - Y/w),$$

with corresponding compensated demands:  $c^c(v, Y, R, w)$  and  $m^c(v, Y, R, w)$ . It is worth emphasizing that both  $V(y, Y, R, w)$  and  $e(v, Y, R, w)$ , and their corresponding demand functions, are conditioned on output,  $Y$ , and productivity,  $w$ , and, therefore, also on the amount of work time,  $l$ .

We introduce a standard assumption on preferences: the *single-crossing* (SC) assumption. The assumption ensures that abler agents choose to produce more. It also plays a key role in simplifying the second order conditions of the agent's maximization. Define the marginal rate of substitution,

$$s(y, Y, R, w) = \frac{V_y(y, Y, R, w)}{V_Y(y, Y, R, w)}.$$

The assumption is that  $s(y, Y, R, w)$  is increasing in  $w$ .

**SC condition:**  $s_w(y, Y, R, w) > 0$  for all  $(y, Y, R, w) \in \mathbb{R}_+^4$ .

This condition implies that agents with higher productivities have flatter indifference curves in the  $(y, Y)$  plane. Throughout the paper we assume SC holds. Some rather weak conditions imply SC (for example, joint normality of  $c$  and  $m$ ) and we do not impose such conditions at a later stage for other reasons (Assumption A).

In the direct mechanism agents report a  $w$  and obtain an allocation as a function of this report:  $(y(w), Y(w))$ . Implicitly, the government's choice of

---

ary. Furthermore, if lotteries are allowed but non-optimal then the optimal allocation is guaranteed to be stationary. Appendix C considers random mechanisms and shows that the results regarding the inflation tax hold in this case as well.

$y(w)$  and  $Y(w)$  summarize the income-tax schedule faced by agents. That is, the budget set available to agents is,

$$B = \{(y, Y) : \exists w' \ y = y(w') \ Y = Y(w')\}.$$

This set can be implemented by a non-linear tax system  $y(Y)$  that corresponds to the frontier of  $B$ . It is in this sense that solving for the optimal mechanism yields a solution of the tax system.

Let  $v(w)$  denote the utility attained by agent- $w$  under the mechanism, we shall setup the problem so that  $v(w)$  is the state variable. Agents choose their report to maximize utility:

$$v(w) \equiv \max_{w'} V(y(w'), Y(w'), R, w).$$

Agents thus truthfully reveal their productivity if and only if,

$$w = \arg \max_{w'} V(y(w'), Y(w'), R, w). \quad (9)$$

By the envelope condition,

$$v'(w) = V_w(y(w), Y(w), R, w). \quad (10)$$

Therefore, (10) is a necessary condition imposed on the government from the agent maximization problem (9).

The maximization in (9) also imposes a local second order condition. It can be shown [see Mirrlees (1976)] that, given the SC condition, the second order condition is equivalent to  $y'(w) \geq 0$  and  $Y'(w) \geq 0$ . Of course, given (10), it is only necessary to impose one of these two conditions. If these constraints bind over an interval,  $[w_0, w_1]$ , agents of different productivities end-up choosing the same bundle,  $(y(w), Y(w)) = (y^*, Y^*)$  for  $w \in [w_0, w_1]$ . This situation is known in the literature as “bunching”. We take full account of the possibility of bunching by imposing the condition that  $Y'(w) \geq 0$  on the government problem below.

For agent- $w$  is to attain utility level  $v(w)$  while producing  $Y(w)$ , his disposable income must be  $y(w) = e(v(w), Y(w), R, w)$ . Therefore, we can think of the government as choosing  $v(w)$  and  $Y(w)$  instead of  $y(w)$  and  $Y(w)$ , leave  $y(w)$  in the background.

We can now restate the government’s problem as:

$$\max_{v(w), Y(w), R \geq 0} \int v(w) f(w) dw, \quad (11)$$

subject to the incentive constraint,

$$v'(w) = V_w(e(v(w), Y(w), R, w), Y(w), R, w), \quad (12)$$

the resource constraint,

$$\int [Y(w) - c^c(v, Y, R, w) - \bar{R}m^c(v, Y, R, w)] f(w)dw = G \quad (13)$$

and the second order condition for agent maximization,

$$Y'(w) \geq 0. \quad (14)$$

This is an optimal control problem with state variables  $v(w)$  and  $Y(w)$ , and control variable  $Y'(w)$  with the additional scalar choice variable  $R$ . As such it can be approached using Pontryagin's principles. However, because our main interest is in the optimality condition for the static variable  $R$ , it is more convenient to work with the conventional Lagrangian,

$$\mathcal{L} = \int \{vf + \mu[v' - V_w] + \lambda(Y - c^c - \bar{R}m^c) f + \phi Y'\} dw$$

(to simplify the notation, whenever obvious, we omit the arguments of functions). Integrating  $\int \mu v' dw$  and  $\int \phi Y' dw$  by parts,

$$\begin{aligned} \mathcal{L} = & \int \{vf - \mu'v - \mu V_w + \lambda(Y - c^c - \bar{R}m^c) f - \phi'Y\} dw \\ & + \mu(\bar{w})v(\bar{w}) - \mu(0)v(0) + \phi(\bar{w})Y(\bar{w}) - \phi(0)Y(0) \end{aligned}$$

Differentiating with respect to  $R$ :

$$\int \mu (V_{wR} + V_{wy}e_R) dw + \lambda \int (c_R^c + \bar{R}m_R^c) f dw = 0. \quad (15)$$

To simplify equation (15) note that by the usual properties of conditional demands,

$$c_R^c + \bar{R}m_R^c = c_R^c + Rm_R^c + (\bar{R} - R)m_R^c = (\bar{R} - R)m_R^c, \quad (16)$$

and that differentiating Roy's identity,  $V_R + V_y m = 0$ , yields:

$$V_{wR} + V_{wy}m = -V_y m_w. \quad (17)$$

Substituting (16) and (17) into (15) and rearranging we obtain:

$$(R - \bar{R}) \int m_R^c f dw = - \int \left( \frac{V_y \mu}{\lambda} \right) m_w dw. \quad (18)$$

This is the fundamental optimality condition required to evaluate the desirability of the Friedman rule in the presence of non-linear income taxes.

In our setup, with positive costs for money, we associate the Friedman rule more broadly as prescribing  $R \leq \bar{R}$ . That is, money balances should not be taxed, if anything they should be subsidized. It is important to determine the signs of the various terms in equation (18). It is easily seen that:

$$m_R^c < 0, V_y > 0 \text{ and } \lambda > 0.$$

Understanding the sign of  $\mu(w)$  requires more discussion. Loosely speaking,  $\mu(w) > 0$  means that the binding incentive constraint deviation is that of under-reporting: of agent  $w + dw$  potentially misreporting to be  $w$  (for  $dw > 0$ ). This is generally true if society at the optimum desires to redistribute income from higher to lower productivity agents. In fact, under the utilitarian welfare function this is precisely the case.

Moreover, where no bunching occurs, the sign of  $\mu(w)$  equals the sign of the marginal tax on agent  $w$  so that non-negative marginal tax rates require  $\mu(w) \geq 0$ . In the context of the two-good case ignoring bunching, Seade (1982) was the first to show that under weak assumptions (normality of leisure and the single crossing condition)  $\mu(w) \geq 0$  for all  $w$  holds, ensuring non-negative marginal tax rates. Ebert (1992) and Brunner (1993, 1990) have since extended Seade's analysis to incorporate the possibility of bunching.

For  $\mu(w) \geq 0$ , then, the sign of  $(R - \bar{R})$  in (18) depends crucially on the sign of  $m_w$ . In particular, if  $m_w \leq 0$ , for all  $w$ , it is optimal to set  $R \leq \bar{R}$ , that is, the Friedman rule is optimal.<sup>4</sup> We state this important result as a proposition.

**Proposition 1** *For  $\mu(w) \geq 0$  if  $m_w(y, Y, R, w) \leq 0$  for  $(y, Y, R, w) \in \mathbb{R}_+^4$  then at the optimum  $R \leq \bar{R}$ . Furthermore, if  $m_w(y, Y, R, w) < 0$  over a positive measure of equilibrium values of  $(y, Y, R, w)$  then at the optimum  $R < \bar{R}$ .*

---

<sup>4</sup>Because the sign of  $R - \bar{R}$  equals the sign of the integral on the right hand side of (19) we do not require  $m_w \leq 0$  for all  $w$ . That is, this condition is not necessary for  $R \leq \bar{R}$ , it is sufficient. However, when the sign of  $m_w$  is independent of where it is being evaluated our results are robust to the particular distribution of  $f(w)$  and other fundamentals.

Note that the proposition does not depend in any special way on the assumption of  $\bar{R} = 0$ . This is important because Correia and Teles (1996, 1999), in a Ramsey setting, have shown that the free-good aspect of money may play a role for the optimality of the Friedman rule. This not true in our setup, the (potential) free-good aspect of money is of no importance in evaluating the optimality of the Friedman rule in our setup.

## 4 Interpretation

We showed above that it is optimal to set  $R \leq \bar{R}$  whenever  $m_w \leq 0$ , for all  $w$ . We now discuss the economic interpretation of  $m_w \leq 0$  and explore the conditions in terms of standard demand elasticities under which it obtains.

In equilibrium money holdings by agent- $w$  are given by,

$$\tilde{m}(w) \equiv m(y(w), Y(w), R, w).$$

It is worth emphasizing then that  $m_w$  represents the *partial* derivative of  $m(y(w), Y(w), R, w)$  with respect to  $w$  and not  $\tilde{m}'(w)$ . In fact, as we shall see, in most cases  $\tilde{m}'(w) > 0$  while  $m_w < 0$ . The correct interpretation of  $m_w$  is that, for  $dw > 0$ ,  $m_w dw$  represents the change in money demand that *would* result from agent  $w + dw$  choosing to under-report his productivity as being  $w$ .

An agent under-reporting his productivity has more non-work time than the agents he impersonates – he produces the same output,  $Y$ , with higher productivity,  $w$ . It is the effect of this extra time that determines the sign of  $m_w$  and, consequently, the optimality of the Friedman rule.

We next characterize the conditions for  $m_w \leq 0$  in terms of demand properties. The money demand  $m(y, Y, R, w)$  obtained from (7) solves the agent's problem conditioning on non-labor time (since it conditions on both  $Y$  and  $w$ ). Consider the following unrestricted problem with full income,  $I$  :

$$\max_{c, m, l} u(c, m, l) \tag{19}$$

$$c + Rm + wl = I$$

with demands  $\tilde{c}(R, w, I)$ ,  $\tilde{m}(R, w, I)$  and  $\tilde{l}(R, w, I)$ . Let  $\eta$  and  $\varepsilon$  represent income and price elasticities, respectively. We will use the following weak assumptions on the unrestricted demands.

**Assumption A:**  $s_c\eta_c + s_m\eta_m > 0$  (i.e.  $\eta_l < 1/s_l$ ) where  $s_i$  are the share of expenditure for good  $i$ .

Appendix A proves the following proposition.

**Proposition 2** Under assumption A,  $m_w(y, Y, R, w) \leq 0$  for  $(y, Y, R, w) \in \mathbb{R}_+^4$  if and only if,

$$\frac{\eta_c}{[s_c\eta_c + s_m\eta_m]} \geq \frac{\varepsilon_{cw}}{[s_c\varepsilon_{cw} + s_m\varepsilon_{mw}]} \quad (20)$$

or the equivalent condition,

$$\eta_c\varepsilon_{mw} \geq \eta_m\varepsilon_{cw}, \quad (21)$$

where  $\varepsilon_{iw}$  and  $\eta_i$  represent the elasticity of good  $i$  with respect to  $w$  (Hicksian or Marshallian) and income, respectively.

To understand the roles played by the elasticities in condition (20) it is useful to perform the following thought experiment, which mimics the steps found in the proof. The extra non-work time made available from the increase in  $w$  can be viewed as a reduction in the shadow price of non-work time,  $w^*$ , and a decrease in virtual income,  $I^*$ . The shadow price of leisure must fall because the composite commodity, of  $c$  and  $m$ , is normal. Virtual income must then rise to keep the desired expenditures on  $c$  and  $m$  constant – as  $m(y, Y, R, w)$  does. The combined decrease in  $w^*$  and decrease in  $I^*$  leads consumption to shift towards  $c$ , away from  $m$ , the greater  $\varepsilon_{mw}$  and  $\eta_c$  are, relative to  $\varepsilon_{cw}$  and  $\eta_m$ .

An equivalent condition for  $m_w \leq 0$  is that the increase in  $l$  lowers the marginal rate of substitution between  $c$  and  $m$ . Conditions like these are common in the optimal taxation literature [Atkinson and Stiglitz (1976)]. Proposition 2 has the virtue of being expressed in terms of demand properties only. Furthermore, proposition 2 can be generalized to  $n$  consumption goods while the statement on marginal rates of substitution cannot<sup>5</sup>.

---

<sup>5</sup>The condition on marginal rates of substitution continues to be useful with more goods if all these goods can be taxed in a joint-non-linear fashion. See Atkinson and Stiglitz (1976) and Mirlees (1976).

## 5 Specific Models

We next consider two specific models of money used in the literature and examine their implications for the optimality of the Friedman rule. The first model we consider is the shopping-time model of McCallum and Goodfriend (1987). The second model is the cash-credit model introduced by Lucas and Stokey (1983).

### 5.1 Shopping-Time Model

In this model agent's basic preferences defined over consumption and leisure only,  $U(c, L)$ . Purchases require time and money economizes on this time. The shopping-time required for real purchases  $c$  using real balances  $m$  is given by  $v(c, m)$ , with  $v_c \geq 0$  and  $v_m \leq 0$ . To map this into our general setup define the utility function over  $(c, m, l)$  to be:

$$u(c, m, l) \equiv U(c, l - v(c, m))$$

To prove the next proposition we use the weak assumption that consumption is a normal good.

**Assumption B**  $\eta_c \geq 0$ .

**Proposition 3** *Under assumptions A and B, in the shopping-time model we have that  $m_w(y, Y, R, w) \leq 0$  for all  $(y, Y, R, w) \in \mathbb{R}_+^4$ .*

The proof relies on checking the conditions for proposition 2 as follows. Consider the problem of choosing  $m$  and  $l$  subject to  $c$  and  $I$  :

$$V(c, I, R, w) \equiv \max_{l, m} U(c, l - v(c, m))$$

$$c + Rm + wl = I$$

with demands:  $m^*(c, I, R, w)$  and  $l^*(c, I, R, w)$ . The maximization of  $m$  implies,

$$-v_m(c, m) = R/w$$

(the second order condition requires  $v_{mm} \geq 0$ ). From this first order condition one can prove that  $m_I^* = 0$  and  $m_w^* > 0$ , or in elasticity form that:  $\eta_m^* = 0$  and  $\varepsilon_{mw}^* > 0$ .

The conditional demand for money,  $m^*$ , is related to the unconditional demands,  $\tilde{m}$  and  $\tilde{c}$ , from the previous section by:

$$\tilde{m}(R, w, I) \equiv m^*(\tilde{c}(R, w, I), I, R, w).$$

Then the elasticities satisfy the following relationship:

$$\eta_m = \varepsilon_{mc}^* \eta_c$$

$$\varepsilon_{mw} = \varepsilon_{mc}^* \varepsilon_{cw} + \varepsilon_{mw}^*$$

Substituting into equation (21),

$$\eta_c (\varepsilon_{mc}^* \varepsilon_{cw} + \varepsilon_{mw}^*) \geq (\eta_c \varepsilon_{mc}^*) \varepsilon_{cw},$$

which after canceling implies,

$$\eta_c \varepsilon_{mw}^* > 0$$

which is satisfied because  $\eta_c > 0$  by assumption  $B$  and  $\varepsilon_{mw}^* > 0$  is implied by the shopping-time model as seen above.

Somewhat surprisingly, no restrictions need to be placed on the shopping technology,  $v(c, m)$ . In contrast, the representative agent Ramsey literature requires restrictions such as the homogeneity of  $v(c, m)$ . The intuition for this result is that the STM imposes a special structure on preferences. In this model  $m_w \leq 0$  because additional leisure lowers the need for time-saving money. This result is interesting in light of the wide popularity of shopping-time models for studies of money demand and inflation [see Lucas (2000) and the references therein].

## 5.2 Cash-Credit Model

In this model agents preferences are defined over two goods and leisure:  $\hat{u}(c_1, c_2, l)$ . The credit-good,  $c_1$ , can be purchased on credit while the cash-good,  $c_2$ , requires money up-front and thus  $c_2 \leq m$ .

The cash-credit model can be mapped into the reduced form model of money as follows. Let  $c \equiv c_1 + c_2$ , preferences over  $c, m, l$  are then given by  $u(c, m, l) \equiv \hat{u}(c - m, m, l)$ . Clearly in this way, the cash-credit good, unlike the shopping-time model, does not necessarily impose any restrictions on  $u(c, m, l)$ .

Consider the important case where  $c_1$  and  $c_2$  are separable from  $l$ . A direct application of propositions 2 and 3 implies that it is optimal to set  $R = \bar{R}$  in this case - the Friedman rule holds with equality.

**Proposition 4** *If  $\hat{u}$  displays separability between  $(c_1, c_2)$  and  $l$ ,*

$$\hat{u}(h(c_1, c_2), l),$$

*then  $m_w(y, Y, R, w) = 0$  for  $(y, Y, R, w) \in \mathbb{R}_+^4$ .*

The result follows by noting that separability implies that the marginal rates of substitution between  $c$  and  $m$  are unaffected by  $l$  or  $w$  so that  $m_w = 0$ . Without separability, the Friedman rule holds, with  $R < \bar{R}$ , as long as leisure shifts consumption towards the credit good.

With separability, a representative-agent Ramsey analysis requires further restrictions on  $h(\cdot)$ . In particular, when  $\bar{R} > 0$ , for  $R = \bar{R}$  to be optimal, homotheticity of  $h(\cdot)$  is required. For  $R \leq \bar{R}$  to be optimal what is required is that the income elasticity of the cash good be weakly greater than that of the credit good.

## 6 Welfare Costs of Inflation

This section, which draws on Werning (2000), reports numerical calculations of the welfare costs of departing from the Friedman rule, when this rule is in fact optimal. For purposes of comparison, in calibrating we follow Lucas (2000) as closely as possible.

When comparing different levels of nominal interest rates our welfare measure will be the resources in the hands of the government. That is, we ask the question: for any  $R > 0$  what additional resources (as a fraction of total output) in the hands of the government are required to obtain the same total welfare level as when  $R = 0$ ?

Based on Lucas' (2000) analysis, we consider the following shopping-time model specification of preferences:

$$u(c, m, n) = \log c + \gamma \log \left( 1 - n - \frac{1}{k} \frac{c}{m} \right)$$

here  $s = v(c, m) = c/mk$  is the shopping-time function. In the computations that follow we set  $k = 800$  and  $\gamma = 1$ , which yields a money demand consistent with the US evidence for M1.

This specification can be shown to imply a demand for money, conditional on labor supply, with unit consumption-elasticity. With respect to the nominal interest rate,  $R$ , is approximately log-log with elasticity of one half. That is, a square root demand:

$$\frac{m}{c} \approx \sqrt{\frac{m}{kg(n)}}$$

Note the dependence on labor supply here,  $n$ , absent in Lucas' analysis.

We work with the continuous heterogenous population indexed by productivity,  $w$ . The resource constraint thus has the form,

$$G + \int c(w)f(w)dw \leq \int n(w)wf(w)dw + A$$

where  $A$  is a variable that captures extra-resources in the hands of the government. Here  $f(w)$  is the density of the productivity level  $w$  in the population. We model  $f(w)$  to be log-normal so that  $\log w \sim N(\mu, \sigma)$ . The parameters were set, following Tuomala (1990), to  $\mu = -1$  and  $\sigma = 0.39$ .

As a first exercise, we computed the model for the benchmark case where  $G = 0$  so that all taxation is for redistribution purposes only. We set  $A = 0$  for the Friedman rule,  $R = 0$ , and compute for each  $R > 0$  the required  $A(R)$  that allows the government to obtain the same welfare – using an un-weighted utilitarian welfare criterion. We report the welfare cost as a fraction of total consumption:

$$\frac{A(R)}{\int c(w)f(w)dw}$$

The welfare results are reported in figure 1, figure 2 shows the money demand function generated by this exercise and figure 3 reports the optimal marginal tax rates<sup>6</sup>. The money demand function matches the one used by Lucas very well. Note that in this exercise the demand for money will depend on the, endogenously determined, supply of labor and on its distribution, thus  $k$  was calibrated through trial and error to yield the right scale for demand.

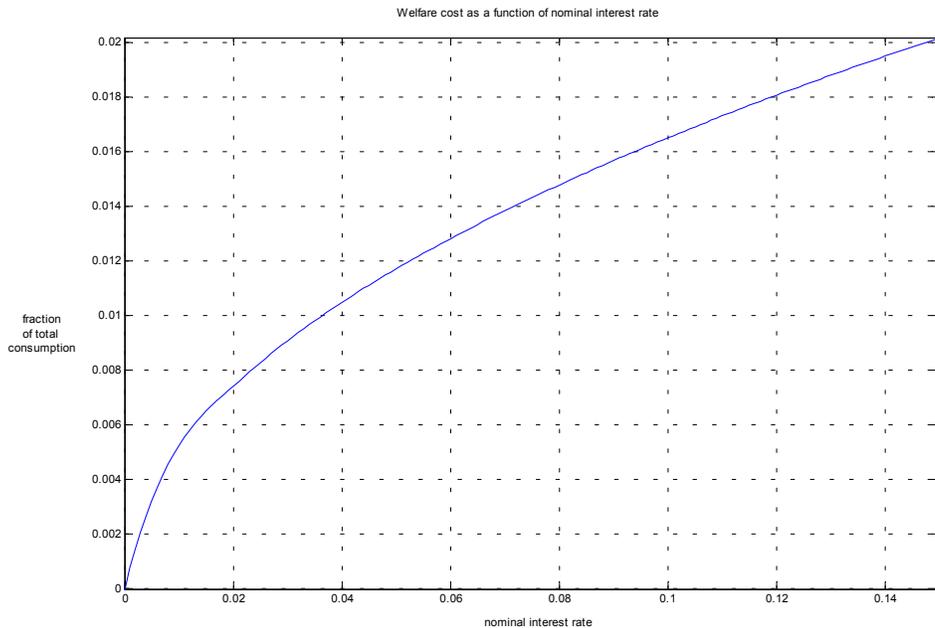
The marginal tax rates structure are consistent with those found by Mirlees (1971), Tuomala (1990) and others. The marginal tax rate is decreasing

---

<sup>6</sup>The optimum can be written as an optimal control problem and the solution can then be found by solving a system of first order ordinary differential equations with non-trivial boundary conditions.

over the population fraction, so that taxation is not redistributive in the sense of increasing marginal tax rates. There is a significant positive subsidy at the zero income level so that taxation is redistributive in the sense that taxes paid as a fraction of income are increasing in income. The average marginal tax in our exercise is relatively low. These marginal rates are higher with positive government expenditure and/or a more concave welfare function.

The welfare cost, relative to the Friedman rule, of a 5% nominal interest rate is about 1.2% points of GDP. A nominal interest rate of 20% has a cost of 2% of GDP. The gains from moving from a nominal interest rate of 6% to 3% are estimated to be about 0.4% of GDP.



Interestingly, the welfare costs coincide almost exactly with Lucas' (2000) calculation, although they are slightly larger. This fact may seem remarkable: Lucas analyzed a first-best representative-agent economy while here we are studying a non-trivial second-best problem with heterogeneous agents. The welfare-cost measure in Lucas is the additional consumption required to compensate the single agent, here it is the additional resources in the hands of the government. However, the similarity of the results can be easily explained.

Let  $W$  denote total welfare, differentiating and using the envelope condition one can show that the change in welfare from a change in  $R$  equals:

$$\frac{dW}{dR} = \lambda R \int \frac{\partial m^c}{\partial R} f dw - \int V_y \mu m_w dw$$

where  $\lambda$  is the multiplier on the resource constraint. This equation is the continuous analog of equation (18) which was set to zero to find the optimal nominal interest rate. Similarly, a change in  $A$  has:

$$\frac{dW}{dA} = \lambda$$

A compensated change thus satisfies:

$$\frac{dW}{dR} dR + \frac{dW}{dA} dA = 0.$$

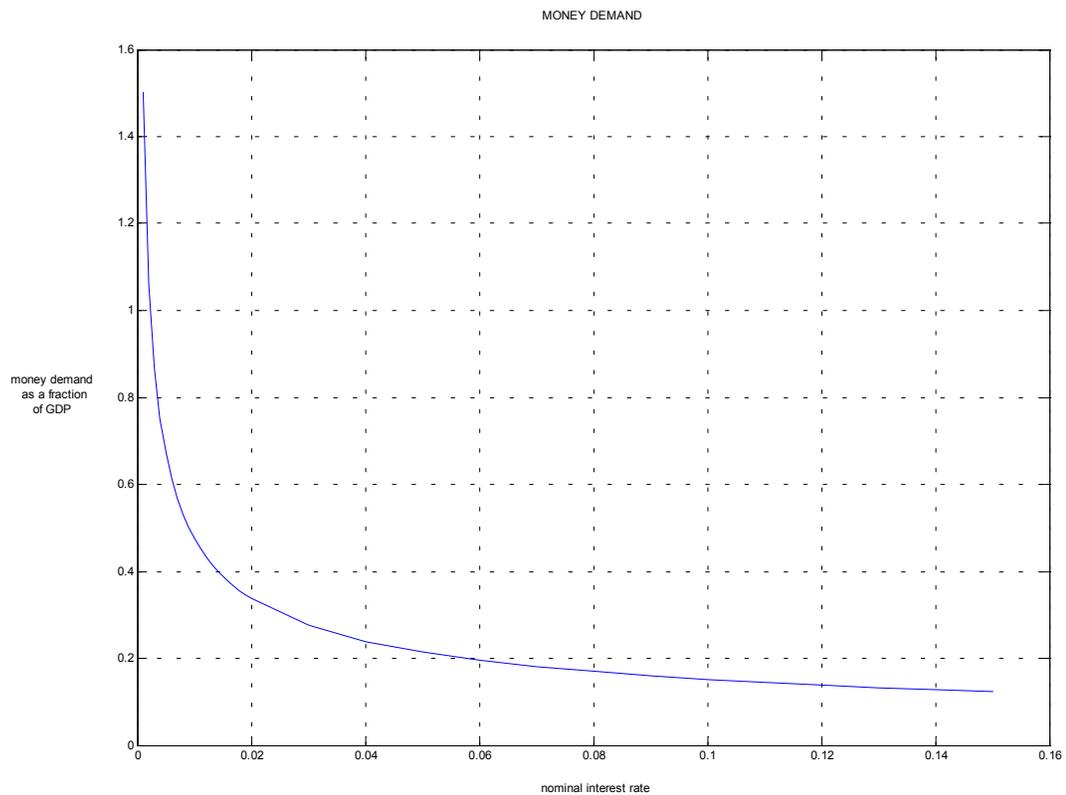
Thus we obtain,

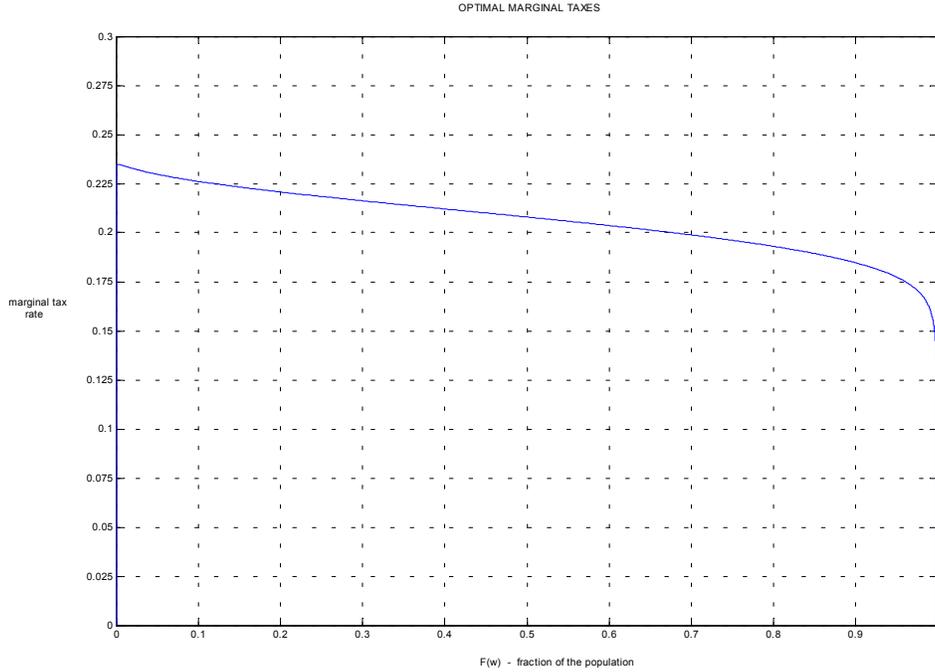
$$\frac{dA}{dR} = R \int \frac{\partial m^c}{\partial R} f dw + \left( -\frac{1}{\lambda} \int V_y \mu m_w dw \right)$$

If the second term is zero the welfare cost is equal to the area under the compensated demand curve – this would be the case in a cash-credit model with separability between the two consumption goods. In the shopping time model the second term is always positive, thus the welfare cost must be greater than the area under the compensated demand curve. Our results are consistent with a small positive second term so that the first term dominates.

The second term becomes larger when the welfare function is more concave and when government expenditure is larger. Both these effects also tend to increase the average marginal income tax. Simulations show that the welfare costs increase by about 5-10% for parametrizations that imply average marginal income tax rates between 30-35%. Thus the second term is generally not trivial. However, the area under the demand curve does provides a

good approximation to the welfare costs.





## 7 Extensions

### 7.1 Multiple Consumption Goods

We now study a generalization with multiple consumption goods:  $c \in \mathbb{R}^n$  and check the conditions under which  $m_w \leq 0$ . We can define unconditional demands here as we did in (19). It turns out that a version of proposition 2 is still available for  $n$  consumption goods. Letting  $e = \sum p_i c_i$ , denote total expenditures on consumption goods, define  $\varepsilon_{ew} = \sum s_i \varepsilon_{c_i w}$  and  $\eta_e = \sum s_i \eta_{c_i}$  as the elasticity of  $e$  with respect to the wage,  $w$ , and full income,  $I$ , respectively.

**Assumption C:**  $s_e \eta_e + s_m \eta_m > 0$  (i.e.  $\eta_l < 1/s_l$ )

**Proposition 5** Under assumption C,  $m_w(y, Y, R, w) \leq 0$  for  $(y, Y, R, w) \in \mathbb{R}_+^4$  if and only if,

$$\frac{\eta_e}{[s_e \eta_e + s_m \eta_m]} \geq \frac{\varepsilon_{ew}}{[s_e \varepsilon_{ew} + s_m \varepsilon_{mw}]} \quad (22)$$

or the equivalent condition,

$$\eta_e \varepsilon_{mw} \geq \eta_m \varepsilon_{ew}, \quad (23)$$

where  $\varepsilon_{iw}$  and  $\eta_i$  represent the elasticity of good  $i$  with respect to  $w$  (Hicksian or Marshallian) and income, respectively.

Assumption  $C$  is the direct analog of Assumption  $A$  and the proof of proposition 5 is a direct extension of that of proposition 2.

To gain insight into the economic aspects of condition (22) we now study in more detail a generalized version of the shopping-time model. We allow the shopping-time function to take the general form  $v(c_1, c_2, \dots, c_n, m)$ . This setup therefore allows some goods to be relatively more money or time intensive than others as in the cash-credit good model.

We also allow preferences over consumption goods and leisure to be defined in a general way by  $U(c_1, c_2, \dots, c_n, L)$ . The reduced form utility function over consumption goods, money and non-work time is then:

$$u(c_1, c_2, \dots, c_n, m, l) \equiv U(c_1, c_2, \dots, c_n, l - v(c_1, c_2, \dots, c_n, m)).$$

We now check condition (23) for this generalized shopping-time model. To do so we decompose the demand for  $m$  as before. That is, conditional on the consumption vector,  $c$ , we solve:

$$V(c, I, R, w, p) \equiv \max_{l, m} U(c, l - v(c, m))$$

$$p'c + Rm + wl = I$$

with demands:  $m^*(c, I, R, w, p)$  and  $l^*(c, I, R, w, p)$ . Here as in section 5.1 the maximization implies that  $m_I^* = 0$  and  $m_w^* > 0$ , or in elasticity form that:  $\eta_m^* = 0$  and  $\varepsilon_{mw}^* > 0$ .

Using the unconditional demands for consumption goods,  $\tilde{c}(R, w, I)$ , we can express the unconditional demand for  $\tilde{m}(R, w, I)$  as:

$$\tilde{m}(R, w, I) \equiv m^*(\tilde{c}(R, w, I), I, R, w).$$

Then the elasticities satisfy:

$$\eta_m = \sum_i \varepsilon_{mc_i}^* \eta_{c_i}$$

$$\varepsilon_{mw} = \sum_i \varepsilon_{mc_i}^* \varepsilon_{c_i w} + \varepsilon_{mw}^*$$

With these decompositions, (23) becomes,

$$\eta_e \left( \sum_i \varepsilon_{mc_i}^* \varepsilon_{c_i w} + \varepsilon_{mw}^* \right) \geq \left( \sum_i \varepsilon_{mc_i}^* \eta_{c_i} \right) \varepsilon_{ew}.$$

Rearranging this expression and assuming the composite consumption good to be normal leads to proposition 6.

**Assumption D**  $\eta_e \geq 0$ .

**Proposition 6** *Under assumptions C and D, in the multiple-good shopping-time model  $m_w(y, Y, R, w) \leq 0$  for  $(y, Y, R, w) \in \mathbb{R}_+^4$  if and only if:*

$$\varepsilon_{mw}^* \geq \sum_i s_i \left( \frac{\varepsilon_{mc_i}^*}{s_i} \right) (\varepsilon_{ew} (\hat{\eta}_i - \hat{\varepsilon}_{c_i w})), \quad (24)$$

where,

$$\hat{\eta}_i \equiv \frac{\eta_i}{\eta_e} \quad \text{and} \quad \hat{\varepsilon}_{c_i w} \equiv \frac{\varepsilon_{c_i w}}{\varepsilon_{ew}},$$

where  $s_i$  is shares of good  $i$ .

Equation (24) generalizes the single good result in proposition 3 to multiple goods. The new term is a ‘‘covariance’’ term between two variables, each of which affords a simple economic interpretation. It is easy to see that  $\varepsilon_{mc_i}^*/s_i$  represents the relative money-intensity of good  $i$ . That is,

$$\frac{\varepsilon_{mc_i}^*}{s_i} = \frac{1}{m} \frac{\partial m^*}{\partial c_i} \frac{e}{p_i},$$

is proportional to the percentage increase in money balances from an additional dollar of purchases on good  $i$ . It can be shown, by a simple extension of the reasoning in appendix A, that  $\varepsilon_{ew} (\hat{\eta}_i - \hat{\varepsilon}_{c_i w})$  is proportional to the percentage change in  $c_i$  from an increase in leisure while holding expenditure on all consumption goods,  $\sum_i p_i c_i$ , constant.

The new term therefore represents the average relationship between money-intensities and the shift in the pattern of consumption from increases in

leisure. Clearly, in general, this relationship can be positive or negative, so that  $m_w \leq 0$  cannot be guaranteed. However, as (24) illustrates, because  $\varepsilon_{mw}^*$  is non-negative,  $m_w > 0$  requires goods to vary in money-intensities and leisure to shift the relative pattern of consumption towards money intensive goods. Of course, even then, the potentially positive “covariance” may or may not overcome the positive  $\varepsilon_{mw}^*$  term.

Consider the following benchmark case, stated as a corollary.

**Corollary:** *Let the shopping-time function be of the form  $\hat{v}(p'c, m)$ , for some function  $\hat{v} : \mathbb{R}^2 \rightarrow \mathbb{R}$ , with the same properties as discussed in section 5.1. Then under assumptions C and D at the optimum  $R \leq \bar{R}$ .*

In this case, the effect of each good on the conditional demand  $m^*(c, I, R, w)$  operates only through total expenditures on consumption goods,  $e$ . The relative money-intensities,  $\varepsilon_{mc_i}^*/s_i$ , are then constant across goods, so that the “covariance” term in (24) vanishes and it simplifies to,

$$\varepsilon_{mw}^* \geq 0,$$

which is guaranteed by the shopping-time model.

This simple example illustrates the kind of departure required for the Friedman rule to be non-optimal. The cash-credit good model offers this possibility. This model has extreme differences in money-intensities, if labor affects the relative consumption towards the cash good then the Friedman rule is not optimal. The case of proposition 4 has a “covariance” term equal to zero because relative consumptions of cash and credit goods are not affected by leisure.

It is important to note that the discussion here, as in the cash-credit literature, assumes that consumption goods are not differentially taxed. The nominal interest rate is then the only instrument affecting prices. If differential taxation of consumption goods were feasible the implications for  $R$  could be important.

This is most notably the case in the cash-credit good case where  $R$  is used to indirectly tax or subsidize the cash good. If taxation were used directly on the cash good  $R$  is indeterminate. This is true because of the complete absence of substitution in the transaction technology. More generally  $R$  is not indeterminate, but its role is greatly affected by the use of differential commodity taxation.

## 7.2 Consumption Taxes Paid with Cash

We have studied the optimality of setting  $R \leq \bar{R}$  for the case where consumption taxes are zero. The motivation for our approach is the well understood point that some normalization is generally required to speak of optimal tax rates. In this subsection we introduce consumption taxes that are partly paid with cash and study the set of equivalent tax structures. Because consumption taxes require cash the set of equivalent taxes is somewhat different from the standard public finance principle of constant relative tax rates.

We return to the main model with a single consumption good. For simplicity we work out the equivalence results in the context of a linear income tax. The results can be extended straightforwardly to non-linear income taxation.

We now characterize the set of equivalent tax structures. We define a tax system by a vector of tax rates  $\tau = (\tau_c, \tau_L, R - \bar{R})$  that affects the budget set of the agent in the following way:

$$(1 + \tau_c)pc + RM = w(1 - L)(1 - \tau_L)$$

both sides are measured in monetary units, e.g. dollars.

Assume a fraction  $\lambda$  of consumption taxes require money. To capture this, we assume the relevant measure for agents of real money balances to be  $m \equiv M/p(1 + \lambda\tau_c)$ . Re-expressing the budget constraint in real terms we obtain:

$$(1 + \tau_c)c + (1 + \lambda\tau)Rm = \frac{w}{p}(1 - L)(1 - \tau_L).$$

Government collects revenues in dollars to finance the purchase of  $G$  units of good  $c$ ,

$$\tau_c pc + (R - \bar{R})M + \tau_L w(1 - L) = pG.$$

In real terms:

$$\tau_c c + (1 + \lambda\tau)(R - \bar{R})m + \tau_L \frac{w}{p}(1 - L) = G.$$

Normalize  $w/p = 1$ . For every  $\lambda$  we have a function  $q_\lambda(\tau) : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ , with  $q_\lambda(\tau) \equiv [(1 + \tau_c), (1 + \lambda\tau_c)R, -(1 - \tau_L)]$ , which defines the prices faced by

agents as a function of tax variables. Letting  $q_\lambda^i(\tau)$  denote the  $i^{\text{th}}$  component  $q_\lambda(\tau)$  and the agent's budget constraint can be written:

$$q_\lambda^1(\tau)c + q_\lambda^2(\tau)m + q_\lambda^3(\tau)(1 - L) = 0 \quad (25)$$

while government's becomes:

$$(q_\lambda^1(\tau) - 1)c + q_\lambda^2(\tau)(1 - \bar{R}/R)m + (q_\lambda^3(\tau) + 1)(1 - L) = G \quad (26)$$

As expected, substituting (25) into (26) yields the resource constraint,

$$c + \bar{R}m + (1 - L) = G, \quad (27)$$

which is independent of  $q_\lambda(\tau)$ . Because equation (25) is homogeneous in  $q_\lambda(\tau)$ , for each tax system we can define an equivalent class as follows:

$$E(\tau) = \{\tau' : q(\tau') = kq(\tau) \text{ for some } k \in \mathbb{R}\}$$

so that all  $\tau' \in E(\tau)$  generate the same allocation. One can characterize  $E(\tau)$  by two scalars,  $\delta_2$  and  $\delta_3$ , defined by:

$$\delta_2 \equiv \frac{(1 + \lambda\tau_c)R}{(1 + \tau_c)} \quad (28)$$

$$\delta_3 \equiv \frac{(1 - \tau_L)}{(1 + \tau_c)}. \quad (29)$$

The interpretation of  $\delta_2$  and  $\delta_3$  is that they represent  $R$  and  $(1 - \tau_L)$  for the case where  $\tau_c = 0$ . This justifies our approach of solving for the case where  $\tau_c = 0$ . Once this problem is solved and  $\delta_2$  and  $\delta_3$  are obtained, one can compute the whole equivalent class that attains the optimal allocation.

For example, when  $\lambda = 0$ , the transformation required is the standard public finance one of scaling  $R$  up by  $1 + \tau_c$ . If  $\lambda = 1$ ,  $R$  is determined independently of the particular tax system within each equivalent class, it is pinned down. This represents a significant difference from the standard public finance principle of constant tax rates. The reason, of course, is that when consumption taxes require money, the definition of  $m$  depends on taxes, so that holding relative tax rates constant does not imply holding relative prices for all goods constant.

As long as  $\lambda < 1$ ,  $R$  does depend on  $\tau_c$ , that is, it is not uniquely determined in each equivalent class. As  $\lambda$  increases, the sensitivity of  $R$  with respect to  $\tau_c$  falls. Because  $\lambda = 1$  is a good benchmark<sup>7</sup>, for reasonable values of  $\lambda$  and  $\tau_c$ , the departures of  $R$  from  $\delta_2$  are likely to be small and not very sensitive to  $\tau$ .

The Friedman rule is usually interpreted in the context of  $\bar{R} = 0$  as prescribing  $R = 0$ . Throughout our analysis we have interpreted the Friedman rule more generally as  $\delta_2 \leq \bar{R}$ . This is a natural extension to positive costs since then  $\delta_2 \rightarrow 0$  as  $\bar{R} \rightarrow 0$ . Furthermore, from (28) we have that  $\delta_2 \rightarrow 0$  implies that  $R \rightarrow 0$  if consumption taxes are bounded.

## 8 Conclusion

Macroeconomics has introduced elements of contract theory and agent heterogeneity to address a number of interesting issues. For normative taxation issues, however, the representative agent Ramsey paradigm still reigns supreme. This framework imposes ad hoc restrictions on the taxation instruments that are allowed. Furthermore, the representative agent assumption used is especially ill suited for the most fundamental issue of taxation: the trade-off between redistribution and efficiency.

Our paper studies the optimal incentive compatible tax-implementable allocation introduced by Mirrlees (1971, 1976). For this problem the non-linear income tax together with linear commodity taxes comprise the natural informationally feasible set of instruments. Instead of ad hoc assumptions it is the economics of private information that restricts the set of available tax instruments.<sup>8 9</sup>

We apply this model to the optimal inflation tax question and derive the general conditions of optimality. For shopping-time models we show that the Friedman rule is optimal under essentially no assumptions on preferences

---

<sup>7</sup>Mulligan and Sala-i-Martin (1997) emphasize the possibility of  $\lambda < 1$  because they do not model income taxes directly as we do. In the model, income taxes are not paid with cash.

<sup>8</sup>Werning (2000) examines other dynamic taxation such as capital income taxation and intertemporal tax smoothing.

<sup>9</sup>We have motivated taxation by redistribution and private information as in Mirrlees (1971), a self-selection setup. Another motivation for taxation is social insurance and hidden-action, a moral-hazard setup. However, da Costa and Werning (2000) show the commodity tax implications of both models to be identical.

or transaction technologies. We then extend these models to allow multiple consumption goods. This analysis highlights the importance of differing money-intensities across goods. In particular, if money intensities are constant across goods the Friedman rule is optimal. If money intensities do vary the Friedman rule is guaranteed as long as leisure does not shift consumption towards money-intensive goods too strongly. The cash-credit model is an extreme version of this extended model, so that the optimality of the Friedman rule for this model depends crucially on the interaction of preferences between leisure and the two consumption goods.

The non-linear income tax was required to obtain clear cut results. If we were to constrain the government, in an ad hoc way, to affine income taxes, the inflation tax may then have the role of mimicking the lost non-linearity, picking up some “leftovers” from the linearly restricted income tax. Yet, there is no telling in which direction this mimicking plays out. Therefore, for this arbitrary, linearly-restricted problem our results may be useful benchmarks.

Our analysis allows the social marginal cost of money to be positive. This is done to address the issue some authors have raised regarding the role of the free-good aspect of money in the optimality of the Friedman rule [Correia and Teles (1996, 1999)]. For our setup the free-good aspect is of no consequence for the optimal taxation of money.

## A Proof of Proposition 2

Given the Marshallian or Hicksian demand system  $\left\{ \tilde{c}(R, w, v), \tilde{m}(R, w, v), \tilde{l}(R, w, v) \right\}$ , where  $v$  represents full income or utility, the shadow wage,  $w^*(x, R, v)$ , and virtual income,  $v^*(x, R, y)$ , are implicitly defined by the identities:

$$\begin{aligned} x &\equiv \tilde{l}(R, w^*(R, x, y), v^*(R, x, y)) \\ y &\equiv \tilde{c}(R, w^*(R, x, y), v^*(R, x, y)) + R\tilde{m}(R, w^*(R, x, y), v^*(R, x, y)) \end{aligned}$$

where  $y$  and  $x$  represent expenditure on  $c$  and  $m$  and leisure, respectively. As functions of  $(R, x, y)$ ,  $c$  and  $m$  are then:

$$\begin{aligned} c(x, R, y) &\equiv \tilde{c}(R, w^*(x, R, y), v^*(x, R, y)) \\ m(x, R, y) &\equiv \tilde{m}(R, w^*(x, R, y), v^*(x, R, y)) \end{aligned}$$

Differentiating these four identities with respect to  $x$  yields:

$$\begin{aligned} 1 &= \varepsilon_{lw} \frac{\partial \log w^*}{\partial \log x} + \eta_l \frac{\partial \log v}{\partial \log x} \\ 0 &= s_c \frac{\partial \log c}{\partial \log x} + s_m \frac{\partial \log m}{\partial \log x}, \\ \frac{\partial \log c}{\partial \log x} &= \varepsilon_{cw} \frac{\partial \log w^*}{\partial \log x} + \eta_c \frac{\partial \log v^*}{\partial \log x} \\ \frac{\partial \log m}{\partial \log x} &= \varepsilon_{mw} \frac{\partial \log w^*}{\partial \log x} + \eta_m \frac{\partial \log v^*}{\partial \log x} \end{aligned}$$

where  $s_i$  are the shares of  $c$  and  $m$  in expenditure,  $y$  (i.e.  $s_c + s_m = 1$ ).

It is then enough to notice that,

$$\frac{\partial \log v}{\partial \log x} = - \frac{[s_c \varepsilon_{cw} + s_m \varepsilon_{mw}]}{[s_c \eta_c + s_m \eta_m]} \frac{\partial \log w^*}{\partial \log x}.$$

We then obtain,

$$\frac{\partial \log c}{\partial \log x} = \left\{ \varepsilon_{cw} - \frac{\eta_c}{[s_c \eta_c + s_m \eta_m]} [s_c \varepsilon_{cw} + s_m \varepsilon_{mw}] \right\} \frac{\partial \log w^*}{\partial \log x}.$$

Given assumption  $A$ , it can be shown that  $\partial \log w^* / \partial \log x < 0$ . Then,

$$\frac{\partial \log c}{\partial \log x} \geq 0 \Leftrightarrow \frac{\varepsilon_{cw}}{[s_c \varepsilon_{cw} + s_m \varepsilon_{mw}]} - \frac{\eta_c}{[s_c \eta_c + s_m \eta_m]} \geq 0$$

which proves the proposition since  $\partial \log c / \partial \log x \geq 0$  if and only if  $\partial \log m / \partial \log x \leq 0$ .

## B Proof of Stationarity

Because our environment is stationary we restricted attention to stationary allocations. We now prove that this is without loss in generality. That is, we prove that a stationary allocation is, in fact, optimal.

If individual savings are not observed, the set of implementable allocations must impose the restriction that agents can trade with each other – that intertemporal marginal rates of substitution are equalized across agents – so that the pattern of expenditures over time for each agent is restricted.

Denote the set of such allocations by  $A$ . The task of characterizing  $A$  may be a difficult one. Fortunately, for our simple stationary environment we do not require much knowledge of  $A$ .

Let  $B$  denote the set of allocations that are implementable when savings are observable, so that each agent's expenditures across periods are unrestricted. We shall work with a set closely related to  $B$  which we denote by  $B'$ . This set replaces the agents reporting maximization problem with the related first and second order conditions. Clearly,  $A \subset B \subset B'$ .

Our strategy is to state the relaxed problem of maximizing welfare subject to  $B'$ . We then argue that the solution to this problem yields a stationary allocation. Because of stationarity we are able to prove that it lies in  $B$ , that is that the first and second order conditions that we impose are also sufficient for the agent's maximization when the allocation is stationary. Finally, all stationary allocations belonging to  $B$  belong in  $A$  as well – since intertemporal marginal rates of substitution are all equal to  $\beta$  they are equalized.

The set  $B$  imposes the resource constraint and that agents reveal their type:

$$w = \arg \max_{\hat{w}} \sum_{t=0}^{\infty} \beta^t V(y_t(\hat{w}), Y_t(\hat{w}), R_t, w). \quad (30)$$

The f.o.c. for (30) is,

$$\sum_{t=0}^{\infty} \beta^t \{V_{y_t} y'_t(w) + V_{Y_t} Y'_t(w)\} = 0. \quad (31)$$

If one defines  $v_t(w)$  by:

$$v_t(w) \equiv V(y_t(w), Y_t(w), R_t, w)$$

then imposing

$$\sum_{t=0}^{\infty} \beta^t v'_t(w) = \sum_{t=0}^{\infty} \beta^t V_w(e(v_t(w), Y_t(w), R_t, w), Y_t(w), R_t, w),$$

is equivalent to (31).

The second order condition is

$$D \equiv \sum_{t=0}^{\infty} \beta^t \{V_{y_t y_t} y_t'^2 + 2V_{y_t Y_t} y_t' Y_t' + V_{Y_t Y_t} Y_t'^2 + V_{y_t} y_t'' + V_{Y_t} Y_t''\} \leq 0.$$

Differentiating (31) we find an alternative expression:

$$D = - \sum_{t=0}^{\infty} \beta^t \{V_{wy_t} y'_t + V_{wY_t} Y'_t\} \leq 0 \quad (32)$$

Before we write the problem it is important to transform from the  $y, Y$  space to the  $v, Y$  space. By definition  $y_t(w) \equiv e(v_t(w), Y_t(w), R_t, w)$  thus,  $y'_t(w) = e_v v'_t + e_Y Y'_t + e_w$ . Substituting this into (32) yields

$$\sum_{t=0}^{\infty} \beta^t \{V_{wy_t} (e_v v'_t + e_Y Y'_t + e_w) + V_{wY_t} Y'_t\} \geq 0,$$

Substituting condition (30) for these first and second order condition leads to the following problem.

**Relaxed problem:**

$$\max_{v_t(w), Y_t(w), R_t \geq 0} \sum_{t=0}^{\infty} \beta^t \int v_t(w) f(w) dw,$$

$$\sum_{t=0}^{\infty} \beta^t v'_t(w) = \sum_{t=0}^{\infty} \beta^t V_w(e(v_t(w), Y_t(w), R_t, w), Y_t(w), R_t, w)$$

$$\int [Y_t(w) - c^c(v_t(w), Y_t(w), R_t, w) - \bar{R}m^c(v_t(w), Y_t(w), R_t, w)] f(w) dw = G \text{ for all } t$$

$$\sum_{t=0}^{\infty} \beta^t \{V_{wy_t} (e_v v'_t + e_Y Y'_t + e_w) + V_{wY_t} Y'_t\} \geq 0,$$

To find the necessary conditions for this problem we will write this problem as an optimal control problem.

We define the controls to be  $\{x_t\}$  and  $\{z_t\}$  and the states  $\{v_t\}$  and  $\{Y_t\}$ . We deal with the (isoperimetric) resource constraints in the usual way by defining another state variables:

$$b_t(\omega) = \int_0^{\omega} [Y_t(w) - c^c(v_t(w), Y_t(w), R_t, w) - \bar{R}m^c(v_t(w), Y_t(w), R_t, w)] f(w) dw,$$

representing the net resources from agents 0 to  $\omega$ . The resource constraints are imposed by adding the boundary conditions  $b_t(0)$  and  $b_t(\bar{w}) = G$ .

We can reduce the notational burden by writing the first and second order conditions as:

$$\sum_{t=0}^{\infty} \beta^t \{x_t - W^1[v_t, Y_t, R_t, w]\} = 0$$

$$\sum_{t=0}^{\infty} \beta^t W^2[v_t, Y_t, R_t, x_t, z_t, w] \geq 0,$$

where  $W^1$  and  $W^2$  are defined in the obvious way.

Letting dots above variables represent derivatives with respect to  $w$  and omitting the dependence on the argument  $w$  in controls and states, we write the following control problem:

$$\max_{\{x_t, z_t, R_t\}} \sum_{t=0}^{\infty} \beta^t \int v_t f dw,$$

$$\dot{v}_t = x_t$$

$$\dot{Y}_t = z_t$$

$$\dot{b}_t = [Y_t - c^c(v_t, Y_t, R_t, w) - \bar{R}m^c(v_t, Y_t, R_t, w)] f dw$$

$$b_t(0) = 0 \text{ and } b_t(\bar{w}) = G \text{ for all } t$$

$$\sum_{t=0}^{\infty} \beta^t \{x_t - W^1[v_t, Y_t, R_t, w]\} = 0$$

$$\sum_{t=0}^{\infty} \beta^t W^2[v_t, Y_t, R_t, x_t, z_t, w] \geq 0,$$

The related (extended) Hamiltonian is,

$$\mathcal{H} \equiv \sum_{t=0}^{\infty} \beta^t \{v_t f + \gamma_t^v x_t + \gamma_t^Y z_t + \gamma_t^b (Y_t - c^c - \bar{R}m^c) + \delta_f (x_t - W^1) + \delta_s W^2\}$$

the necessary conditions are:

$$\begin{aligned} \frac{1}{\beta^t} \frac{\partial \mathcal{H}}{\partial z_t} &= \gamma_t^Y + \delta_s W_{z_t}^2 = 0 \\ \frac{1}{\beta^t} \frac{\partial \mathcal{H}}{\partial x_t} &= \gamma_t^v + \delta_f + \delta_s W_{x_t}^2 = 0 \\ \frac{1}{\beta^t} \frac{\partial \mathcal{H}}{\partial R_t} &= \int \{ \gamma_t^b (c_{R_t}^c + \bar{R}m_{R_t}^c) f - \delta_f W_{R_t}^1 + \delta_s W_{R_t}^2 \} = 0 \\ -\dot{\gamma}_t^v &= f + \gamma_t^b (c_v^c + \bar{R}m_v^c) f - \delta_f W_v^1 + \delta_s W_v^2 \\ -\dot{\gamma}_t^Y &= \gamma_t^b (c_Y^c + \bar{R}m_Y^c) f - \delta_f W_Y^1 + \delta_s W_Y^2 \\ -\dot{\gamma}^b &= 0 \end{aligned}$$

The main observation to make is that these equations do not depend in any way on  $t$ . Thus, in general the solution to these equations is stationary,  $\gamma_t^i = \gamma^i$ ,  $b_t = b$ ,  $v_t = v$ ,  $Y_t = Y$ ,  $R_t = R$  and thus,  $y_t \equiv e(v_t, Y_t, R_t, w) = y$ .

Does this stationary solution to the necessary conditions represent the solution to program? In general, aside from non-convexities, the answer is affirmative, but non-convexities cannot a priori be ruled out. The Ramsey literature encounters a similar problem [see Chari's comment on Mulligan and Sala-i-Martin (1997)]. As in this literature, we assume the problem is sufficiently convex, so that stationary solutions are optimal as we've shown. Alternatively, we introduce lotteries which deal with the non-convexities in a more straightforward manner [see appendix C].

We now prove that this stationary allocations is in  $B$ , and then that it is in  $A$ . To see this we must show that the first and second order conditions are sufficient for the agent's revelation problem in this case. The agent's revelation problem with a stationary allocation is simply:

$$\begin{aligned} w &= \arg \max_{\hat{w}} \frac{1}{1-\beta} V(y(\hat{w}), Y(\hat{w}), R, w) \\ &= \arg \max_{\hat{w}} V(y(\hat{w}), Y(\hat{w}), R, w) \end{aligned} \tag{33}$$

which is the same as in the main body of the paper. As we mentioned there, Mirrlees has shown that given the SC assumption, the first and second order conditions are necessary and sufficient (furthermore, the second order condition boils down to  $Y'(w) \geq 0$ ) for (33). Thus, the allocation we found imposing the first and second order conditions, because it is stationary, satisfies (30) and thus belongs to  $B$ .

Finally, any stationary allocation in  $B$  also belongs in  $A$  since intertemporal marginal rates of substitution are equalized across agents and equal to the discount factor  $\beta$ .

## C Random Allocations

The government now offers a lottery over  $(y, Y)$  for each reported  $\hat{w}$ . We represent the lotteries as follows. After reporting  $\hat{w}$  the agent obtains the allocation pair  $(y(\varepsilon; \hat{w}), Y(\varepsilon; \hat{w}))$  where  $\varepsilon$  is a distributed uniformly  $[0, 1]$  i.i.d. across agents.

Each agent faces a random allocation but the revenue for the government is deterministic. The budget constraint with truth-telling  $\hat{w} = w$  is then:

$$\int \int [w(1 - l^*(w, R; y(\varepsilon; w))) - y(w(1 - l^*(w, R; y(\varepsilon; w)))) + (R - \bar{R})m^*(w, R; y(\varepsilon; w))] f(w) d\varepsilon dw = G.$$

Agents report:

$$\hat{w} \equiv \arg \max_{w'} \int V(y(\varepsilon; w'), Y(\varepsilon; w'), R, w) d\varepsilon,$$

where  $V(\cdot)$  continues to be defined by (7). Applying the envelope theorem yields the incentive compatibility condition:

$$v'(w) = \int V_w(y(\varepsilon, w), Y(\varepsilon, w), R, w) d\varepsilon.$$

For simplicity, we do not incorporate the second order condition of agent's maximization. The Lagrangian is then,

$$\mathcal{L} = \int \int \{vf + \mu [v' - V_w(\varepsilon)] + \lambda (Y(\varepsilon) - c^c(\varepsilon) - \bar{R}m^c(\varepsilon)) f\} d\varepsilon dw.$$

(where only the dependence on  $\varepsilon$  is made explicit). Proceeding as in Section 3 we obtain:

$$(R - \bar{R}) \int \int m_R^c(\varepsilon) d\varepsilon f dw = - \int \int \left( \frac{V_y(\varepsilon) \mu}{\lambda} \right) m_w(\varepsilon) d\varepsilon dw \quad (\text{B1})$$

Equation (B1) is the analog of equation (18) with randomization. The same considerations on the signs of  $m_R^c$ ,  $V_y$ ,  $\lambda$  and  $\mu$  imply that  $R \leq \bar{R}$  if  $m_w \leq 0$ . Hence, the deterministic analysis fully carries over to random policies.

## References

- [1] Atkinson, A.B. and Stiglitz, J.E. (1976) “The Design of Tax Structure: Direct vs. Indirect Taxation” *Journal of Public Economics*, 6: 55-75.
- [2] Brunner, Johann K. (1990), “Utilitarian Income Taxation” *Arbeitspapier Nr. 9003*, Institute für Volkswirtschaftslehre, Universität Linz.
- [3] Brunner, Johann K. (1993), “A Note on the Optimum Income Tax” *Journal of Public Economics* 50, p. 445-451.
- [4] Chari, V.V., Lawrence Christiano and Patrick Kehoe. “Optimality of the Friedman Rule in Economies with Distorting Taxes”, *Journal of Monetary Economics*, 37 (June) p. 203-33.
- [5] Correia, Isabel and Teles, Pedro (1996) “Is the Friedman Rule Optimal when Money is an Intermediate Good?” *Journal of Monetary Economics*, 38 223-244.
- [6] Correia, Isabel and Teles, Pedro (1999) “The Optimal Inflation Tax”, *Review of Economic Dynamics* 2, 325-346.
- [7] da Costa, Carlos and Iván Werning (2000) “Commodity Taxation and Social Insurance”, University of Chicago, mimeo.
- [8] Ebert, Udo (1992) “A Reexamination of the Optimal Nonlinear Income Tax”, *Journal of Public Economics* 49, p. 47-73.
- [9] Faig, Miguel “Characterization of the Optimal Tax on Money When It Functions as a Medium of Exchange” *Journal of Monetary Economics* 22(1), (July): 137-48.

- [10] Friedman, Milton (1969) "The Optimum Quantity of Money" in *The Optimum Quantity of Money and Other Essays* (Aldine, Chicago IL).
- [11] Guidotti, Pablo E. and Vegh, Carlos A. (1993) "The Optimal Inflation Tax when Money Reduces Transaction Costs" *Journal of Monetary Economics* 31 (1993), 189-205.
- [12] Kimbrough, Kent (1986) "The Optimum Quantity of Money Rule in the Theory of Public Finance" *Journal of Monetary Economics*, 18, 277-284.
- [13] Lucas, Robert E. Jr. (2000) "Inflation and Welfare", *Econometrica*, 68, pg. 247-274.
- [14] Lucas, Robert E. Jr. and Stokey, Nancy L.(1983), "Optimal Fiscal and Monetary Policy in an Economy without Capital." *Journal of Monetary Economics*. 12(1), July: 55-93.
- [15] McCallum, Bennett T. and Goodfriend, Marvin S (1987) "Demand for Money: Theoretical Studies" in *The New Palgrave: A Dictionary of Economics*, ed. by John Eatwell, Murray Milgate, and Peter Newman, London: Macmillan; New York: Stockton Press, 775-781.
- [16] Mirrlees, James A. (1971), "An Exploration in the Theory of Optimal Income Taxation," *Review of Economic Studies* 38, 175-208.
- [17] Mirrlees, James A. (1976), "Optimal Tax Theory: A Synthesis," *Journal of Public Economics*, 6: 327-58.
- [18] Mulligan, Casey B. and Sala-i-Martin, Xavier X. (1997), "The Optimum Quantity of Money: Theory and Evidence", *Journal of Money, Credit and Banking*, Vol. 29, No. 4 (November, Part 2).
- [19] Phelps, Edmund S. (1973) "Inflation in the Theory of Public Finance", *Swedish Journal of Economics*. 75(1), March: 67-82.
- [20] Ramsey, Frank (1927) "A Contribution to the Theory of Taxation" *Economic Journal* 37 (March): 47-61.
- [21] Seade, Jesus (1982) "On the Sign of the Optimum Marginal Income Tax", *Review of Economic Studies* 49, 637-643.

- [22] Stiglitz, Joseph (1987) “Pareto Efficient and Optimal Taxation and the New New Welfare Economics” in *Handbook of Public Economics*, vol. II, Chapter 15, p. 991-1042.
- [23] Varian, Hal (1980) “Redistributive Taxation as Social Insurance”, *Journal of Public Economics*, 14, 49-68.
- [24] Werning, Iván (2000) “Dynamic Optimal Taxation”, University of Chicago, mimeo.