The Optimal Inflation Tax*
Isabel Correia and Pedro Teles
Banco de Portugal and Universidade Católica Portuguesa, Lisbon, Portugal
Received August 18, 1997

We determine the second best rule for the inflation tax in monetary general equilibrium models where money is dominated in rate of return. The results in the literature are ambiguous and inconsistent across different monetary environments. We derive and compare the optimal inflation tax solutions across the different environments and find that Friedman's policy recommendation of a zero nominal interest rate is the right one.

Key Words: Friedman rule; inflation tax

1. INTRODUCTION

This paper addresses the issue of the optimal inflation tax in monetary general equilibrium models where money is dominated in rate of return. Friedman (1969) addresses this issue in a first best environment, where lump-sum taxes are available. He proposes a monetary policy rule that generates a nominal interest rate equal to zero, corresponding to a zero inflation tax and to a negative rate of inflation. The intuition is simple: since the marginal cost of supplying money is negligible, the marginal benefit should equal the marginal cost, and so the nominal interest rate should be set equal to zero.

We are interested in the more relevant second best results, i.e., when the government must finance government expenditures without having access to lump-sum taxation. Here, the literature is inconsistent, particularly across different monetary environments. The key inconsistency, which will be our main focus here, is that while in models that explicitly specify transaction technologies the Friedman rule is a general result, in models

* We want to thank V. V. Chari, Mike Dotsey, Rody Manuelli, and Harald Uhlig for very valuable comments. Remaining errors are ours. Financial support by PRAXIS XXI is gratefully acknowledged.

325 1094-2025/99 $30.00
Copyright © 1999 by Academic Press
All rights of reproduction in any form reserved.
with money in the utility function the results are ambiguous. As pointed out by Woodford (1990), “either the Phelps or the anti-Phelps result is possible, depending upon details of specification.” In any case, the conventional wisdom is still the intuition in Phelps (1973), that in a second best environment liquidity is a good that should be taxed, just as any other good. We clarify the issues involved and find that the Friedman rule is the optimal policy. The reason for the generality of this result is the fact that money is a free good. We show that since the cost of producing the good is zero, the optimal unit tax is also zero, under general conditions, translating into a robust optimal rule of a zero nominal interest rate. This result is important in that it translates into a very clean policy recommendation, independent of the parameterization of the economy.

The class of general equilibrium models that incorporate the feature of dominance in rate of return, and in which we perform the welfare analysis, are designed in a somehow ad hoc fashion. Where this is more clearly so is in models where the preferences depend on the real quantity of money, as proposed by Sidrauski (1967) and Brock (1975). The fact that the use of money for transactions is not explicit in these models led Clower (1967) to propose a cash-in-advance restriction. Lucas (1980) and Lucas and Stokey (1983) used this approach in a general equilibrium framework. A more complete transactions technology, where it is assumed that time is substitutable for the use of money, was addressed by McCallum (1983), Kimbrough (1986), and McCullum and Goodfriend (1987).

Two major second best taxation sets of rules in the public finance literature have been used to justify the optimal inflation tax results: the Diamond and Mirrlees (1971) optimal taxation rules of intermediate goods and Ramsey's (1927) taxation rules of final goods, further developed by Atkinson and Stiglitz (1972). The Diamond and Mirrlees (1971) optimal taxation rules, derived for the case of constant returns to scale production functions, are the basis for the results in the literature of monetary models with transactions technologies. In Correia and Teles (1996) we show that the Friedman rule is the optimal solution in these monetary models for all homogeneous transactions costs functions. We also show that the interpretation of this result is not a direct extension of the theorem of Diamond and Mirrlees but is related to the free good characteristic of money and to the special structure of production and taxation implied in this class of models.

1 In contrast, models where the purpose is to generate an equilibrium positive price for fiat money are more fundamentally specified. The seminal papers are Samuelson (1958), Grandmont and Younes (1973), Bewley (1980), Townsend (1980), and Kiyotaki and Wright (1989). In these models, the perfect substitutability between money and bonds implies a zero nominal interest rate, and the policy issue is the determination of the real interest rate.

2 See Kimbrough (1986), Guidotti and Végh (1993), and Chari et al. (1996).
Atkinson and Stiglitz (1972) established that it is optimal not to distort the relative prices between consumption of different goods when the preferences are separable in leisure and homothetic in the consumption goods. These rules were applied to cash–credit goods economies by Lucas and Stokey (1983) and Chari et al. (1996). There, the inflation tax translates into discriminated effective taxes on credit and cash goods. The Friedman rule is optimal under the Atkinson and Stiglitz (1972) conditions for uniform taxation.

The Ramsey (1927) rules were also used to explain the results in models with money in the utility function. Phelps (1973) uses this structure, with exogenous factor prices, and concludes that “the optimal inflation tax is positive.” Chamley (1985) aims at generalizing Phelps (1973) results to a general equilibrium model and concludes that the Friedman rule is the optimum only in the first best case. Siegel (1978) stresses the costless nature of liquidity services but concludes that this characteristic does not affect the result of a strictly positive tax on those services. Drazen (1979) states that the distinction between the costliness and costlessness of the production of goods is important in the determination of the second best solution. Nevertheless, he concludes that “it appears difficult to say even whether the optimal inflation rate will be positive or negative.” These results are disturbing because they are not consistent with the general optimality result of the Friedman rule in transactions technology models.

One other reason for the apparent inconsistency in the optimal inflation tax results is that the approach to the second best problem is not uniform. Some authors impose conditions of stationarity on the second best problem. In this third best solution the Friedman rule is never the optimal solution.

The main contribution of this paper is to show that the Friedman rule is indeed a general result in the set-up where liquidity is modeled as a final good. It turns out that the generalized use of the Ramsey (1927) rules to justify the Phelps result is misleading and explains the ambiguity and the apparent inconsistency in the results in the different monetary environments. The rules on optimal taxation of final goods apply to ad valorem taxes on costly goods. In general, the optimal ad valorem consumption taxes are strictly positive. Since the goods are costly, the corresponding unit taxes are also strictly positive. But money is assumed to have a negligible production cost. If this is the case, then the only tax that can generate positive revenue is a unit tax, and in any case the nominal interest rate is by construction a unit tax. The general result that the ad valorem tax rate on real balances is strictly positive can translate in the limit, when the costs of producing money are made arbitrarily small, into an optimal zero nominal interest rate. This is the best intuition for why the Friedman rule is a general result.
In clarifying the puzzles in this literature, we take into account mainly (i) the free good characteristic of real balances, (ii) the fact that models with money in the utility function are reduced forms of more explicit monetary models, and that (iii) the Ramsey problem is unrestricted. We extend the results for money in the utility function models obtained by Chari et al. (1996) and establish the links between the results obtained in the different types of monetary models. In particular, we relate the results from a model with money in the utility function to those from transactions technology models obtained by Correia and Teles (1996).

The optimal rules in the money in the utility function model are derived in Section 2. We also show, in this section, that the Friedman rule is a general result by establishing an equivalence between the money in the utility function models and the underlying transactions technologies models. In Section 3 we provide the main intuition for the results. In Section 4 we compare the results to those from models with credit goods. In Section 5 we discuss the robustness of the results to alternative specifications of the available taxes and alternative timing structures. Section 6 contains the conclusions.

2. MONEY IN THE UTILITY FUNCTION

We use the general equilibrium model of a monetary economy developed by Sidrauski (1967) and later used by Brock (1975), Woodford (1990), and Chari et al. (1996) to discuss the optimality of the Friedman rule in first best and second best environments. The economy is populated by a large number of identical infinitely lived households, with preferences given by

$$\sum_{t=0}^{\infty} \beta^t V\left(c_t, \frac{M_t}{P_t}, h_t\right),$$

where $c_t$, $M_t$, $P_t$, and $h_t$ represent, respectively, consumption in period $t$, money balances held from period $t$ to period $t+1$, the price of the consumption good in units of money in period $t$, and leisure in period $t$. The utility function shares the usual assumptions of concavity and differentiability; it is increasing in $M_t/P_t$ as long as $M_t/P_t < m^*(c_t, h_t)$ and nonincreasing for $M_t/P_t \geq m^*(c_t, h_t)$. $m^*(c_t, h_t)$ is the satiation function that represents the “satiation level,” that is, the point where “cash balances ... are held to satiety, so that the real return from an extra dollar is zero” (Friedman, 1969). The characterization of the function $m^*(c_t, h_t)$ is crucial for the determination of “the optimum quantity of money,” as will be shown in the next section. Although Friedman (1969) does not fully characterize this point, the examples he presents imply that the point of
satiation is finite. Phelps (1973) also assumes that there is “full liquidity” or “liquidity satiation” when the nominal interest rate is not strictly positive and the demand for real balances is finite. The discussion in Brock (1975) on the optimum quantity of money is also for a finite satiation level. The technology of production of the private good and the public consumption good is linear with unitary coefficients.

The representative household (that implicitly solves the problem of the firm) chooses a sequence \((c_t, h_t, M_t, B_t)_{t=0}^\infty\), given a sequence of prices and income taxes, \((P_t, i_t, \tau_t)_{t=0}^\infty\), and initial conditions for \(W_0 = M_{-1} + (1 + i_{-1})B_{-1}\), to satisfy a sequence of budget constraints:

\[
P_t c_t + M_{t+1} + B_{t+1} \leq (1 - \tau_t) P_t (1 - h_t) + M_t + (1 + i_t) B_t, \quad t \geq 0
\]

\[
M_0 + B_0 \leq W_0,
\]

(2.2)

together with a no-Ponzi games condition. \(B_t\) is the number of bonds held from period \(t\) to period \(t+1\), and \(i_t\) is the nominal return on these bonds. \(1 - h_t\) is the labor supply.

The set of budget constraints can be written as a unique intertemporal budget constraint:

\[
\sum_{t=0}^{\infty} Q_t P_t c_t + \sum_{t=0}^{\infty} Q_t i_t M_t \leq \sum_{t=0}^{\infty} Q_t P_t (1 - \tau_t)(1 - h_t) + W_0,
\]

(2.3)

where \(W_0 = M_{-1} + (1 + i_{-1})B_{-1}\) and \(Q_t = 1/(1 + i_0)\ldots(1 + i_t)\).

In the competitive equilibrium the following marginal conditions must hold:

\[
\frac{V_c(t)}{\beta V_c(t+1)} = (1 + i_{t+1}) \frac{P_t}{P_{t+1}}, \quad t \geq 0
\]

(2.4)

\[(1 - \tau_t)V_{c_t} = V_{h_t}, \quad t \geq 0
\]

(2.5)

\[V_{m_t} = i_t V_{c_t}, \quad t \geq 0,
\]

(2.6)

and the resources constraints are

\[c_t + g_t = 1 - h_t, \quad t \geq 0,
\]

(2.7)

where \(g_t\) is the level of public spending in period \(t\. We assume that \(g_t\) is constant, \(g_t = g\). We proceed to characterize the optimal policy in this environment.
2.1. The Optimal Policy

The first best policy in this economy is given by the maximization of (2.1) subject to the resources constraint (2.7). In this solution \( V_c = V_{h}, \) and \( V_m = 0, \ t \geq 0. \) If the government could collect lump-sum taxes, it would be possible to decentralize this solution by setting a constant nominal interest rate equal to zero. It is clear that the Friedman rule is optimal simply because real balances are a free good, in the sense that they do not require resources to be produced. Since the social marginal cost of money is equal to zero, then the level of money balances that characterizes the first best is \( m^*, \) i.e., the level for which marginal utility is zero. The solution can be decentralized by setting the private marginal cost of holding real balances identical to zero, i.e., a zero nominal interest rate.

The optimal policy problem is more interesting when the government cannot levy lump-sum taxes. To determine the second best (Ramsey) solution, we construct the implementability constraint, substituting (2.4), (2.5), and (2.6) into the intertemporal budget constraint (2.3):

\[
\sum_{t=0}^{\infty} \beta^t (V_{c,t} - V_{h,t}(1 - h_t)) + \sum_{t=0}^{\infty} \beta^t V_{m,t} m_t = (V_{c,0} + V_{m,0}) \frac{W_0}{P_0}. \tag{2.8}
\]

The solution of the maximization of (2.1) subject to the implementability constraint (2.8) and the resources constraints (2.7), and given the initial nominal wealth, \( W_0, \) is the following: \( P_0 \) is set at an arbitrarily large number, and the first-order conditions for \( \{c_t, h_t, m_t\}_{t=0}^{\infty} \) are as follows:

\[
\beta^t V_{c,t} + \beta^t \psi [V_{c,t} + V_{c,h_t} c_t - V_{h,c_t}(1 - h_t) + V_{m,c_t} m_t] = \lambda, \quad t \geq 0 \tag{2.9}
\]

\[
\beta^t V_{h,t} + \beta^t \psi [V_{h,t} + V_{c,h_t} c_t - V_{h,h_t}(1 - h_t) + V_{m,h_t} m_t] = \lambda, \quad t \geq 0 \tag{2.10}
\]

\[
V_{m_t} + \psi [V_{m_t} + V_{c,m_t} c_t - V_{h,m_t}(1 - h_t) + V_{m,m_t} m_t] = 0, \quad t \geq 0, \tag{2.11}
\]

where \( \psi \) is the shadow price of the implementability constraint, i.e., it measures the marginal excess burden of government deficits in this second best world. \( \lambda \) measures the shadow price of resources.

This second best allocation can be decentralized using the instruments \( i_t \) and \( \tau_t, \ t \geq 0. \) Given (2.6), the discussion of whether the Friedman rule is optimal in this environment is equivalent to the discussion of whether \( V_m \) is zero, for \( t \geq 0. \) Conditions (2.9)–(2.11), together with the resources constraint (2.7) and the implementability constraint (2.8), define the sta-
tionary solution for $c_t, h_t, m_t, \psi$, and $\lambda_t/\beta^t$, for $t \geq 0$, since, from the competitive equilibrium condition $V_{m_t} = iV_c$, the solution for the nominal interest rate is stationary. If the Friedman rule holds, it holds for every period. So the issue is whether, for $t \geq 0, V_{m_t} = V_m = 0$ is a solution of the system of equations.

Because money is a free good, as is clear from the resources constraint (2.7), the multiplier of the resources constraint does not show up in condition (2.11). In this second best solution the social marginal benefit of using money, for the households, is equal to the marginal "excess burden," i.e., the marginal cost due to the fact that a change in $m$ affects the budget constraint of the government. This can be seen by rewriting (2.11) as

$$V_{m_t} = -\psi [V_{m_t} + V_{c,m}c_t - V_{h_m}(1 - h_t) + V_{m,m}m_t],$$

given that at the optimum, $\psi > 0$, the relevant issue is the determination of the sign of the term in parentheses, i.e., the impact on government revenue of an increase in $m$, holding the quantities of the other goods constant.

The following is an example of the general properties typically discussed in the literature. An increase in $m$ corresponds to a decrease in the nominal interest rate and has a negative impact on the revenue from seigniorage, $V_m m$. So $V_m + V_{mm}m \leq 0$. The sign of the expression depends then on the cross-derivatives. Suppose $V_{cm} < 0$ and $V_{hm} > 0$. In this case, the expression would be negative, meaning that an increase in $m$ has a negative effect on total government revenue. This means that the marginal "excess burden" would be strictly positive. The implication is that the Friedman rule would not be optimal.

The Friedman rule is optimal if the marginal excess burden of real balances is zero at the optimum. In Proposition 1 we identify conditions in which this is the case. These are local conditions at the point of satiation in real balances. We make the assumption, that we justify fully in Section 2.2, that the satiation point in real balances does not depend on leisure, but on the level of transactions only. This assumption is used to show the main proposition in the paper.

**Assumption 1.** The satiation point in real balances is a function of consumption only, $m^*(c)$.

---

3 The implementability constraint was constructed using the budget constraint of the households, but an equivalent constraint can be obtained using the government budget constraint. The marginal effect on the implementability condition corresponds to a symmetric marginal effect on the condition expressed in terms of the government budget constraint.
**Proposition 1.** In models with money in the utility function, the Friedman rule is the optimal policy when the satiation point in real balances is such that \( m^* = \infty \) or \( m^* = \bar{k}c \), where \( \bar{k} \) is a positive constant.

**Proof.** We verify whether Eq. (2.11) is satisfied when \( V_m = 0 \). \( m^* \) was assumed to be a function of \( c \) only, therefore at the satiation point, \( V_{hm} = 0 \).

When \( m^* = \infty \), we must have \( V_{mn}(c, m^*) = 0 \) and \( V_m(c, m^*) = 0 \). We assume that, when \( V_m = 0 \), and therefore the nominal interest is zero, the inflation tax revenue, \( V'_m m \), is zero. Any reasonable specification of a model with money in the utility function must have this property. Therefore \( \lim_{m \to \infty} \frac{\partial (V_m m)}{\partial m} = 0 \), or \( \lim_{m \to \infty} (V_m + V_{mn} m) = 0 \). Then \( m^* = \infty \) verifies the equation.

When the satiation point in real balances is finite, we find from \( V_m(m^*, c) = 0 \) that

\[
\frac{V_{cm}(c, m^*)}{V_{mn}(c, m^*)} = -\frac{dm^*}{dc}.
\]

So, expression (2.11) evaluated at the satiation point in real balances can be written as

\[
V_m(c, m^*)\left[1 + \psi\right] = -\psi V_{mn}(c, m^*)m^*\left[1 - \frac{dm^*}{dc} \frac{c}{m^*}\right],
\]

where \( V_m(c, m^*) = 0 \).

When \( m^* = \bar{k}c \), we have that \((dm^*/dc)(c/m^*) = 1\), and so \( V_{cm}(c, m^*)c + V_{mn}(c, m^*)m = 0 \). Therefore \( m = \bar{k}c \) also satisfies the Ramsey first-order condition, (2.11).

Notice that when \( V \) is separable in leisure and homothetic in consumption and real balances, then the ratio of real balances and consumption is a function of the interest rate alone, and therefore the relevant elasticity is unitary. From Proposition 1, the Friedman rule is optimal. These are the sufficient conditions for the optimality of the Friedman rule identified by Chari et al. (1996).

In summary, we have shown in this section that the first best and the second best optimal inflation tax rules coincide, when the satiation point in real balances is infinite or when it is characterized by a unitary elasticity with respect to consumption. The reason for this is that in those cases the increase in real balances has a zero effect on government revenues at the satiation point, and consequently the marginal costs are equal to zero in both problems. In the next section we go beyond the reduced form of
models with money in the utility function to inquire, first, how these local properties can be justified, and second, whether it can be argued that the case of the unitary elasticity of real balances at the satiation point is the relevant case.

2.2. Equivalence with a Transactions Technology Model

In this section we show that if the models with money in the utility function are seen as reduced forms of transactions technology monetary models, then the preference specifications must be restricted. These restrictions correspond to the conditions under which the Friedman rule is optimal.

Feenstra (1986) shows that it is possible to establish an equivalence between a monetary model with a transactions technology, where the preferences depend only on consumption net of transaction costs, and a model with money in the utility function. He establishes a correspondence between the set of assumptions characterizing the transactions technology and the assumptions on the utility function expressed as a function of real balances. Here we extend Feenstra’s (1986) results to a world where the original preferences are defined over consumption and leisure, $\sum_{t=0}^{\infty} \beta^t U(c_t, h_t^w)$, and the transaction costs are measured in units of time. The transactions costs function is represented by $s = l(m, c)$, where $s$ is the time spent in transactions. This is the type of shopping time specification of Mccallum (1983).

The maximization problem is as follows:

**Problem 2.** Choose $(c_t, M_t, B_t, h_t^w, s_t)_{t=0}^{\infty}$ to maximize

$$\sum_{t=0}^{\infty} \beta^t U(c_t, h_t^w),$$

subject to

$$P_t c_t + M_{t+1} + B_{t+1} \leq (1 - \tau_t) P_t n_t + M_t + (1 + i_t) B_t, \quad t \geq 0$$

$$s_t = l(c_t, m_t)$$

$$1 = h_t^w + n_t + s_t$$

and $M_0 + B_0 \leq W_0$.

The transactions technology is characterized by the following assumption.

---

$^4$ Mccallum (1990) constructs (footnote 7) the indirect utility function associated with his type of shopping time costs. However, he does not derive the properties of this utility function.
Assumption 2. The transactions costs function $s = l(m, c)$ has the following properties:

(a) $s \geq 0$, $l(m, 0) = 0$.
(b) $l_c \geq 0$.
(c) $l_{cc} \geq 0$, $l_{mm} \geq 0$.
(d) $l_m(m, c) = 0$ defines $m = \bar{m}(c)$, $l_m(m, c) < 0$ when $m < \bar{m}$.
(e) $l$ is restricted to ensure that Problem 2 is a concave problem.

The following are two first-order conditions of Problem 2:

$$\frac{U_t(t)}{U_{h^t}(t)} = \frac{1}{1 - \tau} + l_c(t), \quad t \geq 0$$

$$-l_m(t) = \frac{1}{1 - \tau} i_t, \quad t \geq 0.$$

The optimal choice of $m$ is such that the private agents choose the point where the private value of using money, $-l_m(1 - \tau)$, is equal to its opportunity cost, $i$. The implementation of the Friedman rule, $i = 0$, implies that $l_m = 0$.

We call the private problem defined in the last section Problem 1. The following proposition states the equivalence between the two problems.

**Proposition 2.** Given a monetary model with an explicit transactions costs function (Problem 2), it is possible to construct an equivalent model with money in the utility function (Problem 1), where $h_t = h_t^u + l(c_t, m_t)$ and $V(c_t, m_t, h_t) = U(c_t, h_t - l(c_t, m_t))$. If Assumption 2 is satisfied in Problem 2, then $V(c_t, m_t, h_t)$ is characterized by the following conditions:

(a) $V$ is concave and so $V_{cc} \leq 0$, $V_{mm} \leq 0$.
(b) $V_m(c, m, h) \geq 0$.
(c) $V_{mm}(c, m, h) \geq 0$ and $V_{mm}(c, m, h) = 0$ at the satiation point $V_m = 0$.
(d) $\bar{m} = m^*$ are identical functions of $c$ only.

**Proof.** The equivalence is for a given pair of functions $(U, l)$. The solution of Problem 2 is the vector $(\hat{c}, \hat{m}, \hat{h})$. Then there is a function $V$ such that $(\hat{c}, \hat{m}, \hat{h} = h_t^u + l(\hat{c}, \hat{m}))$ solves Problem 1.

Conditions (a)–(d) are satisfied since

(i) From the concavity of Problem 2, $V$ is concave, $V_{cc} = U_{cc} - 2l_c U_{ch_t} + l_c^2 U_{h_t h_t} - l_{cc} U_{h_t} \leq 0$, and $V_{mm} = -U_{mm} + l_m^2 U_{h_t h_t} \leq 0$.
(ii) $V_m = -U_{h^m}l_m \geq 0$. Therefore $V_m = 0$ if and only if $l_m = 0$.

(iii) $V_{mh} = -U_{h^m}l_m - U_{h^m}l_{mh} = -U_{h^m}l_m \geq 0$ and (ii) both imply that $V_{mh} = 0$, at the point $V_m = 0$.

(iv) $V_m = 0$ defines $m^*(c)$ and $l_m = 0$ defines $\overline{m}(c)$. So, from (ii), $\overline{m} = m^*$. 

As we discussed in the last section, the determination of whether the Friedman rule is optimal in a model with money in the utility function depends crucially on the functional form of the function $m^*(c, h)$, i.e., the function defined by setting the marginal utility of money equal to zero. By the equivalence result we verify that the properties of this function are identical to the properties of the function defined by $l_m(c, m) = 0$. From this, it results that

1. We can justify the hypothesis made in Section 2.1 that leisure is not an argument of the function $m^*$.

2. When the transactions costs function is homogeneous of degree $q$, $l_m$ is homogeneous of degree $q - 1$ and can be represented by

$$l_m = L(m/c)c^{q-1}, \quad (2.12)$$

and, at the point $l_m = 0$, $L(m/c) = 0$ defines $m = m^*(c) = \overline{m}c$.

So, homogeneous transactions costs functions correspond to the case described in Section 2.1, where the function $m^*$ has unitary elasticity and the Friedman rule is always optimal. In fact, as was shown by Correia and Teles (1996), in monetary models with explicit transactions technologies, the Friedman rule is the Ramsey solution for homogeneous transactions technologies.

3. When the transactions costs function, $s = l(c, m)$, is associated with a transactions production function $c = f(s, m)$ which verifies Inada conditions, $l_m = 0$ is equivalent to $m^* = \infty$. In this case at the point $l_m = 0$, we have $l_{mm} = l_{mc} = 0$. Again the marginal condition of the second best problem is satisfied for $l_m = 0$, and so the Friedman rule is optimal. This corresponds to the case of $m^* = \infty$, described in Section 2.1.

5. The case of elasticity of $m^*$ lower or greater than one corresponds to the case of nonhomogeneous transactions costs functions. We do not know of any work where it is argued on theoretical or empirical grounds that the transactions costs function ought to be nonhomogeneous. At the theoretical level the microfoundations of this function are obtained from Baumol (1952) and Tobin (1956), from the generalization of Barro (1976), from Guidotti (1989) or from Jovanovic (1982). All of these forms are homogeneous of degree zero. In Marshall (1992) the proposed and estimated transactions costs function is homogeneous of degree one.
Braun (1994) estimates the degree of homogeneity of the transaction cost function to be 0.98.

In summary we can conclude that once the equivalence between models with money in the utility function and explicit transactions technologies models is established, the local properties used in Section 2.1 to obtain the optimality of the Friedman rule are associated with global properties of the transactions costs functions. Besides, these global properties are not restrictive, since they include all homogeneous transactions costs functions.\(^5\)

In Section 2.3 we provide further arguments in favor of the robustness of the Friedman rule. Even if the elasticity conditions for the optimality of the Friedman rule are not met, i.e., the implicit transactions technology is not homogeneous, the calibrated results are still very close to that prescription.

2.3. How Far Can the Optimal Policy Deviate from the Friedman Rule?

In the previous sections we have shown that the local conditions for the optimality of the Friedman rule in a model with money in the utility function are general, once an equivalence is established between the model with money in the utility function and a transactions technology model. In any case, if those conditions are not met, it is important to know the magnitude of the optimal inflation tax. In this section we compute the optimal policy for calibrated examples where the relevant elasticity is different from one.

When the elasticity is higher than one, \(V_{c m}(c, \mu^*)c + V_{\mu m}(\mu, \mu^*)\mu^* > 0\). If we assume that the marginal benefit of real balances is always nonnegative, then for all of the preference specifications that we have used, the optimal allocation is \(V_m = 0\), corresponding to the Friedman rule.

We have performed a numerical analysis for a class of preferences specification where the elasticity of the satiation function is strictly lower than one, specifically for the limit case, where the elasticity of \(\mu^*\) is equal to zero. The results are depicted in Figs. 1–4. The calibration is made, using as borderline cases the examples of Calvo and Guiddoti [7] and Lucas [26]. The instantaneous utility functions are additively separable: \(U = c_t + H(h_t) + v^i(m_t)\), where \(H(h_t) = h_t - (E/2)(h_t)^2\) and \(i = 1, 2\). We consider

---

\(^5\) The introduction of capital will not alter that result. Suppose that the transactions costs function is defined as before. The introduction of capital as an input in the production of the consumption good has no consequence in the Ramsey solution that defines the optimal choice of \(\mu\). The main difference is that now the marginal utility of labor is not constant and depends on the level of the stock of capital. The Ramsey solution will have a transitional period, but \(l_m = 0\), and consequently \(V_m = 0\), also characterize that transition. If the transactions costs function is modified in a way that capital is also an input in the production of transactions, the results are maintained once we impose that the transaction costs function is homogeneous in consumption, real balances, and capital.
two possible $\nu^i$ functions: $\nu^1(m_i) = m_i(B - D \ln(m_i))$ and $\nu^2(m_i) = -A[1/m_i + m_i/\bar{k}^2]$. $A$, $B$, $D$, $E$, and $\bar{k}$ are parameters, and $\bar{k}$ represents the constant satiation level in real balances. Government expenditures are set at $g = 0.15$. The first utility function, $\nu^1$, is initially calibrated with the numbers provided by Calvo and Guidotti (1993), $B = -0.65$, $D = 0.5$, $E = 1$. Figure 1 shows the resulting welfare cost of the inflation tax, in units of consumption, as well as the corresponding equilibrium schedule for the ratio of real balances to income, as a function of the nominal interest rate. The U.S. line is the log-linear $M1$ to $NNP$ schedule estimated by Lucas (1994) for U.S. data (elasticity is 0.5). The optimal nominal interest rate is large, around 10%, but notice that the calibrated money–income ratio and interest rate schedule are not consistent with the U.S. data. Figure 2 shows the same curves for a different calibration that better fits the U.S. money demand schedule: $B = -0.046$, $D = 0.1429$. The semielasticity is now 7. The optimal nominal interest rate is considerably smaller. Figures 3 and 4 still represent the same two curves, for the
FIG. 3. Welfare cost of the inflation tax and money–income ratio, when the utility from real balances is $u^2(m) = -\lambda [1/m + m/k^2]$, calibrated to fit U.S. data. $\lambda = 1$ is the satiation level in real balances. The curve U.S. is the fitted schedule of the $M1/NNP$ ratio on long-term interest rates, for U.S. data.

second preferences specification, $u^2$. With this utility function, in the limit, for an arbitrarily large $\lambda$, the real balances-to-consumption ratio, as a function of the nominal interest rate, is a log-linear schedule. For $\lambda = 1$, the optimal nominal interest rate is smaller than 0.1% (Fig. 3). Figure 4 shows the optimal nominal interest rate when $\lambda = 0.4$, generating levels of velocity at the satiation point that have been observed for considerably higher nominal interest rates. The optimal nominal interest rate is less than 1%. The conclusion is that for reasonable levels of $\lambda$, the Friedman rule is a very good approximation to the optimum.

FIG. 4. Welfare cost of the inflation tax and money–income ratio, when the utility from real balances is $u^2(m) = -\lambda [1/m + m/k^2]$, $\lambda = 0.4$ is the satiation level in real balances. The curve U.S. is the fitted schedule of the $M1/NNP$ ratio on long-term interest rates, for U.S. data.
3. MONEY IS A FREE GOOD

In Section 2, we have shown that the Friedman rule is optimal when there is no effect on government revenues of changing real balances from the full liquidity level. The arguments for the second best taxation rules of final goods in the public finance literature are different from these. There, the optimal taxes on different goods depend on the comparison of the respective marginal effects on government revenues. In particular, for it to be optimal not to tax final goods, when the alternative choice is an income tax, the marginal effect on government net revenues of a change of one unit of labor used to produce any of the goods should be equal. Atkinson and Stiglitz (1972) derived conditions under which this is the optimal rule.

Our results show that the conditions on preferences to obtain the optimality of the Friedman rule are more general than the ones derived in Atkinson and Stiglitz (1972) and therefore extend the result of Chari et al. (1996) that identified those conditions as sufficient conditions for the optimality of the Friedman rule.\(^6\) The homotheticity and separability conditions of Atkinson and Stiglitz (1972) correspond to utility functions where the marginal rate of substitution between consumption and real balances depends only on the ratio of these two variables. This is one example of the conditions in Proposition 1. The conditions in Proposition 1 are much less restrictive though, since they must hold only in the neighborhood of \(i = 0\).

We now show how the very appealing argument of the distribution of distortions among different goods in the economy can be reconciled with the zero inflation tax result derived for monetary economies.

What distinguishes money from any other consumption good is the fact that an additional unit of real money does not require relevant marginal resources. We think that this is the right way of describing fiat money. In the following exercise we analyze how the taxation rules are affected when the cost of producing a good \(m\) becomes arbitrarily small. Consider a stationary real economy corresponding to the monetary economy we have studied, but for the fact that \(m\) is now a consumption good that is produced with \(\alpha\) units of time, which implies that the price of \(m\) in units of the other consumption good \(c\) is \(\alpha\). The ad valorem tax on \(m\) is \(\tau^m\). The budget constraint of the households is written as

\[
c + \alpha(1 + \tau^m)m = (1 - \tau)(1 - h).
\]

Notice that the equivalent unit tax would be \(T^m = \alpha\tau^m\) (the corresponding equation in the monetary economy is \(c + \imath m = (1 - \tau)(1 - h)\)).

\(^6\) As they point out, the conditions are not necessary.
In the real economy with ad valorem taxation, the optimal taxation rules of Ramsey (1927) and Atkinson and Stiglitz (1972) apply. Atkinson and Stiglitz (1972) derive sufficient conditions for optimality of uniform taxation of consumption goods. When the preferences are homothetic in $c$ and $m$ and separable in leisure, then a tax on labor income and a zero ad valorem tax on both $c$ and $m$ decentralize the second best (Ramsey) solution. This corresponds to $\tau^m = 0$. If instead of the ad valorem tax, a unit tax was used, as is the case with money, then this unit tax is always equal to zero at the optimum. Now suppose that the alternative tax is a tax on $c$, $\tau^c$. The budget constraint is written as

$$(1 + \tau^c)c + \alpha(1 + \tau^m)m = 1 - h.$$ 

The second best allocation for the choice of an inflation tax and an income tax can be decentralized using a consumption tax, equal to the ad valorem tax on money, $\tau^m = \tau^c$. The equivalent unit tax $T^m = \alpha \tau^m$ is positive as long as the cost of producing the good is positive, but the limit is zero, when $\alpha$ converges to zero. This same result applies as long as the optimal ad valorem tax converges to a finite number. This holds whenever, according to the rules derived for the monetary economy, the marginal impact of $m$ on the government revenue, at the satiation point in real balances, is equal to zero.

There is a sense in which the application of the conditions of Atkinson and Stiglitz (1972) to this problem is misleading. When the choice is the optimal mix of an income tax and an inflation tax, then the result that the inflation tax should be zero could be interpreted as a direct application of Atkinson and Stiglitz (1972). However, that same solution is equivalent to a tax on consumption and a zero unit tax on money. Atkinson and Stiglitz (1972) conditions still hold for the equivalent ad valorem tax on money rather than for the relevant unit tax.

In any case, according to the optimal taxation rules derived for the monetary economy (so, for the case where $\alpha$ goes to zero), when the impact of a marginal increase in $m$ on government revenues is not zero, then the optimal unit tax can be positive. This corresponds to an optimal ad valorem tax that becomes arbitrarily large as the cost of producing $m$ is made arbitrarily small.

4. CREDIT GOODS

In this section we compare the results in the model with money in the utility function with the results in models with credit goods. It is well known that it is possible to establish an equivalence between the two
models, by replacing in the model with money in the utility function total consumption with the sum of the consumptions of the two goods and real balances with the consumption of the cash good. A condition that ensures that real balances are smaller than total consumption guarantees nonnegativity of consumption of the credit good.

Chari et al. (1996) identify as sufficient conditions for optimality of the Friedman rule in the cash–credit goods model the conditions of homotheticity in the two goods and separability in leisure. The explanation for this result is simple. Suppose that real money was costly to produce and consider the typical structure in a cash–credit goods model, where consumption of the cash good requires time and real money in fixed proportions. The production functions of the two goods and of real money are linear. It is a direct application of Atkinson and Stiglitz (1972) that under separability and homotheticity, the two goods should be taxed at the same rate. Under the Leontief production structure, a positive tax on money would not distort the production of the consumption good, but would distort the relative consumptions of the two goods. Therefore if the alternative taxes are a tax on income or a uniform tax on consumption, then it is optimal to set the tax on real balances (ad valorem or unit) to zero.

Separability in leisure and homotheticity in the cash and credit goods imply separability in leisure and homotheticity in real balances and total consumption in the equivalent model with money in the utility function. As was seen before, these conditions imply unitary elasticity at full liquidity, and therefore the Friedman rule is optimal.

If the utility function is not homothetic in the two goods, then the inflation tax could be used as a means of achieving the optimal distortion in the two goods. In this case the relevant elasticity is not unitary. The optimal taxation issue here is very different from the one considered before. The issue here is the determination of the optimal distortion between consumption goods. With enough taxation instruments, this issue would not even be present, as seen in Lucas and Stokey (1983).

---

7 If the transactions technology allowed for substitutability between real balances and time, then the distortion of the consumption of the two goods would also imply a distortion in production. Clearly, in this case, it would be preferable to discriminate between the consumption taxes on the two goods.

8 Notice that, under this specification, real money balances must be equal to the consumption of the cash good, and therefore they can never be made arbitrarily large. So the point of satiation cannot be infinity, which is one of the sufficient conditions for the Friedman rule to be optimal.
5. ALTERNATIVE TAXES AND WELFARE CRITERIA

In Section 2 we have constructed a second best environment assuming that there were two alternative taxes, an inflation tax and a tax on labor income. For the purpose of checking the robustness of the results, in this section we discuss the implications of considering a consumption tax instead of the labor income tax. In addition, we will assess the implications of considering alternative timing conventions and welfare criteria in the specification of the second best problem.

5.1. Consumption Taxes

When the level of transactions is measured by consumption net of taxes, the tax on consumption, \( \tau_c \), does not affect the transactions costs function, i.e., \( s = l(c, m) \). In this case the indirect utility function \( V(c, m, h) \) associated with the pair \(( U, l)\) is the same as described in Section 2. As we saw in Section 3 the second best allocation coincides with the one obtained when the alternative tax is an income tax. So the conditions under which the Friedman rule is optimal are the same whether the alternative tax is a tax on labor income or a tax on consumption. In summary, the irrelevance result of the alternative tax in money in the utility function models is extended to models of explicit transactions costs functions, given the equivalence established in Section 2.

A number of authors claim that the introduction of a consumption tax should modify the transactions costs function, in the sense that the amount of transactions ought to be measured by consumption gross of taxes: \( s = l(c(1 + \tau_c), m) \). This introduces some changes. Adopting the same procedure as in Section 2, it is clear that the pair \((U, l)\) corresponds to a utility function \( V\) such that

\[
U(c, h - l(c(1 + \tau_c), m)) = V(c, m, h, (1 + \tau_c)).
\]

Now preferences depend on the tax parameter \( \tau_c \). Under this formulation the Friedman rule is optimal when \( m^* = \infty \). It is also optimal when the elasticity of \( m^*(c(1 + \tau_c)) \) is unitary, if in addition we impose that, at the satiation point, the underlying technology is characterized by either \( l_c(1 + \tau_c) = 0 \) or \( l_c(1 + \tau_c, h, m) = 0 \).

De Fiore and Teles (1998) show that the additional conditions are necessary because the transactions technology has the undesirable property that it is possible to reduce the time used for transactions, without changing real consumption and real money used to buy it, by reducing the tax on consumption. When either the transactions technology does not
have that property or when income taxes are allowed together with consumption taxes, then those additional conditions are no longer necessary.

5.2. Alternative Timing Conventions and Welfare Criteria

It is a standard view (see Woodford (1990), p. 1092) that alternative timing conventions in the decisions of the private agents affect the result of optimality of the Friedman rule. In the previous sections, the private agents are assumed to choose financial assets, in each time period, so that the resulting money balances can be used for transactions that same period. This is the timing assumed by Lucas (1982) and Lucas and Stokey (1983).

Alternatively, Woodford (1990) assumes, in line with Svensson (1985), that the money balances that can be used in any one period are decided the period before. The implication of this timing is that there are real effects of unanticipated monetary shocks. In the beginning of time, period zero, the agents cannot adjust the portfolios, and therefore it is no longer optimal for the benevolent government to completely deplete the real value of outstanding monetary balances. The allocation is stationary from period one on, but in period zero the levels of the variables are, in general, different from the corresponding stationary levels. The stationary optimal allocation from period one on corresponds to the Friedman rule. Woodford (1990), for the sake of tractability, and Braun (1994) and Lucas (1994) propose a third best solution concept: the maximization of welfare restricted to the solution being stationary, i.e., the problem is restricted so that the allocation in period zero is the same as from period one on.

When the stationarity restriction is imposed, the government faces a trade-off between the low level of initial real balances and the high steady-state level. It is intuitive that from this trade-off there results a stationary level of real balances higher than the initial optimal level of the Ramsey solution and lower than the high steady-state level of the same solution. The solution is characterized by less than full liquidity and a strictly positive nominal rate of interest.

---

9. Asset markets open in the beginning of period, so that money balances used for transactions in any period are beginning-of-period balances. Kimbrough (1986) and others assume that end-of-period money balances are used for transactions that same period.

Lucas (1994) has computed numerically, using this criterion, the optimal policy in a transactions technology model. He concludes that the optimal nominal interest rate, although strictly positive, is very close to zero.

6. CONCLUDING REMARKS

In this paper we compute the second best inflation tax rule in models where real balances are an argument in the utility function. We identify local conditions that extend the global conditions of separability and homotheticity in Chari et al. (1996) as sufficient conditions for the optimality of the Friedman rule. Furthermore, we establish an equivalence between the models with money in the utility function and more fundamental models of transactions technologies and show that the wide class of transactions technologies where the Friedman rule is optimal satisfy the local conditions for the optimality of the Friedman rule in the money in the utility function specification.

The characteristic of real balances that is determinant for the general optimality of the Friedman rule is the fact that money is a free good, meaning that the production cost of money is zero, i.e., the production possibilities in these economies are not affected by a change in the quantity of money. For this reason, the usual intuition of the comparison of the marginal excess burdens of alternative taxes that give the same revenue no longer applies. The optimal decision here consists of the following comparison: an increase in the quantity of money generates a benefit for the households in terms of utility and a cost equal to the value of the marginal effect on government net revenues. At the point of satiation in real balances, the marginal utility benefit is by definition equal to zero. We show that under reasonable preferences specifications, the marginal impact on government net revenues is also equal to zero, at that point of full liquidity. Therefore the Friedman rule is optimal. In less adequate specifications for preferences, in terms of its microfoundations, where the Friedman rule is not optimal, it is nevertheless very close to the optimum. These are cases where the optimal implicit ad valorem tax is infinity.

The main conclusion of this paper is that the optimal taxation results in monetary models are much more robust than the public finance results derived in other economic environments. In particular, the Friedman rule, i.e., a zero inflation tax, is a general result for monetary economic structures with reasonable microfoundations. This normative result has no counterpart in the public finance literature where the optimal policies depend on the structure of preferences and technologies. The neat and
A successful practical recommendation of Friedman is reinforced now that it is shown that its optimality extends to a second best environment.

REFERENCES


