

**BANCO DE PORTUGAL**  
**Economics Research Department**

**The Optimal Mix of Taxes on Money,  
Consumption and Income**

*Fiorella De Fiore*  
*Pedro Teles*

WP 2-99

March 1999

The analyses, opinions and findings of this paper represent the views of the authors, they are not necessarily those of the Banco de Portugal

---

Please address correspondence to Pedro Teles, Research Department, Banco de Portugal, Av. Almirante Reis, nº 71, 1150-012 Lisboa, Portugal:  
Tel.#351-1-3130000; Fax#351-1-3143841; e-mail:pmpteles@bportugal.pt.

# The Optimal Mix of Taxes on Money, Consumption and Income\*

Fiorella De Fiore  
European University Institute

Pedro Teles<sup>†</sup>  
Banco de Portugal  
Universidade Católica Portuguesa

August 31, 1998

## Abstract

We show that the Friedman rule is optimal in the standard transactions technology monetary model whether the alternative fiscal instrument is a consumption tax, an income tax, or both, and whether taxes are paid with money or not. These results are at odds with recent literature. We show that an incorrect specification of the transactions technology explains the divergence.

*Key words:* Friedman rule; Inflation tax; Transactions technology.

*JEL classification:* E31; E41; E58; E62

---

\*We thank Isabel Correia for very insightful comments. Remaining errors are ours. Teles gratefully acknowledges financial support by PRAXIS XXI.

<sup>†</sup>Corresponding author: Pedro Teles, DEE, Banco de Portugal, R. Francisco Ribeiro, 1150 Lisboa, Portugal; Tel.# 351-1-3128397; Fax#351-1-3143841; pmpteles@bportugal.pt

# 1 Introduction

Recent literature on the optimal inflation tax has argued that it is optimal not to tax real balances, also in a second best world where only distortionary taxes can be raised to finance government expenditures. Since the marginal production cost of money is commonly assumed to be zero, real balances should then be priced at zero. This recovers the optimal policy rule of zero nominal interest rates, derived by Friedman (1969) in a world with lump sum taxes. It is also at odds with Phelps (1973), where it is argued that liquidity should be taxed as any other good.

The argument for the optimality of the Friedman rule is twofold. First, money should not be taxed because it is an intermediate good, as Kimbrough (1986) pointed out. However this argument is only valid if the technology is constant returns to scale. If the transactions technology is not constant returns to scale, as in the case of the Baumol-Tobin technology<sup>1</sup>, then it is optimal not to tax money only because money is a free good. This is shown in Correia and Teles (1996).

If money was not a free good, the optimal tax on money would be a direct application of the standard optimal taxation rules in the Public Finance literature, as in Ramsey (1927) and Diamond and Mirrlees (1971). When money is modelled as a final good, it should be taxed, as any other good. When instead money is modelled as an intermediate good, it should not be taxed when the technology is constant returns to scale. However, because money is a free good it should not be taxed under general conditions. The intuition is that even if the optimal ad-valorem tax rate on a costly good is positive, the optimal price is in general zero, as the cost approaches zero.

In the standard transactions technology monetary model, as in Kimbrough (1986), Guidotti and Végh (1993) and Chari, Christiano and Kehoe (1996), Correia and Teles (1996) show that, if the government chooses to finance government expenditures with either the inflation tax or an income tax, then the Friedman rule is optimal, for homogeneous transactions technologies of any degree.

If the alternative tax to inflation is a tax on consumption, the conditions under which the Friedman rule is optimal depend on whether taxes are paid with money. Correia and Teles (1997) and Mulligan and Sala-i-Martin

---

<sup>1</sup>V. Baumol (1952) and Tobin (1956)

(1997) show that, when the consumption taxes are not paid with money, the Friedman rule is also optimal for homogeneous transactions functions of any degree. In this case, whether the government uses income or consumption taxes does not affect the optimal inflation tax. The two taxes are equivalent.

When instead it is rightly assumed that the consumption taxes are paid with money, the equivalence between the income tax and the consumption tax is lost. Mulligan and Sala-i-Martin (1997), in line with Guidotti and Végh (1993), argue that, in this case, homogeneity of the transactions technology is not enough to guarantee the optimality of the Friedman rule. The Friedman rule is optimal only if additional *restrictive* assumptions are made on the time spent on transactions at full liquidity. According to Mulligan and Sala-i-Martin (1997), this finding turns the result of optimality of the Friedman rule for all homogeneous functions into a *fragile* result. Both in Guidotti (1993) and Mulligan and Sala-i-Martin (1997), the intuition for why the results under the two alternative taxes differ remains obscure.

In this paper, we show that as long as income taxes are available, together with consumption taxes, the *restrictive* conditions are not necessary. We also argue that the reason for the differing results under the two alternative taxes is that the transactions technology is not well specified when there are consumption taxes and taxes are paid with money. Indeed, under a correct specification of the transactions technology, the Friedman rule is again optimal, for all homogeneous functions. We conclude that the optimality of the Friedman rule is indeed a robust result.

The paper proceeds as follows: In Section 2, we derive the conditions under which the Friedman rule is optimal, when the available tax instruments are the inflation tax and a tax on consumption, and taxes must be paid with money. We find that these conditions are different from those derived by Guidotti and Végh (1993) and Mulligan and Sala-i-Martin (1997). In Section 3 we show that the conditions for the optimality of the Friedman rule, are conditions under which it is not possible to affect time used for transactions, and as a result, it is also not possible to affect profits. We also show that, if money was not a free good, it would always be optimal to tax real balances. It is only because money is a free good, that we obtain conditions for optimality of the Friedman rule, however restrictive they may be.

In Section 4, we allow for the use of both alternative tax instruments, consumption and income taxes, and find that the Friedman rule is again the optimal tax on money for homogenous transactions technologies of any

degree. However, the optimal tax rules for the other two taxes are disturbing: The government should tax income and fully subsidize consumption in order to reduce the value of transactions and therefore save on resources spent on transactions. These results suggest a misspecification of the transactions technology.

In Section 5, we show that the transactions technology, as is specified in the literature, has the undesirable property that it is possible, by manipulating the tax on consumption, to reduce the time used for transactions without actually reducing either the real quantity of the goods transacted, or the quantity of money necessary to purchase these goods. This property is responsible for the additional *restrictive* conditions for optimality of the Friedman rule, when the consumption tax is the single alternative tax to seignorage. In fact, in that case, there is a trade-off generated by the incentive to save on resources used for transactions, so that a slightly higher inflation tax and a lower consumption tax will allow to save on those resources. Under a correct specification of the transactions technology, the Friedman rule is optimal for all homogeneous transactions technologies. Section 6 contains the conclusions.

## 2 The Optimal Combination of the Inflation Tax and the Consumption Tax

The environment is a transactions technology, monetary model as in Chari, Christiano and Kehoe (1996), Correia and Teles (1997) and Mulligan and Sala-i-Martin (1997). The model is a slight variation of the model in Kimbrough (1986), Guidotti and Végh (1993) and Correia and Teles (1996) in that the timing of the cash in advance constraint is different. This distinction is not relevant for the results. We now proceed to compute the optimal fiscal policy, when the available tax instruments are the inflation tax and a tax on consumption, and taxes must be paid with money.

In this economy, there is a large number of identical households, whose preferences are defined over a consumption good and leisure. Each household is endowed with one unit of time that can be allocated to labor, leisure, or transactions. The consumption good is produced using a linear technology with a unitary coefficient. There are two assets, money and nominal

bonds. The government must finance a given constant level of government expenditures with revenues from a consumption tax or from the inflation tax.

In this economy, the households seek to maximize

$$\sum_{t=0}^{\infty} \beta^t U(c_t, h_t), \quad (1)$$

where  $U$  is an increasing concave function,  $c_t$  are consumption goods and  $h_t$  is leisure at time  $t$ . The households supply labor  $1 - h_t - s_t$ , where  $s_t$  is time spent in transactions.

Transactions are costly since they require time that could otherwise be used for production. The amount of time devoted to transactions increases with consumption and decreases with real money balances, according to the following transactions technology:

$$s_t \geq l\left(e_t, \frac{M_t}{P_t}\right), \quad (2)$$

where  $e_t = c_t(1 + \tau_{ct})$  denotes consumption expenditures, gross of taxes.

Let  $m_t = \frac{M_t}{P_t}$ . We assume that the function  $l$  is homogeneous of degree  $k$  and so it can be written as  $l(e, m) = L\left(\frac{m}{e}\right)e^k$ .  $L$  is characterized by the following conditions:  $L : A \rightarrow R^+$ ,  $A \subseteq R^+$ ;  $L' \leq 0$ ,  $L'\left(\frac{\bar{m}}{e}\right) = 0$ , when  $\frac{\bar{m}}{e} = \sup A$ ;  $L'' \geq 0$ .  $\frac{\bar{m}}{e}$  is the point of *full liquidity*, where an additional unit of real balances does not reduce transactions time. Notice that with a consumption tax lower than  $-1$ , expenditures could become negative. Since the problem is well specified only when expenditures are positive, we restrict the tax on consumption to be  $\tau_{ct} \geq -1$ . Increasing returns cannot be too high, or the problem will not be concave. This implies restrictions on the minimum value of the degree of homogeneity of the transactions technology,  $k$ , that depend on the curvature of the utility function. In particular, we restrict  $k \geq 0$ , because otherwise  $l_e$  would be negative at *full liquidity*, for a subset of the class of transactions technologies<sup>2</sup>.

In each period  $t$ , the households choose holdings of money  $M_t$ , to be used for transactions in that same period, and nominal bond holdings  $B_t$ . These bonds entitle the households to  $(1 + i_t)B_t$  units of money in period  $t + 1$ .

---

<sup>2</sup>From the homogeneity assumption we know that  $l_e = e^{k-1} [kL(t) - \frac{m}{e}L'(t)]$ . Since  $L'(t) = 0$  at full liquidity, when  $L(t)$  is strictly positive at this point, then  $k$  must be positive for  $l_e$  to be positive.

Consumption is taxed at a rate  $\tau_{ct}$ . The budget constraints for  $t \geq 0$  are given by conditions

$$(1 + \tau_{ct})P_t c_t + M_{t+1} + B_{t+1} \leq M_t + (1 + i_t)B_t + P_t(1 - h_t - s_t), \quad (3)$$

$$M_0 + B_0 \leq W_0,$$

together with a no-Ponzi games condition. For simplicity we assume that  $W_0 = 0$ . If the initial nominal wealth was strictly positive it would be optimal to set the initial nominal price level to an arbitrarily large number, which would fully devalue the initial real wealth.

The private problem is defined by the maximization of (1), subject to (2) and to (3). The first order conditions include

$$\frac{U_c(t)}{U_h(t)} = (1 + \tau_{ct})(1 + l_e(t)), \quad (4)$$

$$-l_m(t) = i_t, \quad (5)$$

$$\frac{U_h(t)}{\beta U_h(t+1)} = 1 + r_{t+1}, \quad (6)$$

where  $1 + r_{t+1} \equiv (1 + i_{t+1}) \frac{P_t}{P_{t+1}}$ . The optimal choice of real balances requires that the private marginal value of using money,  $-l_m$ , is equal to its opportunity cost,  $i_t$ . The implementation of the Friedman rule,  $i_t = 0$ , implies that  $l_m = 0$ .

The government minimizes the excess burden of taxation by solving a Ramsey problem. The Ramsey solution is an allocation and a set of prices and policy variables such that welfare is maximized and the allocation can be decentralized as a competitive equilibrium.

An implementability condition, obtained using the first order conditions of the private problem to express taxes and prices in terms of quantities, together with the resources constraints, guarantees that the optimal allocation can be decentralized. Let  $d_t = \frac{1}{(1+r_1)\dots(1+r_t)}$ , with  $d_0 = 1$ . Since at the optimum,  $\lim_{t \rightarrow \infty} (d_t m_t + d_t B_t / p_t) = 0$ , the set of budget restrictions can be written as

$$\sum_{t=0}^{\infty} d_t(1 + \tau_{ct})c_t + \sum_{t=0}^{\infty} d_t i_t m_t = \sum_{t=0}^{\infty} d_t(1 - h_t - s_t). \quad (7)$$

Using the homogeneity assumption to obtain  $l(t) - l_m(t)m_t = l_e(t)e_t - (k-1)l(t)$ , and using the first order conditions (4), (5) and (6), to replace taxes and prices in the intertemporal budget constraint (7), we can write the implementability condition as

$$\sum_{t=0}^{\infty} \beta^t \{U_c(t)c_t - U_h(t)[1 - h_t + (1 - k)l(c_t(1 + \tau_{ct}), m_t)]\} = 0. \quad (8)$$

The resources constraints are

$$c_t + g_t \leq 1 - h_t - l(c_t(1 + \tau_{ct}), m_t), \quad (9)$$

where  $\tau_{ct}$  has not been fully substituted in the implementability and resource constraints, for analytical convenience. (4) is a constraint of the Ramsey problem since it implicitly defines the consumption tax as a function of the quantities,  $\tau_{ct} = \tau_c(c_t, h_t, m_t)$ .

The Ramsey problem is the choice of  $\{c_t, h_t, m_t\}_{t=0}^{\infty}$  that maximizes (1), subject to (8), and (9), and where taxes are defined implicitly by (4).

Let  $\psi$  and  $\beta^t \lambda_t$ ,  $t \geq 0$ , be the multipliers associated with conditions (8) and (9) respectively. Since they measure, respectively, the excess burden of taxation and the shadow price of resources in this economy, they must be positive at the optimum.

From the first order condition of the private problem (5), we know that  $i_t = 0$  decentralizes an allocation where real money balances are at full liquidity. Hence, we can analyze the optimality of the Friedman rule by verifying whether  $l_m = 0$  satisfies the first order conditions of the Ramsey problem.

The marginal condition of the Ramsey problem, for real balances, is given by

$$-[\psi U_h(t)(k-1) + \lambda_t] \left[ l_m(t) + l_e(t) c_t \frac{\partial \tau_{ct}}{\partial m_t} \right] = 0, \quad (10)$$

where, from condition (4),

$$\frac{\partial \tau_{ct}}{\partial m_t} = -\frac{(1 + \tau_{ct})l_{em}(t)}{1 + l_e(t) + (1 + \tau_{ct})c_t l_{ee}(t)} = -\frac{U_c l_{em}(t)}{c U_c l_{ee}(t) + U_h (1 + l_e(t))^2}. \quad (11)$$

This optimum marginal condition justifies the following proposition.



**Proposition 1** *Let the government finance expenditures with a consumption tax and an inflation tax, and let the transactions technology take the form in (2). Then, when either  $l_{em}(t)$  or  $l_e(t)$  is zero at the point of full liquidity,  $i_t = 0$ , for all  $t$ , is the optimal solution.*

Proposition 1 extends the results in the literature. Guidotti and Végh (1993) and Mulligan and Sala-i-Martin (1997) argue that, when transactions time is a function of consumption gross of taxes, the Friedman rule only holds if the additional assumption is made, that  $l_e(t)$  is zero at *full liquidity*. For homogeneous functions of degree different from zero, this amounts to assuming that time spent on transactions at *full liquidity* is zero. Instead, Proposition 1 states that even when time spent on transactions is positive at *full liquidity*, the Friedman rule may still be optimal, provided that  $l_{em}(t) = 0$ .

If taxes were not paid with money, then the transactions technology would be

$$s_t = l\left(c_t, \frac{M_t}{P_t}\right). \quad (12)$$

The term  $l_e(t) c_t \frac{\partial \tau_{ct}}{\partial m_t}$  in the first order condition of the Ramsey problem (10) would disappear, and the relevant marginal condition would be

$$[\psi U_h(t)(1 - k) - \lambda_t] l_m(t) = 0. \quad (13)$$

The solution can be decentralized with  $i_t = 0$  and with either a consumption tax, or alternatively an income tax. The equivalent income tax is  $\tau_t = \frac{1}{1 + \tau_{ct}} - 1$ .

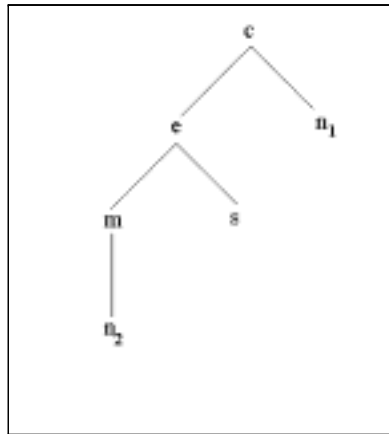
If money was not needed to pay taxes, the Friedman rule would be optimal for homogeneous functions of any degree, irrespective of which is the alternative tax instrument. However, while it is reasonable to assume that income taxes do not have to be paid with money, that is not the case for consumption taxes. Once we rightly assume that consumption taxes must be paid with money, the equivalence of the two alternative tax instruments is lost, and the Friedman rule is optimal under the additional conditions derived above.

In the next section, and through the end of the paper, we provide the intuition for these results.

### 3 Understanding the Conditions for the Optimality of the Friedman Rule

In the last section we showed that, when the government chooses the optimal combination of the inflation tax and the consumption tax, conditions on the derivatives,  $l_e$  and  $l_{em}$ , of an homogeneous transactions technology,  $s = l(e, m)$ , must be satisfied in order for the Friedman rule to be optimal. In order to provide a clear intuition for these results, it is useful to consider an equivalent economy to the monetary economy where real balances can be a costly good, and where all the goods are produced by firms, so that the implicit profits in the production of transactions are made explicit. This is the strategy followed in Correia and Teles (1996).

Since the full dynamic problem is stationary<sup>3</sup>, we will consider that the equivalent real economy is static. The structure of this economy is depicted in Figure 1.



Production structure in an equivalent real economy.

Consumption is produced using time,  $n_1$ , and transactions,  $\frac{e}{1+\tau_c}$ , according to a Leontief technology,  $c = \min\left(\frac{e}{1+\tau_c}, n_1\right)$ . The coefficient of the first argument in the Leontief technology is  $\frac{1}{1+\tau_c}$  because transactions time is a

---

<sup>3</sup>The stationarity of the Ramsey solution is not a restriction but rather a result of optimality. It also does not depend on the assumption that initial total nominal assets are zero. If these were strictly positive, it would still be optimal to set the real value of those assets to zero, and the solution would be the same.

function of transactions gross of the consumption tax, so that  $s = l(e, m)$ , where  $e = (1 + \tau_c)c$  are expenditures. The production of money requires labor,  $n_2$ , according to a linear technology with a fixed coefficient,  $m = \alpha n_2$ . The tax structure in this economy is restricted by the fact that transactions time and labor cannot be taxed. Therefore, revenues can only be raised through a tax on consumption,  $\tau_c$ , and an ad-valorem tax on money,  $\tau_m$ .

Firms maximize profits from the production of consumption, transactions, and money. Profits are distributed to the households, which are also the owners of the firms. The households maximize utility, subject to the budget constraint

$$c(1 + \tau_c) \leq w_n(1 - h) + \Pi(1 + \tau_c). \quad (14)$$

Here  $w_n$  is the common implicit wage in all the activities, and  $\Pi$  are profits in units of the consumption good, obtained from the production of transactions, which is the only activity that may not be constant returns to scale. The optimality conditions<sup>4</sup> require that

$$-l_m = \frac{1 + \tau_m}{\alpha}$$

A transformed Friedman rule would correspond in this context to  $\tau_m = 0$ , and  $m$  would be priced at the cost of production  $\frac{1}{\alpha}$ . Using the first order conditions from the profit maximization of firms, we can derive the following expression for  $\Pi$ , as a function of  $c$ ,  $h$ , and  $m$ :

$$\Pi = \frac{U_h}{U_c}(k - 1)l. \quad (15)$$

The benevolent government maximizes the households' utility, subject to the implementability condition and the resources constraint. The constraints can be expressed respectively as

$$U_c c - U_h(1 - h) - U_c \Pi = 0, \quad (16)$$

and

$$c + g - 1 + l((1 + \tau_c)c, m) + h + \frac{m}{\alpha} = 0. \quad (17)$$

---

<sup>4</sup>See Appendix A for a derivation of the first-order conditions, the implementability condition, and the profit function.

The marginal condition in the government problem for real balances is

$$\psi U_c \Pi_m + \lambda \left[ l_m + l_e c \frac{\partial \tau_c}{\partial m} + \frac{1}{\alpha} \right] = 0. \quad (18)$$

where

$$\Pi_m = \frac{U_h}{U_c} (k - 1) \left[ l_m + l_e c \frac{\partial \tau_c}{\partial m} \right],$$

and

$$\frac{\partial \tau_c}{\partial m} = \frac{-(1 + \tau_c) l_{em}}{1 + l_e + (1 + \tau_c) c l_{ee}}.$$

The first term in the first order condition (18) is the marginal impact of a change in real balances on profits, while the second term is the total marginal impact of the change in real balances on resources. This includes the direct effect  $l_m$ , the indirect effect through expenditures  $l_e c \frac{\partial \tau_c}{\partial m}$ , and the effect on resources from the costly production of the good  $\frac{1}{\alpha}$ . The effect through expenditures  $l_e c \frac{\partial \tau_c}{\partial m}$  results from the impact that a change in real balances can have on both the inflation and the consumption tax. The change in the consumption tax will in turn affect expenditures which can have a marginal impact on time used for transactions.

When the marginal production cost of real balances is zero, so that  $\frac{1}{\alpha} = 0$ , then at *full liquidity*, where  $l_m = 0$ , the marginal impact both on profits and on resources is also zero as long as  $l_e = 0$ , or  $\frac{\partial \tau_c}{\partial m} = 0$ .  $\frac{\partial \tau_c}{\partial m} = 0$ , when  $l_{em} = 0$ .

Notice that the first order condition (18) can also be written as

$$- [\psi U_h (k - 1) + \lambda] \left[ l_m + l_e c \frac{\partial \tau_c}{\partial m} \right] + \lambda \frac{1}{\alpha} = 0,$$

which is the same expression as (10), when  $\frac{1}{\alpha} = 0$ . So the conditions obtained in the last section, for the Friedman rule to be optimal are conditions on the marginal impact on profits and resources.

One case in which clearly there can be no effects on time used for transactions, is when  $k \neq 0$  and  $l_e$  is zero, at full liquidity. In this case, from Euler's theorem  $sk = l_e c (1 + \tau_c) + l_m m$ , it must be true that, at *full liquidity*, the time spent on transactions is also zero. A zero inflation tax is optimal in this case. Consider now that  $k = 0$ , as it is the case when the technology takes the form suggested by the Baumol-Tobin model of the demand for money.

It follows that  $l_e$  must be zero at *full liquidity*, while transactions time may well be positive at that point. Then, even though transactions are costly at the point of *full liquidity*, a marginal change in real balances does not affect time used for transactions. Hence, the Friedman rule is also optimal.

Suppose now that real balances were costly to produce, i.e.  $\frac{1}{\alpha} \neq 0$ . Then a transformed Friedman rule would correspond to a zero tax on the intermediate good  $m$ , so that the price of real balances would be the cost of production,  $l_m = -\frac{1}{\alpha}$ . In this case, the conditions  $l_e = 0$  or  $l_{em} = 0$  would not be satisfied at the point  $l_m = -\frac{1}{\alpha}$ , so that in general there would be an impact on profits and on real resources, and therefore the second best would not be characterized by a zero tax on  $m$ . If the transactions technology was homogeneous of degree 1, the profits would be zero, but the term of the marginal effect of real balances on resources would still not cancel out, and so, also in this case, it would be optimal to tax the intermediate good. Because of the particular production and tax structure in this economy, Diamond and Mirrlees (1971) taxation rules of intermediate rules cannot be applied.

When the cost of producing the intermediate good approaches zero, it may already be reasonable to make the assumptions that zero out the impact of changes in real balances on profits or on the resources. In fact the assumption that  $l_e = 0$ , has been made in the literature (V. Kimbrough, 1986), since it means that time used for transactions is zero at full liquidity (for degree of homogeneity different from zero). Also, when full liquidity is attained at an arbitrarily large value of real balances, then  $l_{em} = 0$ , when  $l_m = 0$ . In fact, at full liquidity,  $l_{em} = 0$  implies that  $l_{mm} = 0$ , and this must be true if  $l_m \leq 0$ .  $l_m \leq 0$  is one of our assumptions on the transactions technology.

In any case, it is still possible that  $l_e(t)$  and  $l_{em}(t)$  are not zero at *full liquidity*. Since, when that is so, the two taxes on consumption and on income are not equivalent, it is necessary to compute the optimal inflation tax when both taxes can be used together with the inflation tax. We do this in the next section.

## 4 Taxes on Money, Consumption and Income

In this section we allow for both the consumption and the income taxes to be used together with the inflation tax. We show that the conditions under which the Friedman rule is optimal are the general conditions of homogeneity

of the transactions function as obtained in Correia and Teles (1996, 1997), using only income taxes as the alternative to the inflation tax. We also show that unless  $l_e = 0$ , at full liquidity, the optimal policy combination requires to fully tax income and subsidize consumption. In that world where the government would be withdrawing all the income to give a part back to the agents as consumption, gross expenditures would be zero and resources used for transactions would be minimized. We proceed to show these results.

The households now maximize (1), subject to (2) and to the budget constraint:

$$P_t(1+\tau_{ct})c_t + M_{t+1} + B_{t+1} \leq P_t(1-\tau_{lt}) [1 - h_t - l((1 + \tau_{ct})c_t, m_t)] + M_t + (1+i_t)B_t \quad (19)$$

$$M_0 + B_0 \leq 0$$

At the optimum the following marginal conditions must be satisfied:

$$\frac{U_c(t)}{U_h(t)} = (1 + \tau_{ct}) \left[ \frac{1}{(1 - \tau_{lt})} + l_e(t) \right], \quad (20)$$

$$-l_m(t) = \frac{1}{1 - \tau_{lt}} i_t, \quad (21)$$

$$\frac{U_h(t)}{\beta U_h(t+1)} = (1 + r_{t+1}) \left( \frac{1 - \tau_{lt}}{1 - \tau_{lt+1}} \right). \quad (22)$$

Notice that there is an indeterminacy of the prices from the private problem. In fact, (20) and (21) are two equations in three unknowns:  $\tau_{ct}$ ,  $\tau_{lt}$ , and  $i_t$ . Therefore, it is only possible to express two prices as functions of the quantities and of the remaining price. We choose to express  $\tau_{lt}$ , and  $i_t$  as a function of  $\tau_{ct}$ .

In this setup the Ramsey problem is defined by the maximization, with respect to the quantities and to the level of the consumption tax, of the utility function, subject to the implementability condition and the resources constraint.

Substituting equations (20)-(22) into the intertemporal budget constraint, rearranging, and using the fact that  $l(t)$  is homogeneous of degree  $k$ , we obtain the same expression for the implementability condition as in the case

when a consumption tax is the only alternative tax to inflation. We rewrite it here for convenience:

$$\sum_{t=0}^{\infty} \beta^t \{U_c(t)c_t - U_h(t) [(1 - h_t) + (1 - k)l(c_t(1 + \tau_{ct}), m_t)]\} = 0. \quad (23)$$

The resources constraints are also identical, and are given by

$$c_t + g_t \leq 1 - h_t - l(c_t(1 + \tau_{ct}), m_t). \quad (24)$$

Notice, however, that now  $\tau_{ct}$  is a choice variable for the government. Let  $\psi$  and  $\beta^t \lambda_t$  be the multipliers of the implementability conditions and the resources conditions, respectively. We will be using the first order conditions of the Ramsey problem, for leisure, real balances, and the consumption tax, that are given respectively by

$$U_h + \psi \{U_{ch}c_t + U_h - U_{hh} [1 - h_t + (k - 1)l(t)]\} = \lambda_t \quad (25)$$

$$- [\psi U_h (k - 1) + \lambda_t] l_m(t) = 0. \quad (26)$$

$$- [\psi U_h (k - 1) + \lambda_t] c_t l_e(t) = 0. \quad (27)$$

**Proposition 2** *Let the government finance expenditures through an income tax, a consumption tax, or an inflation tax, and let the transactions technology take the form in (2). Then, an interior solution of the Ramsey problem requires that  $l_m(t) = l_e(t) = 0$ . Also,*

- i) if  $l_e(t) = 0$  at full liquidity, for any  $e_t$ , then the optimal solution is decentralized by  $i_t = 0$ , and by a continuum of combinations of  $\tau_{ct}$  and  $\tau_{lt}$ . This is the case for the Baumol-Tobin transactions technology;*
- ii) if  $l_e(t) > 0$  at full liquidity, for any  $e_t$ , then the optimal solution is decentralized by  $i_t = 0$ ,  $\tau_{ct} = -1$ , and  $\tau_{lt} = 1$ . In this case expenditures are zero at the optimum.*

**Proof.** From (26) and (27), if the term in square brackets is not zero, then it is clear that an interior solution must satisfy  $l_m(t) = l_e(t) = 0$ . In Appendix B, we show that, indeed, the term in square brackets is strictly positive. We now proceed to show *i)* and *ii)*.

- i) Let  $l_e = 0$  at *full liquidity*, for any  $e_t = c_t(1 + \tau_{ct})$ . Then, equation (26) is satisfied by  $l_m(t) = 0$ , and equation (27) is satisfied for any value of  $\tau_{ct}$ . From the first order condition of the private problem (21), the optimal solution is decentralized by  $i_t = 0$ , and by any combination of  $\tau_{ct}$  and  $\tau_{lt}$  which satisfies the private first order condition (20). For transactions technologies with  $k = 0$ , in particular for the Baumol-Tobin,  $l_e = 0$  at *full liquidity*, for any  $e$ . This is clear since,  $sk = l_e e + l_m m$ , from Euler's theorem.
- ii) Let  $l_e > 0$  at *full liquidity*, for any  $e_t$ . As shown in Appendix B, the term in square brackets in (27) is not zero and therefore, there is no interior solution.

We now check for a corner solution. The effect of a marginal change in the consumption tax on welfare is given by the derivative of the Lagrangian of the Ramsey problem with respect to  $\tau_{ct}$ , which can be rewritten as

$$\frac{\partial L}{\partial \tau_{ct}} = - [\psi U_h(k-1) + \lambda_t] c_t l_e(t). \quad (28)$$

Since  $l_e(t) \geq 0$ , and since the term in square brackets is positive for any degree of homogeneity of the transactions technology,  $\frac{\partial L}{\partial \tau_{ct}}$  is always negative. Therefore, it is always optimal to decrease the consumption tax in order to increase welfare. Hence, we conclude that welfare is maximized at  $\tau_{ct} = -1$ . Beyond this point expenditures would become negative and the problem would not be well specified.

From the first order condition of the private problem (20), rewritten as

$$\tau_{lt} = 1 - \frac{U_h(t)(1 + \tau_{ct})}{U_c(t) - U_h(t)(1 + \tau_{ct})l_e(t)},$$

it follows that the optimal value of the income tax is  $\tau_{lt} = 1$ .

Since, from (26) at the optimum  $l_m(t) = 0$ , the Friedman rule,  $i_t = 0$ , is always optimal ■

The proposition states that, when we allow for the use of both consumption and income taxes as alternative tax instruments to inflation, the Friedman rule is always optimal. The intuition is that, for any fixed level of the consumption tax, it is always preferable to set the inflation tax to zero and



to raise the necessary revenues through the income tax. This is the standard problem analyzed by Correia and Teles (1996, 1997) when they compare the inflation tax with an income tax. The consumption tax is just a parameter in that problem. So it is clear that the Friedman rule is always optimal. The particularly striking lesson from Proposition 2 is that when  $l_e$  is never zero at the point of *full liquidity*, or when  $l_e$  is zero only when  $e = 0$ , then the optimum requires that consumption is completely subsidized and labor is completely taxed. In that solution, households are left with no income, and consumption is free. Since consumption is free, there are no transactions and therefore, resources spent on transactions are minimized. In particular, when  $k > 0$ , transactions time is zero at the optimum, no matter what the level of consumption is.

This objective to minimize resources used for transactions also explains the *restrictive* conditions for the optimality of the Friedman rule, when the choice is between the inflation tax and the consumption tax. Also there, there is an incentive to reduce the tax on consumption in order to save on resources used for transactions. When the income tax is available as well, it is then possible to take this argument to the limit, and as a result we obtain the extreme policy prescription that consumption should be fully subsidized.

At the heart of this problem, there is an incorrect specification of the transactions technology. Notice that in this economy time used for transactions is minimized when the government has to withdraw all the resources in circulation and then give them back as consumption to the agents. We now turn to the discussion of an alternative, more appropriate, specification of the transactions technology.

## 5 Improving on the Specification of the Transactions Technology

The transactions technology has so far been represented by

$$s_t = l \left( c_t(1 + \tau_{ct}), \frac{M_t}{P_t} \right)$$

This specification has an undesirable property that explains the unreasonable policy prescriptions obtained before. In fact, notice that, from homogeneity,

the function can be written as

$$s_t = (1 + \tau_{ct})^k l \left( c_t, \frac{M_t}{(1 + \tau_{ct})P_t} \right) \quad (\text{tct3})$$

where  $k$  is the degree of homogeneity. Notice that, for  $k \neq 0$ , it is possible to change time used for transactions, without changing the real quantity of transactions measured in units of the consumption good,  $c_t$ , and without changing the real quantity of money required to buy those goods,  $\frac{M_t}{(1 + \tau_{ct})P_t}$ . This can be achieved by appropriately changing the tax on consumption and  $\frac{M_t}{P_t}$ . An extreme example of this is that, by changing  $\frac{M_t}{P_t}$  to restore *full liquidity*, it is possible to pick  $\tau_{ct} = -1$ . As a result  $e_t = 0$ , and then time used for transactions is also zero,  $s_t = 0$ . This is clearly a weakness of the specification.

We propose now a specification of the transactions technology that does not share this property for any degree of homogeneity. Indeed, it is a good property of the model that when the real quantity of the goods does not change and when money needed to buy these goods does not change as well, then time used for transactions should not change either. The new transactions technology is described by

$$s_t = \tilde{l} \left( c_t, \frac{M_t}{(1 + \tau_{ct})P_t} \right). \quad (29)$$

The two specifications are equivalent when  $k = 0$ . In this case, at full liquidity,  $l_e(t) = \tilde{l}_c(t) = 0$ , and therefore it is not possible to affect the time used for transactions.

Under the new specification of the transactions technology, the Friedman rule is optimal when the government chooses between taxing consumption or money. We now proceed to show this. We go back to the model economy of Section 3, where the only alternative to inflation is a tax on consumption and where taxes are paid with money. The difference is that we adopt the new specification for the transactions cost technology, as described by (29).

The households then maximize (1), subject to (29) and to the budget constraint (3), where  $W_0 = 0$ . The first order conditions are:

$$\frac{U_c(t)}{U_h(t)} = 1 + \tau_{ct} + \tilde{l}_c(t), \quad (30)$$

$$-\tilde{l}_m(t) = i_t(1 + \tau_{ct}), \quad (31)$$

$$\frac{U_h(t)}{\beta U_h(t+1)} = 1 + r_{t+1}, \quad (32)$$

Now, define  $\tilde{m}_t = \frac{M_t}{P_t(1+\tau_{ct})}$ . Substituting equations (30)-(32) into the intertemporal budget constraint (7), and using the fact that  $\tilde{l}(t)$  is homogeneous of degree  $k$ , we obtain the implementability condition

$$\sum_{t=0}^{\infty} \beta^t \left\{ U_c(t)c_t - U_h(t) \left[ (1 - h_t) + (1 - k)\tilde{l}(c_t, \tilde{m}_t) \right] \right\} = 0. \quad (33)$$

The resources constraints are given by

$$c_t + g_t \leq 1 - h_t - \tilde{l}(c_t, \tilde{m}_t). \quad (34)$$

The marginal condition of the Ramsey problem for  $\tilde{m}_t$  is

$$-[\psi(k-1) + \lambda_t]\tilde{l}_{\tilde{m}}(t) = 0. \quad (35)$$

where  $\psi$  and  $\beta^t \lambda_t$  are the multipliers of the implementability conditions and the resources conditions, respectively.

**Proposition 3** *The Ramsey allocation, when the government can finance expenditures through a consumption tax and an inflation tax, and the transactions technology takes the form in (29), is decentralized by  $i_t = 0$ , for all  $t$ .*

**Proof.** At full liquidity,  $\tilde{l}_{\tilde{m}}(t) = 0$ . From equation (31) it appears that the optimal solution could be decentralized by either  $i_t = 0$  and  $\tau_{ct} > 0$ , or by  $\tau_{ct} = -1$  and  $i_t > 0$ . However, setting  $\tau_{ct} = -1$ , would imply that the relative price of consumption would be zero, which would correspond to satiation in consumption. In fact, we can rewrite the first order condition of the private problem, (30), as

$$\frac{U_c(t) - \tilde{l}_c(t) U_h(t)}{U_h(t)} = 1 + \tau_{ct}.$$

When  $\tau_{ct} = -1$ , it must be that the left hand side of the equation above is set to zero. This is the case when  $U_c(t) - \tilde{l}_c(t) U_h(t) = 0$ , i.e. the marginal benefit of consumption, net of the cost in terms of leisure of higher transactions time,

is zero. The price of consumption becomes so low that households are satiated in consumption<sup>5</sup>. This is not feasible ■

Under the new specification, it is no longer possible, at *full liquidity*, to reduce time used for transactions by increasing the subsidy on consumption. In fact, in this case  $ks_t = \tilde{l}_c(t)c_t + \tilde{l}_m(t)\tilde{m}_t$ , where  $\tilde{m}_t = \frac{M_t}{P_t(1+\tau_{ct})}$ . Thus, decreasing the tax on consumption, at full liquidity, does not affect the time used for transactions.

The results in this section allow us to conclude that the Friedman rule is the optimal policy for all homogeneous transactions technologies. It is therefore a general result. The only remaining argument for *fragility* of the optimality of the Friedman rule is lack of homogeneity. In fact if the transactions technology is not homogeneous the Friedman rule may not be optimal<sup>6</sup>.

Theoretical work on transactions technologies supports the assumption of homogeneity. This is the case in Baumol (1952) and Tobin (1956), the generalization of Barro (1976), Guidotti (1989) or Jovanovic (1982). In all these models the transactions technologies are homogeneous of degree zero. Marshall (1992) proposes and estimates a transactions technology that is homogeneous of degree one; and Braun (1994) estimates the degree of homogeneity to be .98.

Homogeneous functions have the property that the scale elasticity of the money demand is one, at full liquidity. Indeed, for an homogeneous function  $s = l(e, m)$ , we can write  $l = L(\frac{m}{e})e^k$ . At full liquidity, we have  $L'(\frac{m}{e}) = 0$ , so that  $m = \alpha e$ , where  $\alpha$  is a constant. While there is empirical evidence, often contradictory, on the scale elasticity, there is virtually no evidence on the elasticity at full liquidity. The closest evidence are the results in Mulligan and Sala-i-Martin (1996) that suggest that the scale elasticity of the household demand for money approaches one for interest rates that are below the current ones. In any case, in the computation of the optimal inflation tax, Mulligan and Sala-i-Martin (1997) make the assumption of homogeneity, and concentrate on the impact of the restriction that taxes are paid with money on the optimal inflation tax. Not surprisingly they conclude that the optimal inflation tax is very close to the Friedman rule.

---

<sup>5</sup>If there were taxes on income, those taxes could undo this effect, and it could still be optimal to set  $\tau_{ct} = -1$ , as is the case in Section 3.

<sup>6</sup>Correia and Teles (1998), using a calibrated model with US data on the money demand, show that if the transactions technology is not homogeneous, still the optimal inflation tax is very close to the Friedman rule.

## 6 Is the Optimality of the Friedman Rule a Robust Result? Concluding Remarks.

In this paper, we compute the optimal inflation tax in a simple monetary model where money and time are used for transactions. We show that the Friedman rule is the optimal solution whether the fiscal choice is between an income tax and an inflation tax, a consumption tax and an inflation tax or the three tax instruments, and whether taxes are paid with or not. In doing so, we object to the claim in recent literature (V. Mulligan and Sala-i-Martin, 1997) that, when taxes must be paid with money, the conditions under which the Friedman rule is optimal are *restrictive* and that therefore the optimality of the Friedman rule is a *fragile* result. We argue that a misspecification of the transactions technology is the reason for the apparent fragility of the optimality of the Friedman rule, when consumption taxes are used.

When the government faces the problem of choosing the combination of the income and the inflation tax, the optimal solution is to fully use the income tax and abstain from taxing money, for homogeneous transactions functions of any degree. This is shown in Correia and Teles (1996, 1997). The reason for the general result of optimality of the Friedman rule is the free good characteristic of money. Indeed, if money was costly to produce, it would only be optimal not to tax money for constant returns to scale transactions technologies.

The literature on the optimal fiscal choice between the inflation tax and a consumption tax has identified additional restrictions for the Friedman rule to be optimal, when taxes are paid with money. We show that these restrictions are no longer necessary when income taxes are also available, together with consumption taxes. We also show that the reason for the additional restrictions is a misspecification of the transactions technology. Under a more reasonable specification where this property is not present, the result shown by Correia and Teles (1996, 1997) that the optimality of the Friedman rule is a robust result is fully recovered.

The transactions technology, as is specified in that literature, 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. In particular if consumption is fully subsidized, then time used for transactions can be set to zero. Since income taxes are

available, it is feasible, and may be optimal, to fully subsidize consumption by fully taxing income. The same mechanism explains the restrictive conditions identified in the literature for the Friedman rule to be optimal, when the government chooses the optimal combination of the inflation tax and a consumption tax.

The optimality of the Friedman rule is indeed a robust result because the generality of the required conditions of homogeneity can not be found in any other sets of rules in the Public Finance literature. It is the free good characteristic of money that explains this robustness. The relevant question amounts to asking: What is the optimal price charged on a good that has a very low cost of production? Even if the optimal ad-valorem tax is very high, still the price should in general be quite low. This is the reason for the robustness of the Friedman rule as an optimal taxation rule.

## References

- [1] Barro, Robert J., Integral Constraints and Aggregation in an Inventory Model of Money Demand, *Journal of Finance* 31 (1976), 77-87.
- [2] Baumol, William J., "The Transactions Demand for Cash: An Inventory Theoretic Approach", *Quarterly Journal of Economics* 66 (1952), 545-556.
- [3] Braun, Anton, Another Attempt to Quantify the Benefits of Reducing Inflation, *Federal Reserve Bank of Minneapolis Quarterly Review* Fall (1994), 17-25.
- [4] Chari, V. V., Lawrence J. Christiano and Patrick J. Kehoe, "Optimality of the Friedman Rule in Economies with Distorting Taxes", *Journal of Monetary Economics* 37 (1996), 203-223.
- [5] Correia, Isabel and Pedro Teles, "Is the Friedman Rule Optimal when Money is an Intermediate Good?", *Journal of Monetary Economics* 38 (1996), 223-244.
- [6] Correia, Isabel and Pedro Teles, "The Optimal Inflation Tax", Institute of Empirical Macroeconomics of the Federal Reserve Bank of Minneapolis Discussion Paper 123 (1997), forthcoming *Review of Economic Dynamics*.
- [7] Diamond, Peter A. and James A. Mirrlees, "Optimal Taxation and Public Production", *American Economic Review* 63 (1971), 8-27, 261-268.
- [8] Friedman, Milton, "The Optimum Quantity of Money" In *The Optimum Quantity of Money and other Essays* edited by M. Friedman, pp. 1-50, Aldine, Chicago, Il., 1969.
- [9] Guidotti, Pablo E., Exchange Rate Determination, Interest rates and an Integrative Approach to the Demand for Money, *Journal of International Money and Finance* 8 (1989), 29-45.
- [10] Guidotti, Pablo E. and Carlos A. Végh, "The Optimal Inflation Tax when Money Reduces Transactions Costs", *Journal of Monetary Economics* 31 (1993), 189-205.

- [11] Jovanovic, Boyan, Inflation and Welfare in the Steady State, *Journal of Political Economy* 90 (1982), 561-577.
- [12] Kimbrough, Kent P., "The Optimum Quantity of Money Rule in the Theory of Public Finance", *Journal of Monetary Economics* 18 (1986), 277-284.
- [13] Lucas, Robert E., "The Welfare Cost of Inflation", mimeo, The University of Chicago, 1994.
- [14] Marshall, David A., Inflation and Asset Returns in a Monetary Economy, *Journal of Finance* 47 (1992), 1315-1342.
- [15] Mulligan, Casey B. and Xavier X. Sala-i-Martin, "Adoption of Financial Technologies: Implications for Money Demand and Monetary Policy", NBER Working Paper Series #5504, March 1996.
- [16] Mulligan, Casey B. and Xavier X. Sala-i-Martin, "The Optimum Quantity of Money: Theory and Evidence", *Journal of Money, Credit, and Banking* (1997), Vol.29, No. 4, 687-715.
- [17] Ramsey, Frank P., "A Contribution to the Theory of Taxation", *Economic Journal* 37 (1927), 47-61.
- [18] Tobin, James, "The Interest Elasticity of the Transactions Demand for Cash", *Review of Economics and Statistics* 38 (1956), 241-247.



## A Restrictive conditions for the optimality of the Friedman rule

In the equivalent real economy to the monetary economy, illustrated in Figure 1, consumption is produced using labor  $n_1$ , and transactions,  $\frac{e}{1+\tau_c}$ , according to a Leontief technology,  $c = \min\left(\frac{e}{1+\tau_c}, n_1\right)$ . The production of transactions requires time,  $s$ , and money,  $m$ . Money also requires labor,  $n_2$ , according to a linear technology with a fixed coefficient,  $m = \alpha n_2$ . Since transactions time and labor cannot be taxed, revenues can only be raised through a tax on consumption,  $\tau_c$ , and an ad-valorem tax on money,  $\tau_m$ . The wage for labor in all the activities must be the same and is denoted by  $w_n$ , while  $p_j$  is the price of good  $j$  in units of the consumption good.  $\Pi$  are profits, expressed in units of the consumption good, obtained from the production of transactions, that is the only production function that is not constant returns to scale.

*Households' problem*

$$\text{Max} U(c, h),$$

$$\text{s.t. } c(1 + \tau_c) = w_n(1 - h) + \Pi(1 + \tau_c).$$

*The problems of the firms*

$$i) \text{ Max } \Pi^c = c - \frac{w_n n_1}{1 + \tau_c} - p_e e,$$

$$\text{s.t. } c = \min\left(\frac{e}{1 + \tau_c}, n_1\right).$$

$$ii) \text{ Max } \Pi^e = e - \frac{w_n s}{(1 + \tau_c) p_e} - \frac{p_m (1 + \tau_m) m}{p_e},$$

$$\text{s.t. } s = l(e, m).$$

$$iii) \text{ Max } \Pi^m = m - \frac{w_n n_2}{(1 + \tau_c) p_m},$$

$$\text{s.t. } m = \alpha n_2.$$

First order conditions:

$$\frac{U_c}{U_h} = \frac{1 + \tau_c}{w_n} \tag{36}$$

from the household' problem;

$$p_e + \frac{w_n}{(1 + \tau_c)^2} = \frac{1}{(1 + \tau_c)} \quad (37)$$

$$c = n_1 = \frac{e}{(1 + \tau_c)} \quad (38)$$

from firm (i) problem;

$$l_e = \frac{(1 + \tau_c) p_e}{w_n} \quad (39)$$

$$-l_m = \frac{(1 + \tau_m) p_m}{\frac{w_n}{(1 + \tau_c)}} \quad (40)$$

from firm (ii) problem;

$$\frac{p_m (1 + \tau_c)}{w_n} = \frac{1}{\alpha}$$

from firm (iii) problem.

To obtain the implementability condition (16), plug the first-order conditions (36) into the household's budget constraint.

To obtain the expression for profits, in units of consumption,

$$\Pi = p_e \Pi^e = p_e e - \frac{w_n s}{1 + \tau_c} - p_m (1 + \tau_m) m,$$

Using (39) and (40) to replace  $p_e$  and  $p_m (1 + \tau_m)$  respectively, and using (36), to replace  $\frac{w_n}{1 + \tau_c}$ , we obtain

$$\Pi = \frac{U_h}{U_c} (k - 1) l (c (1 + \tau_c), m).$$

## B Proof of Proposition 2

Here we show that, at the optimum,

$$\psi U_h (k - 1) + \lambda_t > 0$$

When  $k \geq 1$ , the term is strictly positive, because the multipliers are strictly positive. In fact,  $\psi > 0$  at the optimum since it measures the excess burden of taxation. When  $k < 1$ , an argument similar to the one used in

Correia and Teles (1996) ensures that the term is still strictly positive. To see how, substitute  $\lambda_t$  with the expression given by (25). Since the problem is stationary, manipulating the implementability condition we obtain  $D(t) c_t = [1 - h_t + (k - 1) l(t)]$ , where  $D(t) \equiv \frac{U_c(t)}{U_h(t)}$  is the marginal rate of substitution between consumption and leisure. Using this expression and rearranging terms, we obtain that

$$[\psi U_h(k - 1) + \lambda_t] = U_h + \psi [U_h(t) D_h(t) c_t + U_h(t) k] > 0.$$

The sign given to the expression follows from the fact that, if consumption is a non inferior good, an increase in leisure increases the marginal rate of substitution, so that  $D_h(t) \geq 0$ .