WORKING PAPER NO. 38

THE OPTIMAL INFLATION TAX WHEN TAXES ARE COSTLY TO COLLECT

BY FIORELLA DE FIORE

November 2000
I wish to thank Pedro Teles, Mike Artis and an anonymous referee for useful comments and suggestions. I am also grateful to Xavier Sala-i-Martin and Steffen Hoernig for helpful discussions and to Andres Manzanares for research assistance. All the remaining errors are mine.
## Contents

Abstract 5

Non-Technical Summary 7

1 Introduction 9

2 The Model 11
   2.1 The Household’s Problem 13
   2.2 The Ramsey Problem 14

3 Alternative Model Specifications 17
   3.1 Marginal Collection Costs are Constant 17
   3.2 The Alternative to Inflation is a Consumption Tax 18

4 The Optimal Inflation Tax when the Friedman Rule is Not Optimal 20

5 The Optimal Inflation Tax in the U.S. 24
   5.1 Evidence on the Relevant Elasticities 24
   5.2 Evidence on the Tax Collection System 27
   5.3 Calibration and Numerical Results 29

6 Conclusion 32

References 33

Appendices 36
   A Proposition 1 36
   B Proposition 4 37

European Central Bank Working Paper Series 40
Abstract

Tax collection costs have been advocated in the literature as a reason to deviate from the Friedman rule, in standard general equilibrium monetary models with flexible prices. This paper shows that there are conditions under which the Friedman rule is optimal despite the presence of collection costs. When these conditions are not satisfied, the optimal inflation tax depends upon the collection costs parameter and schedule, the interest and scale elasticity of money demand, and the compensated labor supply elasticity. Numerical results obtained by calibrating the model on US data suggest that collection costs do not justify substantial departures from Friedman’s prescriptions.

Keywords: Friedman rule, optimal inflation tax, collection costs.
Non-Technical Summary

There has been a long debate in the literature on what should be the optimal inflation tax, for an economy where the government has to finance public expenditures through distortionary taxation. Friedman (1969) shows that, in a first-best environment where lump-sum taxes are available to the government, the optimal monetary policy has to achieve a zero nominal interest rate (a zero inflation tax) or a deflation rate equal to the discount rate of the economy. Phelps (1973) criticizes this result by arguing that governments have only access to distortionary taxes and by claiming that, in a second-best environment, it is generally optimal to set a positive inflation tax. Later works by Kimbrough (1986), Chari, Christiano and Kehoe (1996), Correia and Teles (1996, 1998), and De Fiore and Teles (1999) have overcome this result and have shown that the Friedman rule is optimal under general conditions, even when the government can only finance expenditures through distortionary taxation.

Other arguments have been suggested in the literature to justify a positive inflation tax in a flexible prices environment. One frequently advocated reason to deviate from the Friedman rule is the presence of tax collection costs. It is argued that there is an important difference between the inflation tax and other distortionary fiscal instruments. In fact, while raising revenue from the former is costless, raising revenue from the latter involves substantial costs, due to the burden of organizing a tax system, to the possibility of evasion and to the corresponding necessity of enforcement. Aizenman (1987) and Végh (1989b) claim that, when the presence of collection costs is accounted for, the optimal inflation tax is always positive. However, they do not quantify the extent to which optimal monetary policy should deviate from Friedman’s prescription of zero nominal interest rates.

This paper reconsiders the importance of the collection costs argument in generating optimal deviations from the Friedman rule, both from a theoretical and an empirical perspective.

The first part of the paper presents a general transactions technology model, where government expenditures are financed through revenues from the inflation tax and from an alternative tax that is costly to collect. It is shown that there are conditions under which the Friedman rule is optimal despite the presence of collection costs. When these conditions are not satisfied, the optimal inflation tax is shown to depend upon the collection cost schedule, as well as the interest elasticity of money demand, the consumption elasticity of money demand, and the compensated labor supply elasticity.
In the second part of the paper, it is assessed how important collection costs are in determining the optimal inflation tax in the United States. The numerical results obtained within the calibrated model show that the optimal inflation tax is not very sensitive to different assumptions. Under the most unfavourable case, where all collection costs are variable costs, marginal collection costs are increasing, and tax collection requires throwing away 20 percent of total government revenues, the computed optimal inflation tax remains below one percentage point. These results suggest that the presence of collection costs does not justifies substantial deviations from Friedman’s prescription.
1 Introduction

There has been a long debate in the literature on what is the optimal inflation tax, for an economy where the government has to finance public expenditures by collecting revenues from distortionary taxes. Friedman (1969) shows that, in a first-best environment with lump-sum taxes, the optimal monetary policy has to achieve a zero nominal interest rate (a zero inflation tax) or a deflation rate equal to the discount rate of the economy. Phelps (1973) criticizes this result by arguing that, in a second-best environment where the government has only access to distortionary taxes, the optimal inflation tax is always positive. The author obtains this result by using the optimal taxation approach of Ramsey (1927), which allows to derive the optimal combination of ad-valorem tax rates on costly final goods. In line with Ramsey, Phelps finds that optimal ad-valorem tax rates are proportional to each good’s price elasticity and thus he concludes that it is generally optimal to tax money. Later works by Kimbrough (1986), Guidotti and Végh (1993), and Chari, Christiano and Kehoe (1996) have challenged Phelps’ conclusions by showing that, under certain conditions on what alternative tax to inflation is available and on either preferences or the transactions technology, the Friedman rule is also optimal in a second-best environment. More recently, Correia and Teles (1996, 1998) and De Fiore and Teles (1999) have shown that the Friedman rule is optimal under general conditions. An intuitive explanation for the robustness of the Friedman rule is that the inflation tax is a unit tax on a costless good. Even though the optimal ad-valorem tax on real balances is generally positive, the optimal unit tax on money is zero because the cost of producing money approaches zero.

Other arguments have been advocated to justify a positive inflation tax. Végh (1989a) shows that the presence of currency substitution generates a positive optimal inflation tax. When foreign money can also be used to reduce transactions costs, a positive foreign interest rate acts as a tax on consumption and thus it distorts the consumption-leisure choice. In this case, the government should partly offset this distortion by setting a positive domestic interest rate and by subsidizing consumption. Nicolini (1998) finds that a positive optimal inflation tax also arises in the presence of an underground economy, since inflation is the government’s only instrument able to tax this sector. Aizenman (1987) and Végh (1989b) argue that inflationary finance can also be justified by the presence of tax collection costs. While collecting the inflation tax is a costless process, there are substantial costs associated with collection of all other taxes, due to the burden of organizing a collection system, the possibility of evasion, and the corresponding necessity of enforcement. When these costs are accounted for, the optimal inflation tax
becomes positive. However, the authors do not quantify the extent to which optimal monetary policy should deviate from Friedman’s prescription of zero nominal interest rates.

This paper reconsiders the importance of collection costs in justifying deviations from the Friedman rule, both from a theoretical and an empirical perspective.

In the first part of the paper, I set up a general transactions technology model, where government expenditures are financed through revenues from the inflation tax and from an alternative tax that is costly to collect. In the limiting case where satiation is attained at an arbitrarily large level of real balances, the Friedman rule is shown to be optimal for any level of collection costs and for any assumption on the collection cost schedule. This result arises, for instance, for specifications of the transactions technology as derived by Baumol (1952) or Tobin (1956) within their model of the agents’ cash management problem. When satiation occurs at a finite level of real balances, the Friedman rule may still be optimal, under restrictive assumptions on the transactions costs technology. Finally, when the conditions under which the Friedman rule is optimal are not satisfied, the optimal inflation tax is shown to depend upon the collection cost schedule, as well as the interest elasticity of money demand, the consumption elasticity of money demand, and the compensated labor supply elasticity.

In the second part of the paper, I assess how important collection costs are in determining the optimal inflation tax in the United States. The numerical results obtained within the calibrated model show that the optimal inflation tax is not very sensitive to different assumptions. Moreover, the computed values never exceed 1 percent, suggesting that the presence of collection costs does not justify substantial deviations from the Friedman rule.

The paper proceeds as follows. In section 2, I present a shopping time model, where the alternative to inflation is a tax that is costly to collect. It is first assumed that the alternative fiscal instrument is an income tax, and that marginal collection costs are increasing with the amount of revenues collected. In section 3, I consider different specifications of the model and I show that these specifications do not affect the conditions under which the Friedman rule is optimal; the case of constant marginal collection costs is presented in section 3.1, while the case where the alternative tax instrument is a consumption tax is presented in section 3.2. In section 4, I analyze the determinants of the optimal inflation tax, when the Friedman rule is not optimal. In section 5, I present evidence on the money demand, the relevant elasticities and the tax collection system in the United States. Then, I calibrate the model under different assumptions and I compute the optimal inflation tax in each case. In Section 6, I conclude.
2 The Model

The model builds on Kimbrough (1986), Guidotti and Végh (1993), Chari, Christiano and Kehoe (1996), and Correia and Teles (1996), the main difference being the presence of costs in collecting revenues from the income tax. The setup is more general than in Végh (1989) in that no restriction is imposed either on the degree of homogeneity of the transactions technology or on the level of satiation in real money balances. Végh (1989) only considers the case when the alternative to inflation is a tax on consumption. In this paper, it is first assumed that the alternative to inflation is a tax on income. This provides a convenient benchmark because, when collection costs are zero, the model collapses to the one in Correia and Teles (1996) and the Friedman rule is always optimal.\(^1\) The case when the alternative to inflation is a tax on consumption is subsequently analyzed to show that the choice of the alternative fiscal instrument is not relevant for the results.

The economy is inhabited by a large number of identical infinitely lived households whose lifetime utility is given by

\[
\sum_{t=0}^{\infty} \beta^t u(c_t, h_t),
\]

where \(c_t\) is consumption, \(h_t\) is leisure, \(\beta\) is the discount factor, and the usual assumptions on the utility function apply. The representative agent is endowed with one unit of time, which can be used in three different activities: leisure, \(h\), labor, \(n\), and shopping, \(s\).

Shopping requires time, as described by a transactions costs technology\(^2\) that is assumed to be homogeneous of degree \(k\):

\[
s_t = l(c_t, m_t) = L \left( \frac{m_t}{c_t} \right)^k.
\]

Here \(s_t\) is time spent shopping, \(m_t = M_t/P_t\) are real money holdings, \(M_t\) are nominal money holdings, and \(P_t\) is the price level in terms of the consumption good. For convenience, I adopt the notation \(u(t) = u(c_t, h_t)\), \(u_x(t) = u_x(c_t, h_t)\), \(l(t) = l(c_t, m_t)\), and \(l_x(t) = l_x(c_t, m_t)\), where the subscript \(x\) denotes the derivative of the function with respect to the argument \(x\). The following assumptions on the transactions technology are made:

\(^1\)The timing assumed here is different. However, this distinction is not important for the analysis.

\(^2\)Correia and Teles (1999) show that a monetary model with a transaction function that is homogeneous of any degree is isomorphic to a model with money in the utility function, when the satiation point in real balances has a unitary elasticity with respect to consumption. The results obtained in this section would therefore extend to that class of models.
i. \( L(t) : A \to R^+ \), \( A \subseteq R^+ \); \( L'(t) \leq 0 \); \( L''(t) \geq 0 \). This implies that \( l_c(t) \geq 0 \), \( l_m(t) \leq 0 \), \( l_{mm}(t) \geq 0 \). All these partial derivatives are continuous.

ii. \( k \geq 0 \).

iii. \( l(t) \) is such that the consumer’s problem is concave.

I define the satiation point in real balances \( \frac{m}{c} \) as the point where it is not possible to reduce transactions time by increasing money holdings, so that \( L' \left( \frac{m}{c} \right) = 0 \).

Assumptions i) requires transactions time to be non-negative, increasing in the amount of consumption and decreasing in the amount of money balances. Also, increases in real money balances decrease shopping time at a decreasing rate. Assumption ii) is made because otherwise \( l_c(t) \) would be negative at satiation, for a subset of the class of transactions technologies. In fact, from the homogeneity assumption we know that \( l_c = \frac{c}{k} - \frac{c}{k} L'(t) \).

At satiation, \( mL'(t) = 0 \)\textsuperscript{3} When \( L(t) \) is strictly positive at satiation, it must be that \( k \geq 0 \) for \( l_c \) not to become negative.

The production technology transforms one unit of labor into one unit of the good. At each period \( t \), the consumer’s time constraint is given by

\[
h_t + n_t + s_t = 1.
\]

I assume that initial nominal wealth is zero, or that \( M_{t-1} + (1 + i_{t-1}) B_{t-1} = 0 \).\textsuperscript{4} The sequence of budget constraints is then given by

\[
M_0 + B_0 \leq 0,
\]

\[
P_0 c_t + M_{t+1} + B_{t+1} \leq M_t + (1 + i_t) B_t + P_t \left( 1 - \tau_t \right) (1 - h_t - s_t), \ t \geq 0,
\]

where \( M_t \) are holdings of money chosen at the beginning of \( t \), to be used to finance transactions in that period, and \( B_t \) is the amount of one-period

\textsuperscript{3}When \( m \) is finite, \( mL' \left( \frac{m}{c} \right) = 0 \) because \( l_m(c, m) \equiv L' \left( \frac{m}{c} \right) c^{k-1} = 0 \). When \( m \to \infty \), \( \lim_{m \to \infty} mL' \left( \frac{m}{c} \right) = 0 \) because it is reasonable to assume that seignorage, defined as \( S = \frac{m}{c} \), is zero when \( i \), the unit tax on money, is also zero. In fact, in equilibrium, \( i = \frac{1 - \tau}{1 - \tau} l_m(c, m) = - \frac{1 - \tau}{1 - \tau} L' \left( \frac{m}{c} \right) c^{k-1} \), where \( \tau \) is the tax on income. It follows that \( i \) is only zero when real balances are at the point of satiation. Therefore, it must be that \( \lim_{m \to \infty} S = \lim_{m \to \infty} \frac{m}{c} l_m(c, m) = 0 \), so that \( \lim_{m \to \infty} mL' \left( \frac{m}{c} \right) = 0 \).

\textsuperscript{4}If this was not the case, the government would have an incentive to set \( P_0 \) arbitrarily large, and that way to fully destroy the real value of outstanding liabilities. However, it would still have to use distortionary taxes in order to finance the future sequence of government expenditures.
nominal government bonds, which entitle the household to \((1 + i_t) B_t\) at \(t + 1\). Here, \(i_t\) is the opportunity cost of holding one unit of real money. Under the timing adopted for households’ decisions, it is also the inflation tax. Finally, the income tax is given by \(\tau_t \in [0, 1]\).

The government finances a constant flow of public expenditures \(g\) through an income tax and an inflation tax, and it issues debt. There is a cost in collecting revenue from the income tax, which is denoted by \(z_t\). I will first consider the case where marginal collection costs are increasing with the amount of revenue collected, so that \(z_t = \gamma \left[\tau_t (1 - h_t - s_t)\right]^2\), where \(\gamma \in [0, \infty)\) is the collection cost parameter. The government’s sequence of budget constraints is then given by

\[
M_0 + B_0 \geq 0,
\]

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

The presence of collection costs generates a loss of the economy’s resources, as reflected in the resource constraints:

\[
c_t + g \leq (1 - h_t - s_t) - z_t.
\] (6)

2.1 The Household’s Problem

For a given sequence of prices and taxes \(\{P_t, i_t, \tau_t\}_{t=0}^{\infty}\), the household chooses a sequence of quantities \(\{c_t, h_t, M_t, B_t, s_t\}_{t=0}^{\infty}\) that maximizes (1), subject to \(c_t \geq 0, M_t \geq 0, 0 \leq h_t \leq 1\), (2), (4), and (5).

From the first order conditions, the consumer’s optimal choice must satisfy:

\[
-(1 - \tau_t) l_m (t) = i_t,
\] (7)

\[
\frac{u_c (t)}{u_h (t)} = \frac{1}{1 - \tau_t} + l_c (t),
\] (8)

\[
(1 + r_{t+1}) = \frac{U_h (t)}{\beta U_h (t+1)} \frac{(1 - \tau_{t+1})}{(1 - \tau_t)},
\] (9)

where \(1 + r_{t+1} \equiv (1 + i_{t+1}) \frac{P_{t+1}}{P_t}\) and \(r_t\) is the real interest rate. Equation (7) says that the consumer’s optimal holdings of real balances are at the point where the opportunity cost of an additional unit equals the marginal benefit of it. Equation (8) equates the marginal rate of substitution between consumption and leisure to the ratio of their real prices, which can be written as \(l_c (t) / (1 - i_t)\). The real price of consumption is given by a direct component, the production cost of a unit of the consumption good, plus an indirect component, the cost in terms of lost wage income of the transactions time
necessary to enjoy an additional unit of the consumption good. \(1 - \tau_t\) is the real price of leisure, net of taxes. Finally, equation (9) determines the real interest rate. In steady state, this is constant and equal to \(\frac{1}{\beta} - 1\).

### 2.2 The Ramsey Problem

Since the government’s financing instruments are distortionary, finding the optimal inflation tax and income tax amounts to solving a second best, Ramsey problem.

For given initial conditions and government expenditures, the solution to the Ramsey problem is an allocation that maximizes social welfare, subject to the restriction that it can be decentralized as a competitive equilibrium with taxes.

The resources constraints together with an implementability condition ensure that this restriction is satisfied. The implementability condition is obtained by using the first order conditions of the household’s problem to replace the taxes and prices in terms of quantities into the consumer’s intertemporal budget constraint. Let \(Q_t = \frac{1}{(1+q_0)\ldots(1+q_t)}\) and \(d_t = \frac{Q_t P_t}{P_0}\). Since at the optimum \(\lim_{t \to \infty} (Q_t M_{t+1} + Q_t B_{t+1}) = 0\), we can write the intertemporal budget constraint as

\[
\sum_{t=0}^{\infty} d_t c_t + \sum_{t=0}^{\infty} d_t i_t m_t = \sum_{t=0}^{\infty} d_t (1 - \tau_t) [1 - h_t - l(t)].
\]

(10)

Using the homogeneity assumption and first order conditions (7), (8) and (9), to replace taxes and prices in the intertemporal budget constraint (10), the implementability condition can be written as

\[
\sum_{t=1}^{\infty} \beta^t \{c_t u_c(t) + (1 - k) u_h(t) l(t) - u_h(t) (1 - h_t)\} = 0.
\]

(11)

The optimal allocation \(\{c_t, m_t, h_t\}\) maximizes (1) subject to (11) and to the resources constraints

\[c_t + g \leq (1 - h_t - s_t) - \gamma [\tau (t) (1 - h_t - s_t)]^2.\]

(12)

Notice that \(\tau(t)\) denotes the expression for the income tax as a function of the quantities, as implied by the household’s optimality condition (8).

---

5Since the household’s budget constraint and first order conditions are not affected by the presence of collection costs, the implementability condition is also unchanged.
The first order condition of the Ramsey problem with respect to real money balances is given by

\[
\left[ \beta \psi u_h(t) (1 - k) \right] l_m(t) = \zeta_t l_m(t) - 2 \zeta_t \gamma \tau(t) [1 - h(t) - l(t)] l_m(t) + 2 \zeta_t \gamma \tau(t) \tau_m(t) [1 - h(t) - l(t)]^2,
\]

where \( \psi \) is the marginal excess burden of government expenditures (or of taxation), while \( \zeta_t \) is the shadow price of the economy’s resources, for \( t \geq 0 \). Both \( \psi \) and \( \zeta_t \) are greater than zero at the optimum.

The relevant question now is what is the optimal inflation tax when there are positive collection costs. Equation (7) ensures that the Friedman rule is optimal if \( l_m(t) = 0 \) satisfies (13), for all \( t \).

**Proposition 1.** Suppose that the alternative tax to inflation is an income tax and marginal collection costs are increasing. Then,

i) if \( \overline{m} \rightarrow \infty \), the Friedman rule is optimal for any finite level of the collection cost parameter \( \gamma \).

ii) if \( \overline{m} = \alpha c, \alpha \in [0, \infty) \), the Friedman rule is optimal when the transactions technology is such that \( l_m(c, \overline{m}) = 0 \) or when collection costs are zero (\( \gamma = 0 \)).

**Proof.** (Sketch) From equation (7), we know that the Friedman rule decentralizes a solution where \( m = \overline{m} \). Therefore, a necessary condition for the Friedman rule to be optimal is that \( \overline{m} \) satisfies equation (13). If the problem is globally concave, this is a necessary and sufficient condition. If the problem is not globally concave, this condition is only necessary. As it is common in this literature, I assume that, when \( \overline{m} \) satisfies equation (13), the allocation decentralized by the Friedman rule achieves the global maximum even when the problem is only locally concave. This finds support in numerical computations under several alternative specifications of the model. To prove the proposition, I distinguish two cases.

(i) \( \overline{m} \rightarrow \infty \). We know that \( \lim_{\overline{m} \rightarrow \infty} l_m(c, \overline{m}) = 0 \), so that the LHS of equation (13) tends to zero at satiation. If \( \gamma = 0 \), then the RHS approaches zero as \( \overline{m} \rightarrow \infty \), and the Friedman rule is optimal for any degree of homogeneity \( k \). If \( \gamma \in [0, \infty) \), the Friedman rule is still optimal since \( \lim_{\overline{m} \rightarrow \infty} \{ \tau_m \}_{m=\overline{m}} = 0 \). To show this, I differentiate (8) to obtain

\[
\tau_m|_{m=\overline{m}} = \frac{-u^{2} I(c, \overline{m})}{[u_{c} - u_{l} I(c, \overline{m})]}.\]

From homogeneity of the transactions technology and from the assumption that, when the unit tax on money

\footnote{See Section A in the Appendix for a complete proof of case (i).}
\( i \) tends to zero, revenues from seignorage approaches zero, it can be shown that \( \lim_{\overline{m} \to \infty} l_{cm}(c, \overline{m}) = 0 \), so that \( \lim_{\overline{m} \to \infty} \tau_m|_{m=\overline{m}} = 0 \). It follows that the RHS of equation (13) also approaches zero as \( \overline{m} \to \infty \).

(ii) \( \overline{m} = \alpha c, \alpha \in [0, \infty) \). Since \( l_m(c, \overline{m}) = 0 \), the LHS of equation (13) is zero at satiation. When \( \gamma = 0 \), the RHS is zero as well, and the Friedman rule is optimal for any degree of homogeneity \( k \). When \( \gamma \in [0, \infty) \), the Friedman rule is still optimal if \( l_m(c, \overline{m}) = L''(\frac{\overline{m}}{c}) = 0 \), because this implies that \( \tau_m|_{m=\overline{m}} \) and the RHS, evaluated at \( \overline{m} \), are zero. ■

When collection costs are zero, the general result in the literature is that money should not be taxed. A zero inflation tax is optimal for any degree of homogeneity of the transactions technology, as shown in Correia and Teles (1996).

When there are collection costs, the optimal inflation tax is generally different from zero. When equation (13) is not satisfied at the point of full liquidity, it is optimal to set a positive inflation tax and to correspondingly reduce the tax on income. The optimal policy mix minimizes the welfare losses due to the use of the two distortionary tax instruments. A positive inflation tax introduces a distortion in the households’ decisions to hold real balances. A decrease in the income tax reduces both the distortion in the households’ consumption-leisure choice and the loss in social resources due to the tax collection process. The two cases listed in Proposition 1, where a zero inflation tax is still optimal, are cases in which a marginal decrease in real balances does not allow to increase the revenue collected from seignorage, \( S = im \), and to correspondingly reduce the income tax. Therefore, it is not possible to save on the resources wasted in the tax collection process. Moving away from satiation can only increase the distortions in the economy, so that the optimal policy does not tax money.

Consider the case when \( \overline{m} \to \infty \). Since at that point the unit tax on money, \( i_t \), is zero, it is reasonable to assume that seignorage is zero. Moreover, as real balances approach satiation, the increase in seignorage associated to a marginal decrease in real balances also tends to zero. This is discussed in Appendix A. Since the authorities cannot save on the resources lost in the collection process, a zero inflation tax is optimal, as in the case with no collection costs.

A similar reasoning explains the optimality of the Friedman rule when \( \overline{m} \) is finite and \( l_{cm}(c, \overline{m}) = 0 \). From the assumptions on the transactions technology, \( l_{cm}(c, \overline{m}) = -C^\delta -2C^\gamma L''(\frac{\overline{m}}{c}) \). This can only be zero if \( L''(\frac{\overline{m}}{c}) \) is also zero. Using the definition of seignorage, \( S_m|_{\overline{m}} = -\overline{m} \frac{L''(\overline{m})}{c} c^{k-2} \). It
follows that, whenever $l_m(c, m) = 0$, the increase in seignorage associated to a marginal decrease in real balances tends to zero. Also in this case, it is optimal to follow Friedman’s prescriptions and not to tax money.

3 Alternative Model Specifications

It is useful to analyze how different model assumptions affect the conditions under which the Friedman rule is optimal, when there are collection costs. In what follows, I will derive the Ramsey solution when marginal collection costs are fixed, and when the alternative instrument to inflation is a tax on consumption.

3.1 Marginal Collection Costs are Constant

In section 2, I have analyzed the optimality of the Friedman rule under the assumption that total collection costs are quadratic, or that marginal collection costs increase with the revenues raised by the income tax. As it will be argued in section 5.2, evidence on the functional form of the cost schedule associated with the administration of the tax system is not conclusive, and different assumptions could be made about the losses in resources incurred in the economy. For instance, it could be argued that collection costs are mainly fixed costs so that marginal costs are negligible, or that marginal costs are positive but independent of the amount of revenue raised.

When collection costs are fixed, $z_t = \gamma$, for all $t$, there are levels of $\gamma$ so high that it is optimal not to use the income tax at all. In this case, the government would only use the inflation tax, as long as revenues from seignorage are large enough to cover public expenditures. When the level of collection costs is low enough as to make it optimal to use a mix of the inflation tax and the income tax, then the optimal policy is always given by the Friedman rule. In this case, it is optimal for the government to set a zero inflation tax and to finance government expenditures through the income tax, because there is no trade-off between a higher inflation tax and lower collection costs.

Consider now the case when the collection cost schedule is increasing, but marginal costs are constant, i.e. $z_t = \gamma \tau(t) [1 - h_t - l(t)]$. The first-order conditions corresponding to equation (13) are given by

$$\left[ \beta \frac{\psi}{\kappa_t} u_h(t) (1 - k) - 1 + \gamma \tau(t) \right] l_m(t) = \gamma \tau_m(t) [1 - h_t - l(t)], \ t \geq 0.$$  

(14)
Proposition 2. Suppose that the alternative tax to inflation is an income tax. If total collection costs rise with tax revenues but marginal costs are constant, Proposition 1 still holds.

Proof. Consider equation (14), and notice that $\tau_m|_{m=m}$ and $\theta_m(c, m)$ are still given by $\tau_m|_{m=m} = \frac{-\theta_m^2 \theta_m|_{c=m}}{|m_m - \alpha_m|c,m|}$ and $\theta_m(c, m) = -c^{k-2}m^2 L''(\frac{m}{c})$. Therefore, the conditions under which $m$ solves the Ramsey problem are unchanged.

Although the solution to the Ramsey problem is different under a schedule with constant marginal collection costs, the conditions under which the Friedman rule is optimal are unchanged. These are the conditions under which the trade-off between a positive inflation tax and lower collection costs disappears.

3.2 The Alternative to Inflation is a Consumption Tax

When the alternative to inflation is a consumption tax, this tax appears in the transactions technology. The reason is that income taxes are generally paid either as tax deductions on wages, or out of checking accounts every fiscal year. Instead, consumption taxes are paid each time a transaction is carried, and this requires households to increase the amount of cash that is necessary to purchase the same amount of the consumption good.

In the absence of collection costs, the general result when the alternative to inflation is a tax on income is the optimality of the Friedman rule. Mulligan and Sala-i-Martin (1997) argue that this is a fragile result because, when the alternative to inflation is a tax on consumption and taxes are paid with money, specific assumptions are necessary for the Friedman rule to hold. These assumptions relate to the time spent on transactions at satiation in real balances. De Fiore and Teles (1999) show that the claimed fragility of the Friedman rule is due to the peculiar specification of the transactions technology used in the literature, $s_t = l(1 + \theta_t)c_t, \frac{M_t}{P_t}$, where $\theta_t$ is the tax on consumption. Under this specification, a lower consumption tax helps saving on the resources used for transactions, so that a positive inflation tax becomes optimal. However, this formulation has an undesirable property. To see it, use the homogeneity property to rewrite the function as

$$s_t = (1 + \theta_t)^k (c_t, \frac{M_t}{P_t(1 + \theta_t)}).$$

(15)

Now consider a reduction of the consumption tax $\theta_t$. For a given real quantity of transactions in units of the consumption good, $c_t$, it is possible to change
\( \frac{M}{P_t} \) in such a way as to leave unchanged the real quantity of money required to buy those goods, \( \frac{\hat{M}}{\hat{P}_t (1 + \hat{\theta}_t)} \). Nonetheless, for any \( k > 0 \), purchasing the same real quantity of goods with the same real quantity of money requires less transactions time. Under the alternative specification proposed by De Fiore and Teles, \( s_t = \hat{\ell} \left( c_t, \frac{M}{\hat{P}_t (1 + \hat{\theta}_t)} \right) \), this undesirable property disappears. Since reducing the consumption tax cannot help saving on transactions time, it is always optimal to set a zero inflation tax, as in the case when the alternative is an income tax.

In what follows, I consider an economy where collection costs are positive, the government’s instruments are the inflation tax and a consumption tax, and the specification of the transactions technology is

\[
s_t = \hat{\ell} \left( c_t, \frac{m_t}{(1 + \theta_t)} \right) = \hat{L} \left( \frac{m_t}{c_t (1 + \theta_t)} \right) c_t^k,
\]

(16)

The assumptions made in Section 2 on the transactions technology \( l \) also apply to the function \( \hat{l} \). Notice that \( \hat{l} (t) \) and all its partials are now a function of \( \theta_t \), so that both the household’s problem and the Ramsey problem are changed. The household’s optimality condition are now given by

\[
- \frac{\hat{l}_m (t)}{(1 + \theta_t)} = i_t,
\]

(17)

\[
\frac{u_c (t)}{u_h (t)} = (1 + \theta_t) + \hat{l}_c (t).
\]

(18)

The Ramsey solution maximizes (1) subject to the resource constraints

\[
c_t + g \leq \left[ 1 - h_t - \hat{l} (t) \right] - \gamma \theta (t)^2 c_t^2, \quad t \geq 0,
\]

and the implementability condition (11), where \( \theta (t) \) defines the consumption tax in terms of quantities, as implied by the household’s optimality condition (18). The first-order condition with respect to real money balances can be expressed as

\[
\left[ \beta \frac{\psi}{\zeta_t} u_h (t) (1 - k) - 1 \right] \left[ \frac{1 + \theta (t) - m_t \theta_m (t)}{[1 + \theta (t)]^2} \right] \hat{l}_m (t) = 2 \gamma \theta (t) \theta_m (t) c_t^2.
\]

(19)

**Proposition 3.** Suppose that the alternative tax to inflation is a consumption tax, marginal collection costs are increasing, and taxes are paid with money. Then, Proposition 1 still holds.
Proof. Consider equation (19). From the first-order condition (18),

\[ \theta_m|_{m=\overline{m}} = -\frac{\tilde{l}_{cm}(c, \frac{m}{1-\theta})}{(1 + \theta)^2 - \overline{m} l_{cm}(c, \frac{m}{1-\theta})} . \]

Again, I consider two cases.

(i) \( \overline{m} \to \infty \). When \( \overline{m} \to \infty \), \( \tilde{l}_{cm}(c, \frac{m}{1-\theta}) \) approaches zero and so does the RHS, evaluated at satiation. Hence, the Friedman rule is optimal for all finite values of \( \gamma \).

(ii) \( \overline{m} = \alpha c, \alpha \in [0, \infty) \). When \( \gamma = 0 \), the RHS is zero as well, and the Friedman rule is optimal. When \( \gamma \in [0, \infty) \), the Friedman rule is still optimal if \( \tilde{l}_{cm}(c, \frac{m}{1-\theta}) = 0 \). ■

4 The Optimal Inflation Tax when the Friedman Rule is Not Optimal

In the previous analysis, I have derived the specific conditions under which the Friedman rule is optimal, despite the presence of collection costs.

The assumption of an arbitrarily large level of satiation is implicit in many models within the inventory-theoretic literature, which provides the micro-foundation of the transactions technology model. For instance, Baumol (1952) and Tobin (1956) formalize the cost of consumption in terms of time and money. They derive a technology of the form \( s_t = A \frac{m}{m^*} \), where \( A \) is the cost of withdrawing money from the bank in unit of time. Miller and Orr (1966) show that a similar technology, \( s_t = D \left[ \frac{c_t}{m_t} \right]^2 \), describes the costs of a firm’s cash management problem in a stochastic setting. Proposition 1 ensures that, under these specifications of the transactions technology, the Friedman rule is optimal for any finite level of the collection costs parameter and for any assumption on the marginal collection costs schedule. Nonetheless, an infinite satiation level should be seen as a conceptual limiting case rather than as a realistic assumption. Empirical evidence, as discussed in section 5, suggests that satiation occurs at a finite and relatively low level.

It is difficult to assess whether \( l_{cm}(c, \overline{m}) = 0 \) is a realistic assumption or not. One could think that it approximates the behaviour of economies where seignorage is low and where a change in the nominal interest rate is not likely to allow for substantial reductions in the alternative costly taxes. This seems to be the case for the United States. In 1993, the short-term commercial
paper rate was at 3.3 percent and seignorage was around 3 percent of all federal tax receipts.\footnote{Seignorage is computed as the short-term commercial paper rate times M1 deflated by the GDP deflator.} In 1995, the short-term commercial paper rate had almost doubled, being at 5.9 percent, while seignorage had only increased at around 4.5 percent of all federal tax revenues.

What happens when the conditions stated in Proposition 1 fail to hold? In this section, I analyze what are the main determinants of the optimal inflation tax, when the Friedman rule is not optimal. In order to do so, I start from a (not necessarily Pareto optimal) tax system, where both \( \tau \) and \( i \) are positive. Then, I consider a revenue neutral marginal tax reform that increases the inflation tax and correspondingly reduces the income tax, and I ask whether this reform leads to a Pareto improvement. Similar conditions are derived in a different environment by Faig (1988) and Mulligan and Sala-i-Martin (1997).

I carry the analysis under the assumption that the alternative to inflation is a tax on income. For analytical convenience, I use a dual approach to the solution of the Ramsey problem,\footnote{In the dual approach, the government’s problem is solved by choosing the taxes (rather than the quantities) that maximize the representative agent’s utility, subject to the individual and government budget constraints, and subject to the optimality conditions of the consumer’s problem.} and I assume that marginal collection costs are constant, i.e. \( z = \gamma \tau (1 - h - l) \).\footnote{When marginal collection costs are constant, it must be that \( 0 \leq \gamma \leq 1 \) for the government to use the income tax.} The results should extend to the case with increasing marginal collection costs, as suggested by the numerical exercise of section 5.3.

Since the solution is stationary, I will drop time dependence in the notation. The consumer now maximizes utility \( U(c, h) \), subject to the steady state form of the budget constraint

\[
c + im + \tau (1 - h - s) = 1 - h - s \equiv n.
\]

The first order conditions are given by (7) and (8), where the latter can be rewritten for convenience as

\[
D(c, h) \equiv \frac{u_c}{u_h} = l_c + \frac{1}{1 - \tau}.
\]

(20)

The second order conditions require that

\[
\Delta = -(D_c - DD_h) + l_{xc} - \frac{\gamma^2}{l_{mm}} > 0.
\]

(21)
The government maximizes the representative agent’s utility, subject to the consumer’s budget constraint, the government’s budget constraint, and the optimality conditions of the consumer’s problem. Its indirect utility function is given by

$$V(\tau, i) = U[e(\tau, i), h(\tau, i)]$$  \hspace{1cm} (22)

and the government budget constraint can be written as

$$G(\tau, i) \equiv \tau (1 - \gamma) [1 - h(\tau, i) - l(\tau, i)] + im(\tau, i) = g.$$  \hspace{1cm} (23)

I assume that the government is never on the downward sloping part of the Laffer curve, or that \( G_\tau, G_i \geq 0 \).

Consider a tax system \( \{\tau, i\} \) that finances a given flow of government expenditure \( g \), where both taxes are strictly positive. A marginal tax reform that increases the inflation tax and correspondingly reduces the income tax, leaving unchanged government revenues, is optimal when

$$\left. - \frac{di}{d\tau} \right|_{V=V} = \frac{V_x}{V_i} > \frac{G_\tau}{G_i} = - \left. \frac{di}{d\tau} \right|_{G=G}.$$  \hspace{1cm} (24)

Intuitively, it is optimal to increase the nominal interest rate when, compared to the inflation tax, the income tax leads to a large decrease in utility relative to the revenue it raises.

Define \( \varepsilon_{mi} \) as the absolute value of the interest rate elasticity of the money demand, \( \varepsilon_{nt} \) as the absolute value of the compensated labor supply elasticity with respect to the income tax,\(^{10}\) and \( \varepsilon_{me} \) as the scale elasticity of money demand.

**Proposition 4.** Suppose that the alternative tax to inflation is an income tax, marginal collection costs are constant, \( \bar{m} \) is finite, and the economy is at a point where \( \tau > 0 \) and \( i > 0 \). Then, a revenue neutral marginal tax reform that increases the inflation tax and decreases the income tax is optimal when

$$\frac{\tau c \varepsilon_{mi}}{(1 - \tau)^2 n} < \frac{\gamma \tau (1 + \frac{im}{c} \varepsilon_{me})^2 + (1 - \varepsilon_{me}) \left[ \tau (1 - \gamma) + (1 - \gamma \tau) \frac{im}{c} \varepsilon_{me} \right] \varepsilon_{nt}}{(1 - \gamma \tau) (1 + \frac{im}{c} \varepsilon_{me})^2 + (1 - \varepsilon_{me}) \left[ \tau (1 - \gamma) + (1 - \gamma \tau) \frac{im}{c} \varepsilon_{me} \right] \varepsilon_{nt}}.$$  \hspace{1cm} (25)

\(^{10}\)It is possible to express the compensated change in labor supply following a change in the income tax as \( \frac{\partial n}{\partial \tau} = \frac{\partial n}{\partial i} + \frac{\partial n}{\partial \tau} \), where \( W \) is the amount of transfers in units of the consumption good that is given to the agent in order to keep his utility unchanged. Then, the absolute value of the compensated labor supply elasticity with respect to the income tax is given by \( \varepsilon_{nt} = - \left( \frac{\partial n}{\partial \tau} + \frac{\partial n}{\partial i} \right) \).
Proof. (Sketch) The consumer’s first-order conditions implicitly define the demand functions $c(\tau, i)$, $h(\tau, i)$, and $m(\tau, i)$. Using the household’s budget constraints and first order conditions, I obtain expressions for $h_i$, $h_{\tau}$, $m_i$, $m_{\tau}$, $c_i$, and $c_{\tau}$. A marginal tax reform that increases the inflation tax and correspondingly reduces the income tax, leaving unchanged government revenues, is optimal when $V_{\tau} = \frac{|V_i|}{|V_m|} > \frac{G_i}{G_m}$. Taking derivatives of (22) and of (23) with respect to both $\tau$ and $i$, I can derive expressions for $V_{\tau}$, $V_i$, $G_{\tau}$, and $G_i$. Under the assumption that the government is on the upward sloping part of the Laffer curve, using the expressions for $m_i$, $h_i$, $c_i$, $m_{\tau}$, $h_{\tau}$, $c_{\tau}$, and the definitions of $\varepsilon_{mi}$, $\varepsilon_{n\tau}$ and $\varepsilon_{mc}$, the condition for the optimality of the marginal tax reform can be rewritten as condition (25).

Condition (25) is entirely stated in terms of quantities and elasticities. Proposition 4 says that, when $\varepsilon_{mc} < 1$, an increase in the inflation tax is optimal when the interest rate elasticity of money demand is small relative to the compensated labor supply elasticity. However, when $k = 1$, the level of the optimal inflation tax only depends on the interest elasticity of money demand and on the level of collection costs. The level of the compensated labor supply elasticity becomes irrelevant. Finally, when $\varepsilon_{mc} > 1$, the RHS of condition (25) is not necessarily increasing in $\varepsilon_{n\tau}$. An increase in the inflation tax could be optimal when the interest rate elasticity of money demand is large relative to the compensated labor supply elasticity.

Condition (25) confirms the optimality of the Friedman rule in the absence of collection costs. When $\gamma = 0$ and $i = 0$, then $\varepsilon_{mc} = 1$ and the RHS in condition (25) goes to zero. Thus, it is never optimal to increase the inflation tax and to deviate from the Friedman rule.

---

11See Section B in the Appendix for a complete proof.
5 The Optimal Inflation Tax in the U.S.

The aim of this section is to evaluate empirically whether the presence of collection costs justifies substantial deviations of the optimal inflation tax from Friedman’s prescriptions. In what follows, I first discuss available evidence from U.S. data on key parameters. Then, I calibrate the model presented in section 2 and I compute numerically the optimal inflation tax under alternative assumptions.

5.1 Evidence on the Relevant Elasticities

The analysis above makes it clear that the interest elasticity of money demand is an important determinant of the optimal inflation tax. Lucas (2000) argues that the best fit for the U.S. money demand (deflated by gross nominal income) for the period 1900-1994 is given by a log-log function with a constant interest elasticity, equal to one half. Mulligan and Sala-i-Martin (1996) criticize the fit of a log-log function because it extrapolates the behavior of agents at historically observed nominal interest rates to rates below 1 percent, which have not been experienced in the U.S. In order to estimate the interest elasticity of money demand at low interest rates, Mulligan and Sala-i-Martin consider a cross section of agents and their holding of assets. They argue that agents, when deciding whether to invest or not in interest bearing assets, look at the product of the interest rate times the total amount of their assets. This because there is a cost in acquiring a "financial technology", i.e. the information necessary to make accurate investment decisions, so that households might not want to invest in the technology when their total level of assets is low, particularly when interest rates are close to zero. The estimates they obtain do not support a log-log functional form. The interest elasticity of money demand is found to be close to 0.5 at a nominal interest rate of 6 percent, but close to zero at very low nominal interest rates.

Rejecting the log-log functional form for the money demand also implies imposing that satiation in real balances occurs at a finite level. To shed some light on possible values for the ratio $\frac{G}{y}$, one can look at the demand for money (deflated by gross nominal income) at values of the nominal interest rate very close to zero. Figure 1 plots the velocity of money in the U.S., computed as $V_t = \frac{M_t}{P_t y_t}$, for the period 1919-1996. $M_t$ is M1 and $P_t y_t$ is nominal income at current prices.\footnote{M1 is in billions of US dollars, not seasonally adjusted. For the period 1919-1970, it is from Friedman and Schwartz (1982) "A monetary History of the US: 1867-1960", Column 7, Table A-1, values in June of each year. For the period 1971-1996, it is from the Federal Reserve Bank of St. Louis, FRED Database, values in June of each year, series H.6.} The figure also plots a short-term nominal interest rate,
which captures the opportunity cost of holding money. Similar data are used in the analysis of the welfare costs of inflation of Lucas (2000). The inverse of velocity in periods of very low interest rates provides a lower bound for the ratio \( \frac{\pi}{\nu} \). Figure 1 shows that in the U.S., when the interest rate is at its lowest levels, around 1 percent, velocity remains close to two. This suggests a lower bound for \( \frac{\pi}{\nu} \) of 0.5. Figure 2 plots M1 deflated by nominal GDP as a function of the short-term nominal interest rate, showing a deflated money demand which is consistent with \( \frac{\pi}{\nu} \) being above 0.5.

When \( k \neq 1 \), the scale elasticity of money demand and the compensated labor supply elasticity also become important in the determination of the optimal inflation tax. Evidence on the scale elasticity of money demand is extensively discussed in Mulligan and Sala-i-Martin (1997). The estimates reported by the authors support values less than or equal to one. Evidence on the compensated labor supply elasticity is less conclusive. Mulligan and Sala-i-Martin (1997) argue that the compensated labor supply elasticity can be estimated by looking at the labor force participation of women, because there can be no wealth effect of a higher wage for a woman who initially does not work. They claim that existing evidence on female labor force points at elasticities around one or higher. Faig (1988) argues for a lower level of the compensated labor supply elasticity, based on existing estimates. For instance, Pencavel (1986) suggests values for the U.S. prime-age men of around 0.1. Killingsworth and Heckman (1986) report a wide range of results for female labor force, some being close to those of men. Stuart (1984) estimates an aggregate value for the compensated labor supply elasticities in the U.S. at around 0.5.


13 The nominal interest rate is the short-term commercial paper rate, in percentage per annum. For the period 1867-1975 it comes from Milton Friedman & Anna J. Schwartz "Monetary Trends in the United States and the United Kingdom." For the period 1976-1996 it is taken from the Economic report of the President (1997), Appendix B, table 71.
Figure 1: Velocity and the nominal interest rate

Figure 2: US Money Demand
5.2 Evidence on the Tax Collection System

Two assumptions about the tax collection system are important for the computation of the optimal inflation tax. The first is the exact definition of collection costs, or of the inefficiency in the tax system that we want to address. The second is the functional form of the marginal collection cost schedule and the definition of collection costs as fixed or variable costs.

I define collection costs as the sum of the government’s budgetary costs of administering the tax system and of the taxpayers’ costs of complying with the tax law. The government’s budgetary costs are incurred by the government while writing fiscal laws, by the Tax Court in judicails relating to tax issues, and by the IRS while providing services to the taxpayers, collecting tax revenues, enforcing payments or implementing audit. The taxpayers’ compliance costs are incurred while getting informed about tax requirements and possible deductions, keeping records throughout the year, filling tax returns, seeking the IRS assistance, or asking for professional advice.

Some of the government’s budgetary costs reported above are related to the possibility of evasion. It could also be argued that evasion is a cost for the government which derives from the use of taxes other than inflation, and that it should therefore be explicitly modeled when analyzing the costs of administering an income tax system. Including evasion in the representative agent framework of this paper does not require changing the structure of the model. Suppose that only a fraction α (τ), where α' (⋅) ≥ 0, of total tax liabilities is actually collected by the fiscal authorities. It can easily be verified that the model of section 2 remains unchanged, except that the tax rate τ is replaced by τ_E ≡ α (τ) τ, the effective tax rate.\(^\text{14}\) The main difference is that now we need to solve the model for the optimal combination of the inflation tax i* and of the effective income tax τ_E*, which corresponds to a higher value for the optimal official tax rate τ*. Notice that evasion reduces the amount of government revenue, as collection costs do. Nonetheless, the presence of collection costs generates both a reduction in government revenues and a loss in the economy wide resources. On the contrary, the amount of revenue evaded does not translate into a loss of resources for the economy, because it increases the net income of the representative household. This can be seen by summing the steady state versions of the household and government budget constraints in the model that allows for evasion, to obtain the economy

\(^{14}\text{A more general formulation would allow evasion to be a function both of the tax rate and of the level of enforcement and auditing, so that }\alpha (τ, e). \text{ However, the government would choose } e \text{ as a function of the amount of evasion; hence, at the optimum, } e^* = e (τ).\)
resource constraint
\[ c(\tau_E, i) + g = [1 - h(\tau_E, i) - l(\tau_E, i)] - \gamma \{\tau_E [1 - h(\tau_E, i) - l(\tau_E, i)]\}^2. \]
(26)

Equation (26) suggests that, in a representative agent model, the amount of revenue evaded does not matter for the computation of the optimal inflation tax. Evasion only matters as long as it imposes additional administrative costs, due to the necessity of enforcing payments or implementing audit. These additional costs are accounted for in the definition of collection costs given above.\(^\text{15}\)

To derive an estimate of collection costs as a percentage of tax revenues, I use data from the Internal Revenue Service, as reported in the Appendix of the 1997 Budget of the U.S. Government, and data from the 1999 Statistical Abstract of the United States. In 1995, total net revenue collected from the IRS amounted to approximately 18 percent of GDP. IRS reports a series of measures apt to capture the overall Service performance during the fiscal year 1995. One measure gives the amount of revenue collected per dollar of IRS budget, which is reported to be 172. This implies that around 0.6 percent of the revenue raised by the IRS was spent in administrative costs. The second measure gives the revenue collected per dollar of taxpayers’ burden, which is reported to be 10.97. This implies that the compliance costs paid by the taxpayers was around 9.1 percent of the revenue raised.

The figures obtained from the IRS performance measures on the taxpayers’ compliance costs are larger than those reported in the empirical literature on the costs of collecting taxes. A survey of the available evidence is given by Alm (1996). Slemrod and Sorum (1984) use evidence from a survey of Minnesota taxpayers to estimate the aggregate cost of complying with federal and state income tax returns. They find that in 1982 this cost was between 5 to 7 percent of the revenues raised by the federal and state income tax systems combined. Pitt and Slemrod (1989) calculate that the compliance costs of itemizing deductions in 1982 was around 0.5 percent of total revenues. There is little empirical work on the government’s administrative costs. Vaillancourt (1989) estimates these costs for Canada. He finds that the cost of collecting individual income, corporate income, and sales taxes generally exceeds one percent of the revenue that each tax raises.

In the benchmark case, I will calibrate the model such that collection costs are 10 percent of government revenues, i.e. \(\xi_g = .1\), in line with the IRS

\(^{15}\)In a model with heterogeneous agents or multiple sectors, evasion could call for a positive inflation tax because the underground sector could only be taxed through inflation. This is the standard “evasion” argument to a positive inflation tax, as considered by Nicolini (1998), but it is different from the collection cost argument analyzed here.
estimates. These estimates only refer to federal taxes, so that the calibration implicitly assumes that a similar percentage applies to state and local tax collection. The empirical evidence reviewed above would suggest that this figure is an upward biased estimate of total collection costs. Nonetheless, to check that the computed optimal inflation tax is not sensitive to errors in the measure of collection costs, I will also compute it when the share of collection costs to government revenues is doubled, i.e. $\tilde{\gamma} = .2$.

The second crucial assumption in the computation of the optimal inflation tax relates to the functional form of the marginal collection costs schedule and to the allocation of collection costs into a fixed and a variable component. To shed light on the second point, it may be useful to consider the outlays reported in the U.S. budget that relate to the administration of the tax system. At a federal level, the largest outlays are devoted to processing returns, the provision of assistance to taxpayers, inspection, investigation, enforcement activities, the provision of statistics on income and compliance, and the management of services. Most of these outlays appear to be mainly fixed costs, as small changes in tax rates do not significantly affect each single budget outlay. While changes in tax rates are likely to affect the government’s enforcement and auditing expenditure, this latter amounts to less than 1 percent of total government revenues from taxes. Slemrod (1985) estimates that small changes in tax rates do not significantly affect compliance costs either. Nevertheless, increases in government revenues are often achieved through new taxes or through a different regulation of the existing taxes and deductions, rather than through higher tax rates. There are significant fixed administrative and compliance costs associated with this alternative way of raising revenues. Hence, it is likely that the true functional form of the collection costs schedule, when there are several taxes alternative to inflation, is stepwise increasing.

5.3 Calibration and Numerical Results

In the numerical examples, I consider a CES instantaneous utility function

$$U (c_t, h_t) = \left( \frac{c_t^{\frac{\gamma - 1}{\gamma}} + v h_t^{\frac{\gamma - 1}{\gamma}}}{\gamma} \right)^{\frac{\gamma}{\gamma - 1}}.$$

I assume a transactions costs technology such that satiation can occur at any positive value of real balances. This allows to test the sensitivity of the optimal inflation tax to different levels of satiation. Also, the technology is such that the money demand shows a variable interest rate elasticity, which decreases with the level of the nominal interest rate. The interest rate elasticity depends upon the parameter values of the transactions technology;
however, it is always zero at a zero nominal interest rate. Finally, the transactions technology is assumed to be homogeneous of degree $k$, which allows to experiment with alternative steady state values of $\varepsilon_{mc}$. The functional form is given by

$$l_t = \left[ \eta \frac{c_t}{m_t} + \zeta \frac{m_t}{c_t} - \varepsilon \right] c^k,$$

where the function is defined over the range of parameters for which $l_t \geq 0$. I assume that $l(c, m) = 0$, or that the time spent transacting is zero at satiation. This requires setting $\varepsilon = 2\sqrt{\eta \zeta}$. The satiation point is given by $m = \sqrt{\frac{\eta}{\zeta}} c$, so that it changes with the values of the parameters. Figure 2 suggests that $\frac{m}{y}$ approaches 0.5 when $i$ is around 1%. Since in equilibrium $c < y$, $\left. \frac{m}{c} \right|_{i=1\%}$ has to be greater than 0.5, and even more so $\frac{m}{c}$. For a given value of the satiation level, there is a continuum of possible values for $\eta$. I choose a value of $\eta$ such that the absolute value of the interest elasticity of money demand is around 0.5 at a 6 percent nominal interest rate, in line with the estimate of Lucas (1998) and Mulligan and Sala-i-Martin (1997).

Concerning the tax system, I assume that marginal collection costs are increasing and that the fixed component is zero, for two reasons. First, I am interested in calculating an upper bound for the optimal inflation tax. Considering all fixed costs as variable will result in an upward biased estimate. If this estimate is low, I can be confident in concluding that collection costs do not justify large deviations from the Friedman rule. The second reason is that increasing marginal costs can be regarded as a smooth approximation of the stepwise function that would arise under several alternative taxes.

In the benchmark case, I set the parameter $\gamma$ such that, under the optimal policy mix, $z_{A^*} = .1$. Government expenditures are fixed at 18 percent of GDP. Figure 3 shows the solution of the model for this case, under the preferred parameterization $\gamma = 1.05$, $\frac{m}{c} = 1$, $k = 1$, $\eta = .006$, $\sigma = 1.8$ and $v = 1.3$. The parameters in preferences are set so that the absolute value of the compensated labor supply elasticity is between 0.5 and 1 in all the numerical examples reported below. At a 6 percent nominal interest rate, this benchmark parameterization implies that the absolute value of the interest rate elasticity is at .47, the compensated labor supply elasticity is at .56 and velocity $\frac{\mu}{m}$ is at 5.8. The scale elasticity is unitary given the assumption of $k = 1$. Figure 3 plots welfare as a function of the nominal interest rate, and the actual and fitted money demand. The resulting optimal inflation tax is very close to Friedman’s prescription, being at 0.2 percent.
To check the sensitivity of this result to the various assumptions, I compute the optimal inflation tax in five additional cases, as reported in table 1. These computations keep constant the share of government expenditures to GDP, \( \frac{G}{y} = .18 \), and the parameter values in preferences, \( \sigma = 1.8 \) and \( v = 1.3 \).

The results are obtained under two alternative values of the degree of homogeneity \( k \). Since the empirical evidence on \( \varepsilon_{mc} \) suggests values below or equal to one, I choose \( k = 1 \) and \( k = .5 \). Recall that, when \( k = 1 \), \( \varepsilon_{mc} = 1 \) for all values of the nominal interest rate. When \( k = .5 \), \( \varepsilon_{mc} < 1 \) for all strictly positive values of the nominal interest rate. Columns two to four list the results of the computations where either \( \gamma \), the collection cost parameter, or \( \zeta \), the transactions technology parameter, are changed, under the assumption that \( k = 1 \) and \( \eta = .006 \). Column two corresponds to the benchmark case represented in figure 3. Columns five to seven list the results when the parameters are changed in an identical way, under the assumption that \( k = .5 \) and \( \eta = .0025 \). The change in \( k \) is accompanied by a corresponding change in \( \eta \) which leaves the absolute value of the interest elasticity of money demand unchanged, at a nominal interest rate of 6 percent.

For each set of assumptions on \( k \) and \( \eta \), the optimal inflation tax is computed by varying \( \zeta \) in such a way that \( \frac{n^{\text{sat}}}{\gamma} \) takes a value of either 1 or .7. The table shows that lowering the satiation level has an effect on the steady state level of the interest elasticity of money demand. However, the optimal inflation tax, denoted as \( i^* \), only increases from 0.2 percent to 0.4 percent in
both cases. The corresponding optimal tax on income, $\tau^*$, remains around 32 percent.

For each set of assumptions on $k$ and $\eta$, the optimal inflation tax is also computed by varying the collection cost parameter $\gamma$. In the benchmark case, $\gamma$ is set such that, under the optimal policy mix, $\frac{\hat{z}}{g} = .1$. In the alternative case, I choose a value for $\gamma$ such that $\frac{\hat{z}}{g} = .2$. Under both sets of assumptions on $k$ and $\eta$, the optimal inflation tax remains low. The largest value, obtained when $k = .5$, $\eta = .0025$ and $\frac{\hat{z}}{g} = .2$, is still below 1 percent.

<table>
<thead>
<tr>
<th>$k = 1, \eta = .006$</th>
<th>$k = .5, \eta = .0025$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma$</td>
<td>1.05</td>
</tr>
<tr>
<td>$\frac{\hat{z}}{g}$</td>
<td>.1</td>
</tr>
<tr>
<td>$\zeta$</td>
<td>.006</td>
</tr>
<tr>
<td>$\frac{m}{\sqrt{t}}$</td>
<td>1</td>
</tr>
<tr>
<td>$\tau^*$</td>
<td>.319</td>
</tr>
<tr>
<td>$i^*$</td>
<td>.002</td>
</tr>
</tbody>
</table>

Table 1

<table>
<thead>
<tr>
<th>$\varepsilon_{mi}$</th>
<th>$\varepsilon_{mc}$</th>
<th>$\varepsilon_{nt}$</th>
<th>$V$</th>
</tr>
</thead>
<tbody>
<tr>
<td>.47</td>
<td>.76</td>
<td>.56</td>
<td>5.8</td>
</tr>
<tr>
<td>.44</td>
<td>.78</td>
<td>.56</td>
<td>6.1</td>
</tr>
<tr>
<td>.47</td>
<td>.76</td>
<td>.56</td>
<td>6.4</td>
</tr>
<tr>
<td>.43</td>
<td>.78</td>
<td>.56</td>
<td>6.2</td>
</tr>
<tr>
<td>.46</td>
<td>.78</td>
<td>.72</td>
<td>6.8</td>
</tr>
</tbody>
</table>

Two points emerge from the numerical exercise. The first is that the optimal inflation tax in the U.S. is not very sensitive to different assumptions about the satiation level in real money balances. In the limiting case when satiation occurs at an arbitrary large level of real balances, the optimal inflation tax is zero. In the benchmark case, where collection costs are 10 percent of government expenditures, the optimal inflation tax does not exceed half percent when satiation occurs at $\frac{m}{\sqrt{t}} = .7$. The second point is that, even when tax collection generates losses as large as 20 percent of government revenues, the optimal inflation tax only increases to less than 1 percent. For reasonable values of the real interest rate, this number implies a negative inflation rate in the economy.

## 6 Conclusions

In this paper, I have reconsidered the argument that, when it is costly to collect taxes alternative to inflation, it is always optimal to set a positive inflation tax. In the first part, I have shown that there are specific conditions
under which the Friedman rule is optimal despite the presence of collection costs. When these conditions do not hold, the level of the optimal inflation tax depends upon the collection cost schedule, as well as the interest elasticity of the money demand, the scale elasticity of money demand and the compensated labor supply elasticity.

In the second part of the paper, I have argued that the presence of collection costs does not justify substantial deviations from the Friedman rule in the United States. Despite the extreme assumptions that all collection costs are variable costs, that marginal collection costs are increasing, and that the tax collection process requires throwing away 20 percent of total government revenues, the computed optimal inflation tax remains below one percentage point.

References


A Proposition 1

Here, I give a complete proof of case i) in proposition 1. Consider the case when \( \overline{m} \to \infty \). We know that \( \lim_{\overline{m} \to \infty} l_m(c, \overline{m}) = 0 \), so that the LHS of equation (13) tends to zero at the satiation point. Consider the RHS. If collection costs are zero, i.e. \( \gamma = 0 \), then the RHS approaches zero as \( \overline{m} \to \infty \), and the Friedman rule is optimal for any degree of homogeneity \( k \).

If collection costs are positive and finite, i.e. \( \gamma \in [0, \infty) \), the Friedman rule is still optimal since \( \lim_{\overline{m} \to \infty} \{ \tau_m|_{m=\overline{m}} \} = 0 \). To see why it is so, implicitly differentiate (8), to get

\[
\tau_m|_{m=\overline{m}} = \frac{-u_n^2 l_{cm}(c, \overline{m})}{[u_c - u_h l_c(c, \overline{m})]^2}.
\]  

(27)

From the homogeneity assumption of the transactions technology, it can easily be verified that

\[
l_{cm}(c, \overline{m}) = -c^{k-2} \frac{\overline{m}}{c} L'' \left( \frac{\overline{m}}{c} \right).
\]  

(28)

From the assumption that \( l_m \leq 0 \), it follows that \( \lim_{\overline{m} \to \infty} L'' \left( \frac{\overline{m}}{c} \right) = 0 \). Then, one needs to determine \( \lim_{\overline{m} \to \infty} \overline{m} L'' \left( \frac{\overline{m}}{c} \right) \). Define seignorage as \( S = im \). As the nominal interest rate, which is the unit tax on money, tends to zero, it is reasonable to assume that revenue from seignorage also tends to zero. Then, \( \lim_{i \to 0} S = \lim_{\overline{m} \to \infty} (1 - \tau) ml_n(c, \overline{m}) = 0 \). From the assumptions that \( l_m \leq 0 \) and \( l_{mm} \geq 0 \), and that the derivatives are continuous, it must be that \( S \) is a smooth function, which approaches zero as \( \overline{m} \to \infty \). It follows that the first derivative of seignorage with respect to money tends to zero at the point of satiation (at \( i = 0 \)), i.e. \( \lim_{i \to 0} S_m = 0 \). Partially differentiating with respect to \( m \) implies that \( \lim_{\overline{m} \to \infty} \overline{m} l_{mm}(c, \overline{m}) = \lim_{\overline{m} \to \infty} L'' \left( \frac{\overline{m}}{c} \right) \frac{\overline{m}}{c} = 0 \), so that

\[
\lim_{\overline{m} \to \infty} \tau_m|_{m=\overline{m}} = \lim_{\overline{m} \to \infty} l_{cm}(c, \overline{m}) = 0.
\]

Hence, the RHS approaches zero at the limit, and the Friedman rule holds for all positive values of the collection cost parameter \( \gamma \). \( \blacksquare \)

---

\(^{16}\) See Correia and Teles (1996) for a proof that the Friedman rule is a solution for any positive \( k \), when there are zero collection costs. Note that when \( \gamma = 0 \), equation (13) is identical to equation (2.7) in Correia and Teles.
B Proposition 4

(i) The consumer’s first-order conditions implicitly define the demand functions \( c(\tau,i) \), \( h(\tau,i) \), and \( m(\tau,i) \). From equation (7), we know that the household’s demand for money is given by \( m = F(c, \frac{i}{1-\tau}) \). Hence, the household’s first order condition can be rewritten as

\[
l_m \left(c, F \left(c, \frac{i}{1-\tau}\right)\right) = -\frac{i}{1-\tau}.
\]

Now, take derivatives of both sides with respect to \( i \) and to \( \tau \), to get \( F_i = -1/[(1-\tau)l_{mm}] \) and \( F_{\tau} = -l_{cm}/l_{mm} \. Also, define the scale elasticity of money demand as \( \varepsilon_{mc} \equiv \frac{\partial m}{\partial c} = F_c c/m \), and the absolute value of the interest rate elasticity of money demand as \( \varepsilon_{mi} = -iF_i/m \).

(ii) The consumer’s budget constraint can be written as

\[
c(\tau,i) + im(\tau,i) = (1-\tau)[1-h(\tau,i) - l[c(\tau,i), m(\tau,i)]].
\]

(29)

Take derivatives of both sides of (29) with respect to \( i \), use equation (7) to substitute for \( i \), and get

\[
h_i = - \left(Dc_i + \frac{m}{1-\tau}\right).
\]

(30)

Take derivatives of both sides of (29) with respect to \( \tau \) to get

\[
h_{\tau} = - \left(Dc_{\tau} + \frac{(1-h-l)}{1-\tau}\right).
\]

(31)

Rewrite money demand as \( m(\tau,i) = F(c(\tau,i), \frac{i}{1-\tau}) \). Take derivatives of both sides of the money demand, \( m(\tau,i) = F(c(\tau,i), \frac{i}{1-\tau}) \), with respect to \( \tau \), substitute the expression obtained for \( F_{c} \) and \( F_{i} \), and get

\[
m_{\tau} = \frac{l_{m}}{(1-\tau)l_{mm}} - \frac{l_{mc}}{l_{mm}}c_{\tau}.
\]

(32)

Take derivatives of both sides of the money demand with respect to \( i \), substitute the expression obtained for \( F_{c} \) and \( F_{i} \), and get

\[
m_i = - \left[ \frac{1}{(1-\tau)l_{mm}} + \frac{l_{mc}}{l_{mm}}c_i \right].
\]

(33)

Now, rewrite the consumer’s first order condition with respect to consumption as

\[
D[c(\tau,i), h(\tau,i)] = l_c[c(\tau,i), h(\tau,i)] + \frac{1}{1-\tau}.
\]

(34)
Take derivatives of the consumer’s first order condition with respect to \( \tau \), substitute the expressions obtained above for \( h_\tau \) and \( m_\tau \), and use the definition given in (21) to obtain

\[
c_\tau = -\frac{1}{(1 - \tau) \Delta} \left[ D_h (1 - h - l) + \frac{l_m l_{en}}{l_{mn}} + \frac{1}{(1 - \tau)} \right]
\]

(35)

Take derivatives of the consumer’s first order condition with respect to \( i \), substitute the expressions obtained above for \( h_i \) and \( m_i \), and use (21) to obtain

\[
c_i = -\frac{1}{(1 - \tau) \Delta} \left[ D_h m - \frac{l_{en}}{l_{mn}} \right]
\]

(36)

(iv) A marginal tax reform that increases the inflation tax and correspondingly reduces the income tax, leaving unchanged government revenues, is optimal when

\[
\frac{V_\tau}{V_i} = \frac{|V_\tau|}{|V_i|} > \frac{G_\tau}{G_i}
\]

(37)

\( V(\tau, i) \) is the government’s indirect utility function, while \( G(\tau, i) \) is the government’s revenue raised by taxing money and income minus the revenue lost in collecting the income tax. Taking derivatives of (22) and of (23) with respect to both \( \tau \) and \( i \), we obtain:

\[
|V_\tau| = \frac{-(Dc_\tau + h_\tau)}{u_h} > 0,
\]

(38)

\[
|V_i| = \frac{-(Dc_i + h_i)}{u_h}, > 0
\]

(39)

\[
G_\tau = (1 - \gamma) (1 - h - l) - \tau (1 - \gamma) (h_\tau + l_c c_\tau + l_m m_\tau) + i m_\tau > 0,
\]

(40)

\[
G_i = -\tau (1 - \gamma) (h_i + l_c c_i + l_m m_i) + m + i m_i > 0.
\]

(41)

Notice that \( G_\tau \) and \( G_i \) are positive because I assume that the government is on the upward sloping part of the Laffer curve.

Now, plug the four expressions above in (37), use equations (30)-(33), (35), and (36) to substitute for \( m_k, h_\tau, c_i, m_\tau, h_\tau, \) and \( c_\tau \). Taking into account the sign of expressions (38) to (41), condition (37) can be expressed as

\[
[nc_i - mc_\tau] \left[ \frac{\tau (1 - \gamma)}{1 - \tau} + \frac{(1 - \gamma \tau) i m \varepsilon_{mc}}{c} \right] + \gamma n m > m \varepsilon_{mi} \left( \frac{1 - \gamma \tau}{1 - \tau} \right) \frac{c}{1 - \tau}.
\]

(42)
From equations (35) and (36), we have that

\[ nc_i - mc_r = \frac{m(1 - \varepsilon_{mc})}{(1 - \tau)\Delta}. \]  

(43)

Also, recall that \( F_c = -\frac{m}{m_{mc}} = \varepsilon_{mc}m/c_\text{c} \) and \( \varepsilon_{mi} = -iF_i/m \). Finally, notice that the compensated change in labor supply following a change in the income tax can be written as \( \frac{\partial n^c}{\partial \tau} = \frac{\partial n}{\partial \tau} + \frac{\partial n}{\partial W} n \), where \( W \) is the amount of transfers in units of the consumption good that is given to the agent in order to keep his utility unchanged. Then, the absolute value of the compensated labor supply elasticity with respect to the income tax is given by \( \varepsilon_{n\tau} = -\left( \frac{\partial n}{\partial \tau} + \frac{\partial n}{\partial W} n \right) / \left( \frac{\partial n}{\partial W} n \right) \). To derive \( \frac{\partial n}{\partial W} \), substitute \( h \) from the budget constraint in the first-order condition (34) and apply the implicit function theorem to obtain \( \frac{\partial c}{\partial W} = \frac{\partial h}{\partial (1-\tau)\Delta} \). From the budget constraint, it follows that \( \frac{\partial m}{\partial W} = \frac{\partial h}{\partial (1-\tau)\Delta} \left( 1 + \frac{im}{c} \varepsilon_{mc} \right) - \frac{1}{(1-\tau)\Delta} \). Use (31), (32) and (35) to obtain an expression for \( \frac{\partial n}{\partial \tau} \), then use \( \frac{\partial n}{\partial \tau} \) and \( \frac{\partial n}{\partial W} \) to derive \( \varepsilon_{n\tau} \) and rearrange to get

\[(1 - \tau)\Delta = \frac{\tau \left( 1 + \frac{im}{c} \varepsilon_{mc} \right)^2}{(1 - \tau)^2 n \varepsilon_{n\tau} - \tau im \varepsilon_{mi}}. \]  

(44)

Substituting (43) and (44) into (42), I obtain condition (25). ■
European Central Bank Working Paper Series


7. “A cross-country comparison of market structures in European banking” by O. de Bandt and E. P. Davis, September 1999.


18. “House prices and the macroeconomy in Europe: Results from a structural VAR analysis” by Matteo Iacoviello, April 2000.
19 “The euro and international capital markets” by Carsten Detken and Philipp Hartmann, April 2000.

20 “Convergence of fiscal policies in the euro area” by O. de Bandt and F. P. Mongelli, May 2000.


22 “Regulating access to international large value payment systems” by C. Holthausen and T. Rønde, June 2000.


27 “This is what the US leading indicators lead” by M. Camacho and G. Perez-Quiros, August 2000.


30 “A small estimated euro area model with rational expectations and nominal rigidities” by G. Coenen and V. Wieland, September 2000.


32 “Can indeterminacy explain the short-run non-neutrality of money?” by F. De Fiore, September 2000.

33 “The information content of M3 for future inflation” by C. Trecroci and J. L. Vega, October 2000.

34 “Capital market development, corporate governance and the credibility of exchange rate pegs” by O. Castrén and T. Takalo, October 2000.


36 “Measuring Core Inflation in the euro area” by C. Morana, November 2000.

"The optimal inflation tax when taxes are costly to collect" by F. De Fiore, November 2000.