

## OPTIMAL INFLATION\*

*Isabel Correia\*\***Pedro Teles\*\**

*How should monetary policy be carried out in the long-run? Recent literature proves the optimality of the Friedman rule, implying that the nominal interest rate should on average be zero. Prices are then expected to decrease over time, since the real interest rate is positive. According to the Friedman rule, the Government should not tax money, despite the need to resort to distorting taxes to finance public expenditure.*

## 1. INTRODUCTION

How should monetary policy be carried out in the long-run — and also in the short-run — in response to economic shocks? The study of the desirable monetary policy requires the prior identification of both long- and short-run effects of money. Only then follows the identification of the best policy strategy, i.e., the one leading to a better resource allocation in the economy. To attain this, it will also be necessary to use models that both reproduce the relevant facts and that address issue of optimality.

Long-run effects of monetary policy are well known and adequate models are available to measure such effects. On the contrary, research on the short-term effects of money still arouses some controversy as regards both the adequate theoretical model and the identification of the facts. Therefore, the answers to the question of how short-run monetary policy should be carried out are far from being definitive. Partly on these grounds, this article focuses on the presentation of the known findings on long-run optimal monetary policy.

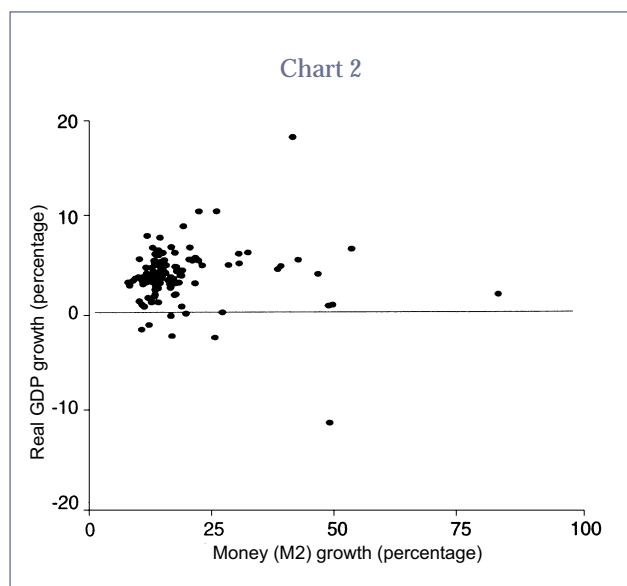
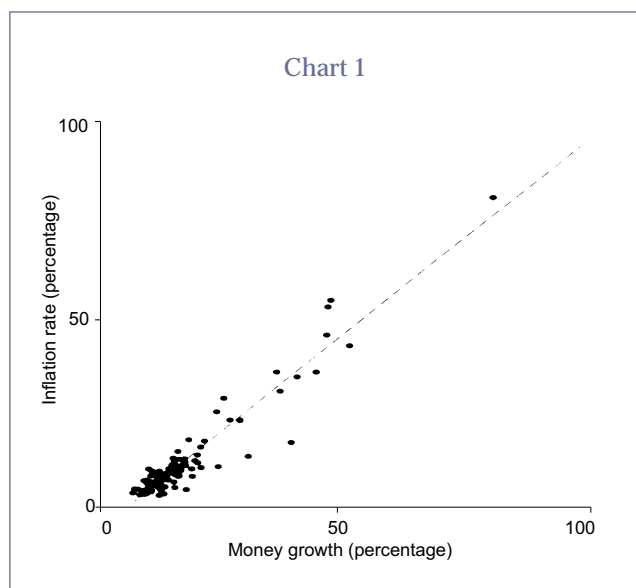
In the identification of the long-run effects of money, the quantitative theory is consensual.

Economies with higher money growth rates are economies with higher inflation rates and higher nominal interest rates. Effects on real interest rates and on growth rates are negligible (charts 1 and 2<sup>(1)</sup>). Even if there are no effects on growth there will be significant effects of inflation on resource allocation. Hence the issue of long-run monetary policy optimality is justified. Since it is a long-run policy, decisions concern the average growth rates of money and prices and secular averages of nominal interest rates. The distortions generated by a high average inflation rate or by high nominal interest rates are similar to those of any other tax. High inflation taxes transactions that use money, making consumption and investment costlier and deviating resources to leisure or to alternative ways of making transactions. As any other tax, a high average inflation also enables the government to collect more revenue, as it may finance deficits through the issuing of money, instead of paying high interest rates on the public debt. The objective of long-run monetary policy is to minimise the effect of distortions generated by the inflation tax, while taking into account that if this tax is cut, other taxes — also distorting — will have to be raised to finance public expenditure. Answers to the following issues are given in this article:

\* The opinions of the paper represent the views of the authors, and are not necessarily those of the Banco de Portugal. All errors and omissions are the authors' responsibility.

\*\* Economic Research Department.

(1) Lucas (1996).



what is the optimal average inflation level, when the government needs to collect distorting taxes to finance public expenditure? What should be the average growth rate of money? What is the nominal interest rate on longer-term bonds, resulting from optimal long-run monetary policy?

In the *Optimum Quantity of Money* (1969), Milton Friedman proposed a monetary policy rule able of generating the lowest nominal interest rates possible: “the rule on the optimum quantity of money is met through an inflation rate generating a nominal interest rate equal to zero”. The arguments defended by Friedman are simple Pareto optimality arguments which only hold true if taxes are non-distorting. A good with a zero production cost — indeed, money has very low marginal production costs — should have a price also equal to zero. As the nominal interest rate is the price of holding money — since it corresponds to private agents’ foregone revenue when they opt for holding this more liquid asset — the long-term nominal interest rate should, according to Friedman, be equal to zero. This rule for the nominal interest rate means that prices should decrease on average at a rate equal to that of the long-term real interest rate: the money stock must decrease at a rate consistent with the required deflation.

The leading criticism to the Friedman rule was made by Phelps (1973), who used the optimum taxation principles of Ramsey (1927): in the absence of non-distorting taxes, the optimal taxation problem consists of financing an exogenous se-

quence of public expenditure in the less distorting way. In this context, the marginal distortion caused by one unit of tax revenue should be equal for all taxes. Implicitly it would seem that money should also be taxed, as any other good, and therefore the price of money should be higher than its production cost. The long-term nominal interest rate should, therefore, be higher than zero.

Recent developments in the general equilibrium monetary theory questioned the intuition of Phelps and recovered the bounty of the Friedman rule. Despite the need to resort to distorting taxes, money should not be taxed. This finding holds true in an economic environment where money is necessary for transaction purposes, explicitly through a transactions function where money can be replaced with other production factors<sup>(2)</sup>. Considering that in this environment money is an intermediate good, the optimal taxation results of Diamond and Mirrlees (1971) concerning intermediate goods could apparently be invoked. These results suggest that under certain conditions intermediate goods should not be taxed. However, the conditions of the theorem of Diamond and Mirrlees (namely the linearity condition of the production function) do not occur necessarily in monetary models. For example if we think, as it is

(2) Kimbrough (1986), Faig (1986, 1988), Guidotti and Végh (1983), Chari, Christiano and Kehoe (1983) demonstrate the result, imposing restrictive conditions. Correia and Teles (1996) generalise these conditions.

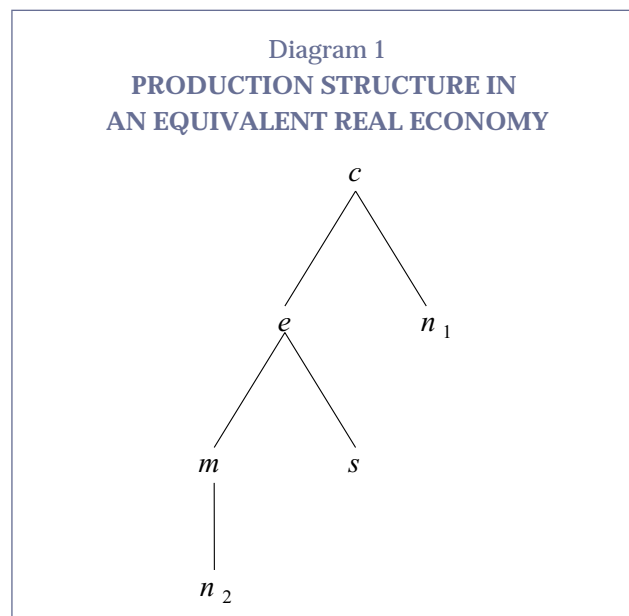
reasonable, that the transactions technology proposed by Baumol (1952) and Tobin (1956) is a good description of the transactions process, the structure of production no longer has constant returns to scale, hence making taxation on intermediate goods desirable.

In an environment closer to that used by Phelps (1973), where money is an input to the provision of liquidity services — modelled as a final good — Correia and Teles (1999) derived optimal taxation rules and concluded also in this context, that the Friedman rule is the general rule of optimum (non) taxation of money. Therefore, they showed that the intuition of Phelps cannot be applied to money, because money is a zero-cost good, taxed through a specific tax, the nominal interest rate. The optimum taxation results of Ramsey (1927), or Diamond and Mirrlees (1971), refer to *ad valorem* tax rates on goods with a positive production cost. It follows that the general result according to which these tax rates should be positive does not imply that the specific tax must also be positive, when the production cost of the good is close to zero. Indeed, in this case the optimal is also close to zero.

## 2. OPTIMAL INFLATION

This section describes in detail the optimal money taxation finding in Correia and Teles (1996). To address the issue of what should be long-term optimal inflation when all taxes are distorting, Correia and Teles (1996) use a monetary model where money is used in transactions in a way that the time spent in transactions is a function of the volume of transactions and the stock of money. In this model, money is an intermediate good necessary to carry out transactions. A possible justification for the transactions function — and the only one with a microeconomic theoretical foundation — is the transactions technology proposed in Baumol (1952) and Tobin (1956), according to which time spent in transactions is a function of the ratio of the volume of transactions per currency unit to the number of visits to the bank. This transactions function is homogeneous of degree zero.

In the model, there is a large number of households with endowments of time that can be used



for leisure, the production of an aggregate good, the production of transactions, or the production of money itself. Transactions have a cost, measured in terms of the time dedicated to that activity. Money can reduce this cost. This friction allows money to have a value. Households have preferences over consumption goods and leisure. Markets for goods and labour and markets for assets, money and nominal bonds exist in every period. A benevolent government chooses the optimal combination of income tax and inflation tax, which finance an exogenous sequence of public expenditures.

In this economic environment, where money is an intermediate good, Correia and Teles (1996) concluded that, when money has a negligible production cost, it is desirable that money is no longer taxed, regardless of the degree of homogeneity of the transactions function. If on the contrary money requires significant production costs, then taxing money would be optimal, the tax rate depending on the level of homogeneity of the transactions function.

The result that intermediate goods should not be taxed in a second best environment — when technology exhibits constant returns to scale — is well known since Diamond and Mirrlees (1971). They proved that production efficiency is a characteristic of the second best solution when taxes on consumption are available. As a corollary to this result, intermediate goods should not be taxed.

Production efficiency means that labour is optimally allocated — as in the first best — between different uses. This means that the marginal productivity of labour used in the production of a given good equals the marginal productivity of the intermediate good (used in producing that good), times the marginal productivity of labour (used in producing the intermediate good). In the monetary model with one aggregate good and no capital, the consumption tax proposed by Diamond and Mirrlees (1971) is equivalent to a single tax on labour, the intermediate goods being untaxed. These optimal tax rules are rules on the value of *ad valorem* taxes.

The taxation rules of Diamond and Mirrlees (1971) do not apply directly to the monetary economy for two reasons: because the production structure is a specific one, and because there are natural constraints on the taxes that can be collected. The distinctive features of the production structure in the monetary model are: first, the consumption good is produced using labour and transactions according to a Leontief production structure; second, the interesting transactions functions — like the Baumol-Tobin one — do not present constant returns to scale. The other distinctive feature is that time spent in the production of transactions cannot be taxed, since the activity of transactions does not feed through to the market.

In the context of the monetary model, efficiency in production is attained when money and time spent in its production are not taxed, and only time spent in the production of the good is taxed. If the transactions function presents constant returns to scale, production efficiency is desirable, and so money is not to be taxed. If, however, the transactions function does not exhibit constant returns to scale — as is the case of the Baumol-Tobin function — then distorting production would already be first best, and the optimal *ad valorem* tax on money would no longer be zero. But a positive *ad valorem* tax corresponds to a zero unit tax when the production cost of money tends to be zero. Since the inflation tax is a unit tax, the result of the optimality of the Friedman rule is ultimately explained by the free good characteristic of money.

To understand deeply these optimal money taxation results, it is useful to think of the monetary economy in terms of an equivalent real economy

with three vertical levels of production. The equivalent real economy is represented in Diagram 1. Here, economic agents have preferences over consumption,  $c$ , and leisure,  $h$ .  $c$  is produced using transactions,  $e$ , and labour,  $n_1$ , according to a Leontief production function,  $c = \min(e, n_1)$ . The production of  $e$  requires time,  $s$ , and an intermediate good,  $m$ . The intermediate good  $m$  is produced with labour,  $n_2$ , at a constant marginal rate ( $m = \alpha n_2$ ). Total time available in the economy is normalised to one unit. The taxation structure is such that  $c$ ,  $n_1$ ,  $n_2$  and  $m$  can be taxed, but  $e$  and  $s$  cannot be taxed. These constraints on the taxation capacity are natural constraints of the equivalent monetary model because transactions are not marketed.

Assuming that function  $s = l(e, m)$  is homogeneous of degree  $k$ , the optimal taxation solution is characterised by the following *ad valorem* tax rates on money

$$\begin{aligned} \tau_m &= 0, & \text{when } k &= 1 \\ \tau_m &> 0 & \text{when } k < 1 \\ \tau_m &< 0 & \text{when } k > 1 \end{aligned}$$

when the tax rate on labour used in money production is zero,  $\tau_2 = 0$ .

In this case, efficiency in the production of transactions,  $e$ , is optimal only when the production function of  $e$  is constant returns to scale. When there are profits, i.e., when the transactions function does not exhibit constant returns to scale, the effect of taxes on profits explains the deviations from production efficiency in the second best solution. When  $k \neq 1$ , the possibility of non-zero profits and the absence of a tax on those profits justifies optimal taxation rules that induce a reduction in profits. The reduction of profits, even when these are negative, is equivalent to a lump-sum tax. Thus the second best solution allows for a production distortion, through the taxation of intermediate goods, so as to reduce profits implicit in transactions production.

The reason why efficiency in the production of  $e$  is attained when  $\tau_2$  and  $\tau_m$  are equal to zero is that  $\tau_s$  is usually equal to zero. Therefore, suppressing taxation of labour used in the production of  $m$ , and taxation of  $m$  itself, maintains the efficiency in this transactions production branch.

Since transactions and hours worked are used in fixed proportions in the production of the con-

sumer good, production is not distorted by the taxation of  $n_1$ . For this reason the Ramsey solution, even with the above mentioned specific constraints of the tax system (i.e.,  $s$  and  $e$  cannot be taxed), is a second best, not a third or fourth best. If imposing  $\tau_2$  equal to  $\tau_1$ , production efficiency would imply a negative  $\tau_m$ . Therefore we can assert that the result achieved for constant returns to scale technologies (i.e., that  $m$  should not be taxed) ensures production efficiency, but due to the constraints of the taxation instruments it does not provide a natural extension to the Diamond and Mirrlees result. In this case, the intermediate good is not taxed but labour income is taxed at very different rates, depending on the sector where they were originated.

Income from labour in the production of money and in the production of transactions is not taxed, while income from labour in the production of the consumer good is taxed at a positive rate.

When  $m$  is a free good, if the nominal interest rate equals zero — meaning that money is being fully used ( $I_m = 0$ ) — the marginal effect of  $m$  on profit is zero. Despite the fact that, for transactions functions that are homogeneous of order  $k \neq 1$ , the level of implicit profits is different from zero and  $m$  generally exerts a marginal effect on profits. Moreover, at the satiation point of real money (i.e., the point where the free good has zero marginal productivity) this effect is null. Therefore the satiation point defines the optimum quantity of money. This finding can be interpreted as the limit result of the optimum unit tax charged on an intermediate good that uses resources, when the costs of producing the good become arbitrarily small. The intuition is that the unit tax equivalent to a finite *ad valorem* tax on a good with an arbitrarily low production cost, is arbitrarily low.

In any case, zero variable production costs of money stand as the essential assumption to derive the optimality of the Friedman rule. We take this assumption for granted despite the evidence of significant fixed costs associated with money creation. Therefore, money as a free primary input — and not as an intermediate good — is the relevant quantitatively reasonable assumption, as well as the fundamental theoretical justification behind the robustness of Friedman's optimality rule.

### 3. CONCLUSIONS

Long-run average inflation has real effects on the level of economic activity. To lessen these effects, the literature on long-run monetary policy rules recommends a policy that is consistent with close to zero nominal interest rates. According to the Friedman rule, of 1969, this corresponds to deflation. This result is surprising, since it holds even when the need the government has to resort to distorting taxes, to finance public expenditure, is taken into account (Correia and Teles, 1996; 1999). The basic intuition of the finding is that the nominal interest rate is a unit tax rate on a good (money) with a very low production cost. Therefore, even if in proportional terms it were optimal to tax money at a high rate, the equivalent specific tax is very low. Once optimum policy is defined, the quantitative issue of what are the welfare gains from reducing nominal interest rates to virtually null levels should be addressed. Correia and Teles (1994) calculate that the gain from reducing the nominal interest rate from 5 per cent to the Friedman rule amounts to about 1 per cent for GDP<sup>(3)</sup>.

This limit result — i.e., the optimality of non-taxation of money — can, however, be adjusted according to various considerations — for instance, taxation of the underground economy, high tax administration costs, or costs due to price changes. Since the underground economy is precisely one sector that cannot be taxed through the tax system, for efficiency and equity reasons, the inflation tax should be used to this end. In quantitative terms, the optimal long-run inflation levels are marginally positive<sup>(4)</sup>. High costs of collecting taxes on consumption or income can also explain a deviation from the Friedman rule — still a minor deviation, amounting to about one percentage point in the nominal interest rate<sup>(5)</sup>. Costs of price changes can also explain deviations from the Friedman rule, towards the price stability objective.

(3) See also Lucas (1994).

(4) Nicolini (1998).

(5) De Fiore (1998).

### REFERENCES

- Baumol, W.J., 1952, The Transactions Demand for Cash: An Inventory Theoretic Approach, *Quarterly Journal of Economics* 66, 545-556.
- Chari, V.V., L. J. Christiano and P.J. Kehoe, 1996, Optimality of the Friedman Rule in Economies with Distorting Taxes, *Journal of Monetary Economics* 37, 203-223.
- Correia, I. and P. Teles, 1994, Money as an Intermediate Good and the Welfare Cost of the Inflation Tax, Banco de Portugal, *Working Paper* no. 10, 1994.
- Correia, I. and P. Teles, 1996, Is the Friedman Rule Optimal when money is an Intermediate Good?, *Journal of Monetary Economics*, 38, 223-244.
- Correia, I. and P. Teles, 1999, The Optimal Inflation Tax, *Review of Economic Dynamics* 2, 325-246.
- De Fiore, F., 1998, The Optimal Inflation Tax with Costs of Collecting Taxes, mimeo, European University Institute.
- Diamond, P.A. and J.A. Mirrlees, 1971, Optimal Taxation and Public Production, *American Economic Review* 63, 8-27, 261-268.
- Faig, M., 1986, Optimal Taxation of Money Balances, Ph. D. Dissertation (Stanford University, Stanford, CA).
- Faig, M., 1988, Characterization of the Optimal tax on Money when it Functions as a Medium of Exchange, *Journal of Monetary Economics* 22, 137-148.
- Friedman, M., 1969, The Optimum Quantity of Money, in M. Friedman, ed., *The Optimum Quantity of Money and other Essays* (Aldine Chicago, IL).
- Guidotti, P.E. and C.A. Végh, 1993, The Optimal Inflation Tax when Money Reduces Transactions Costs, *Journal of Monetary Economics* 31, 189-205.
- Kimbrough, K.P., 1986, The Optimum Quantity of Money Rule in the Theory of Public Finance, *Journal of Monetary Economics* 18, 277-284.
- Lucas, Robert E., Jr., 1994, On the Welfare Cost of Inflation, mimeo, The University of Chicago.
- Lucas, R. 1996, Monetary Neutrality, *Journal of Political Economy*, 4, 104, 661-682.
- Nicolini, J. P., 1998, Tax Evasion and the Optimal Inflation Tax, *Journal of Development Economics*, 55 (1) 215-232.
- Phelps, E. S., 1973, Inflation in the Theory of Public Finance, *Swedish Journal of Economics* 75, 37-54.
- Ramsey, F.P., 1927, A Contribution to the Theory of Taxation, *Economic Journal* 37, 47-61.
- Tobin, J., 1956, The Interest Elasticity of the Transaction Demand for Cash, *Review of Economics and Statistics* 38, 241-247.