# Government Revenue from Inflation

Milton Friedman

*University of Chicago*

What rate of inflation will yield the greatest steady state command over real resources to a government having a monopoly on the issue of fiat money? The usual answer—the rate at which the inflation elasticity of demand for real balances is unity—is correct if real income is constant but wrong if real income is rising. The answer then depends also on the growth rate and on the income elasticity of demand for real balances. The revenue-maximizing rate of inflation is generally lower for growing than for constant real income and may even be negative, that is, deflation. Many actual rates of inflation seem higher than the revenue-maximizing rate.

It has become common to regard inflation produced by the issue of fiat money as a tax on cash balances. The simplest case arises when the government is the sole issuer of money and all money is non-interest bearing.[1] In this case, the real yield from the tax in equilibrium, when holders of cash are fully adjusted to the inflation, is typically taken to be equal to the rate of price rise times the real stock of money, which product in turn is taken as equal to the real value of the new money issued (see Friedman 1953; Bailey 1956; Cagan 1956; Johnson 1967). The rate of price rise is the rate of the tax. The real stock of money is the base of the tax. By strict analogy with an excise tax on a commodity, the yield is the product

The key idea of this paper originated in discussions at the Bank of Korea on the appropriate monetary policy for Korea. I am indebted to Governor Jin Soo Suh of the Bank of Korea and other members of the bank staff for hospitality that made my visit, though brief, highly productive. I am indebted also to members of the Money and Banking Workshop of the University of Chicago for helpful comments on an early draft of this paper.

[1] This involves no essential loss of generality. If there is competition in the creation of deposits and no restraints on the payment of interest on deposits, the more complex case involving deposits reduces to this one with "high-powered money" replacing the fiat money of our model. If there is an effective restriction of interest payments on deposits, a broader total than high-powered money replaces the fiat money of our model. In this latter case, the tax yield is shared between the issuers of high-powered money and the issuers (or favored borrowers) of other components of the broader total.

of the two. As for an excise tax, the yield is a maximum at the rate of tax (that is, rate of inflation) at which the elasticity of demand for real money balances with respect to the rate of inflation is equal to unity. For still higher rates of tax, the tax base declines by a larger percentage than the tax rate rises, so that the product declines.

This analysis is entirely correct for a stationary economy with fixed real income. But it is seriously misleading for a growing economy. For such an economy, the issuer of money obtains a yield from two sources: a tax on existing real cash balances; and provision of the additional real cash balances that are demanded as income rises. As in the stationary economy, the yield from the tax part is maximized at the unit-elasticity rate of price rise. However, the yield from the second part declines monotonically as the rate of price rise increases. As a result, the rate of price rise that will give maximum total yield is lower, and may be drastically lower, for a growing economy than the unit-elasticity rate.[2]

The question of the rate of inflation that will yield maximum revenue is especially topical for underdeveloped countries embarked on programs of development. For such countries, it has often been argued that inflation is desirable or inevitable. This argument has many facets. The one that is relevant here is that inflation is a form of taxation, the proceeds from which can be devoted to investment. Mundell (1965) and Marty (1967) have demonstrated that the possibilities along this line are extremely limited if there are no other sources of growth. However, when other sources of growth exist, their argument is in one sense too favorable to inflation, in another sense too adverse. It is too favorable because the revenue-maximizing rates of inflation are much lower than they implicitly assume. It is too adverse because growth increases the revenue that can be obtained by the issuance of money.

Of course, the revenue potentiality of money issue is only one element relevant to the desirability of using this device to raise revenue or of using the revenue so obtained to finance investment. This note is devoted solely to the technical issue of the revenue from inflation, not to the broader issue.[3]

[2] This point is implicit in Mundell (1965), but he does not recognize its importance because the only growth he considers is the growth which is produced by investment financed by government from the proceeds of the money issue. He assumes all other sources of growth to be zero. Marty (1967) makes explicit what Mundell leaves implicit and even gives the correct formula for total government revenue, but again he fails to recognize the importance of allowing for growth because he follows Mundell in considering only that growth produced by investment financed by the money issue. Both Mundell and Marty conclude that such growth is at best very small.

[3] My own personal view is that inflation is neither desirable nor necessary, that the most effective road to development is through free enterprise and private investment, and that the government can serve best by limiting itself to essential government functions, keeping taxes of all kinds low, refraining from intervention into the economy, and providing a stable monetary framework.

**Formal Analysis**

Let

$$M = \text{nominal quantity of money},$$

$$P = \text{price level},$$

$$N = \text{population},$$

$$m = \frac{M}{PN} = \text{real quantity of money per capita},$$

$$Y = \text{nominal income},$$

$$y = \frac{Y}{PN} = \text{real income per capita}.$$

Let $g$ with a subscript designate a *logarithmic* time derivative of the variable in the subscript, so $g_m = d(\log m)/dt = (1/m)(dm/dt) =$ percentage rate of growth of real quantity of money per capita.

Since we are interested in the equilibrium position after full adjustment to the inflation produced by money issue, we shall regard $g_P$ as simultaneously the actual and the anticipated rate of change of prices.

Let the demand for real money balances per capita be given by

$$m^D = f(y, g_P) . \tag{1}$$

Interest rates are omitted for simplicity, though insofar as the real interest rate is affected by the rate of inflation it should be taken into account, since different rates of inflation will then also affect the total yield by altering interest rates and thereby the amount of money demanded. However, this effect seems clearly of a smaller order than those that we shall consider. In the same spirit, we shall neglect any effect of the rate of inflation on the level or rate of growth of real income.[4] Equation (1) also assumes that the demand for money is homogeneous of first degree with respect to prices and population, so that the demand for aggregate nominal money is

$$M^D = NP \cdot f(y, g_P) . \tag{2}$$

Take logs of both sides, differentiate with respect to time, and assume equilibrium, so $M^D = M^S$ (money supplied) $= M$. This gives

$$g_M = g_N + g_P + \eta_{my} g_y , \tag{3}$$

where $\eta_{my} =$ elasticity of real per capita money balances with respect to real per capita income. In obtaining equation (3), the time derivative of $g_P$ is treated as zero, which means that equation (3) is valid only for alternative steady states of inflation. The yield from money issue in real

[4] That is, we shall neglect the only effect Mundell and Marty consider.

terms, call this $R$ (for revenue), is

$$R = \frac{1}{P}\frac{dM}{dt} = \frac{M}{P}\cdot g_M = \frac{M}{P}(g_N + g_P + \eta_{my}g_y)$$
$$= N\cdot f(y,g_P)(g_N + g_P + \eta_{my}g_y) . \tag{4}$$

Note that in the stationary case in which $g_N = g_y = 0$, equation (4) reduces to

$$R(g_N = g_y = 0) = \frac{M}{P}\cdot g_P , \tag{5}$$

or to the usual formulation, with $g_P$ the tax rate and $M/P$ the base.

To determine the value of $g_P$ which maximizes the revenue, differentiate $R$ as given by equation (4) with respect to $g_P$ and set equal to zero:

$$\frac{dR}{dg_P} = N\cdot f(y,g_P)\cdot\left(1 + g_y\frac{d\eta_{my}}{dg_P}\right) + N(g_N + g_P + \eta_{my}g_y)\frac{df(y,g_P)}{dg_P}$$
$$= \frac{M}{P}\left(1 + g_y\frac{d\eta_{my}}{dg_P}\right) + \frac{M}{P}\cdot\frac{1}{f(y,g_P)}(g_N + g_P + \eta_{my}g_y)\frac{df(y,g_P)}{dg_P} \tag{6}$$
$$= \frac{M}{P}\left[1 + g_y\frac{d\eta_{my}}{dg_P} + (g_N + g_P + \eta_{my}g_y)\frac{d\log m^D}{dg_P}\right] = 0 .$$

The rate of inflation that yields the maximum revenue is then the value of $g_P$ which satisfies the equation,

$$(g_N + g_P + \eta_{my}g_y)\frac{d\log m^D}{dg_P} + g_y\frac{d\eta_{my}}{dg_P} = -1 . \tag{7}$$

For a stationary economy, where $g_N = g_y = 0$, this reduces to the usual solution

$$g_P\cdot\frac{d\log m^D}{dg_P} = \eta_{mg_P} = -1 , \tag{8}$$

or the point of unit elasticity.

For a growing economy, for which $g_N$ and $g_y$ are both positive, it is clear that equation (7) would not generally be satisfied at the value of $g_P$ for which equation (8) is satisfied. To begin with, assume that the income elasticity is not affected by the rate of inflation, so the final term on the left-hand side of equation (7) is zero. Then at the value of $g_P$ at which equation (8) is satisfied, the first term on the left-hand side would be less than $-1$, that is, negative and larger in absolute value than unity. It follows that a lower value of $g_P$ (which would lower the absolute value of the left-hand side) would be required to maximize revenue. Now, let the second term on the left-hand side differ from zero. If it is negative—which is to say if the demand for money with respect to income is less elastic at high than at low rates of inflation—this term will make the whole left-hand side even larger in absolute value and require a still lower $g_P$. The only case in which the value of $g_P$ satisfying equation (8) might also satisfy equation (7) is if the income elasticity of money is greater at high than at low rates of inflation. In that case, the positive second term might offset the excess over unity of the absolute value of the first term.

There is no way to solve equation (7) as it stands explicitly for $g_P$. Similarly, we cannot determine explicitly the rate of monetary growth that will maximize yield. We can solve equation (7) for $g_N + g_P + \eta_{my} g_y$ and substitute the result in equation (3). That gives

$$g_M(\text{for maximum } R) = \frac{-1 - g_y \dfrac{d\eta_{my}}{dg_P}}{\dfrac{d \log M^D}{dg_P}} . \tag{9}$$

However, in general, the derivatives on the right-hand side are functions of $g_P$ and must be evaluated at the value of $g_P$ that satisfies equation (7). Hence this expression is not of much use.

## A Special Case

To get farther, let us consider the special case in which the demand function for money[5] is

$$m^D = l(y) e^{-bg_P} . \tag{10}$$

In this special case

$$\frac{d \log m^D}{dg_P} = -b , \qquad \frac{d\eta_{my}}{dg_P} = 0 , \tag{11}$$

so equation (7) can be solved for $g_P$:

$$g_P = \frac{1}{b} - g_N - \eta_{my} g_y ; \tag{12}$$

and equation (9) becomes

$$g_M(\text{for maximum } R) = \frac{1}{b} . \tag{13}$$

This final result is rather remarkable. For this special case, the rate of monetary growth is determined entirely by the slope of the semilogarithmic straight line giving the demand function for real money balances for a particular per capita real income. It does not depend at all on the rate of real income growth or the income elasticity of demand. However, these factors determine what rate of inflation will be associated with the

[5] This is the functional form used by Cagan (1956) to allow for the effect of $g_P$. He treats real income as constant, so does not deal with the effect of income.

revenue-maximizing rate of monetary growth. The higher the rate of growth of population and of real per capita income, and the higher the income elasticity of demand, the lower the rate of inflation.

Although the revenue-maximizing rate of monetary growth is in this special case independent of the other factors considered, the revenue obtained from monetary growth is not. The revenue is the rate of monetary growth times the real stock of money, or, as a percentage of income, the rate of monetary growth divided by income velocity. The higher the rate of inflation, the higher will tend to be the velocity and hence the lower the revenue. So whatever reduces the rate of inflation associated with the revenue-maximizing rate of monetary growth also raises the revenue.

## Illustrative Calculations

The importance of these factors can be illustrated by calculating equation (12) for plausible numerical values of the parameters.

*Value of* b.—The derivative of log $m^D$ with respect to the expected rate of change of prices should, in principle, be roughly equal to the derivative of log $m^D$ with respect to "the" nominal interest rate, since that interest rate, under the equilibrium conditions assumed here, will equal the expected rate of inflation plus the real rate of interest. Accordingly, we can use evidence from studies of the effect on the quantity of money demanded of either differential rates of inflation or different rates of interest.

Elsewhere, I have used an estimate of $b = 10$ as "rather on the high side" (Friedman 1969, p. 42).[6] This estimate, largely based on computed slopes for interest rates, is over twenty times as large as Cagan's average estimate of $b$ for the hyperinflations he studies (Cagan 1956, p. 45). However, Bailey has shown that the slope is almost surely much higher at low rates of inflation than at the rates that prevailed during hyperinflations (Bailey 1956, pp. 98–99). More recently, Klein has shown that the slopes generally computed for interest rates are biased downward by failure to allow for interest paid directly or indirectly on demand deposits (Klein 1970, chap. 3).

Klein himself obtains estimates of $b$ that range from 7 to 78, when he makes allowance for both interest paid on demand deposits and the regression bias (Klein 1970, table 10). A collection of studies (Meiselman 1970) yields estimates of the slope with respect to the rate of change of expected prices that vary from 0 to 3, with most in the interval from 1 to 3. These estimates are derived from time series studies for Chile (Deaver

[6] Note that $b$ has the dimensions of time so its value depends on the time units in which growth rates are expressed. Throughout, I treat them as per year.

1970), Argentina (Diz 1970), Korea and Brazil (Campbell 1970), and from a cross-section study for forty-seven countries (Perlman 1970). However, these estimates are probably underestimates of the number we seek for two reasons. First, they are computed for periods when the rate of inflation varied considerably and hence probably do not capture the full reaction to maintained inflation at a given rate. Second, they are all subject to a downward regression bias.

Accordingly, let us use three alternative values of $b$: 2, 10, and 20. These imply that the revenue-maximizing rates of monetary growth are 50 percent per year, 10 percent per year, and 5 percent per year, respectively. These would also be the rates of inflation if there were no growth.

*Income elasticity.*—The evidence suggests that the income elasticity is high in countries that are in the stage of rapid economic development and approaches unity for the more developed countries in which the financial structure is highly developed. For the United States from 1870 to 1914, the income elasticity was about 2.0, and similar values have recently prevailed in Japan and Korea. The studies referred to in connection with the value of $b$ generally give estimates of income elasticity between 1 and 2. Accordingly, we may use 1 and 2 as alternative values of the income elasticity.

*Population growth.*—Alternative values of 0 and 2 percent per year span much of the range that can be anticipated to prevail.

*Real per capita income growth.*—Experience here is very variable, from a negligible rate of increase in countries like India, to 2 percent a year in the United States over long periods, to 10 percent a year or so over a decade or more in Hong Kong, Taiwan, Japan, and Korea. Let us take three values to illustrate the magnitude of the effect: 0, 5, and 10.

Table 1 gives the revenue-maximizing rates of inflation corresponding

TABLE 1

RATES OF INFLATION YIELDING MAXIMUM REVENUE FOR ALTERNATIVE ASSUMPTIONS

| Assumptions about Population and Real Income Growth (% per Year) | | Revenue-maximizing Rate of Inflation (% per Year) If Value of $b$ (Demand Slope) Is | | | | | |
|---|---|---|---|---|---|---|---|
| | | 2 | | 10 | | 20 | |
| | | and Income Elasticity Is | | | | | |
| $g_N$ | $g_y$ | 1 | 2 | 1 | 2 | 1 | 2 |
| 0 | 0 | 50 | 50 | 10 | 10 | 5 | 5 |
| 0 | 5 | 45 | 40 | 5 | 0 | 0 | −5 |
| 0 | 10 | 40 | 30 | 0 | −10 | −5 | −15 |
| 2 | 0 | 48 | 48 | 8 | 8 | 3 | 3 |
| 2 | 5 | 43 | 38 | 3 | −2 | −2 | −7 |
| 2 | 10 | 38 | 28 | −2 | −12 | −7 | −17 |

to these assumed values of the parameters. One striking result from this table is the importance of the assumed value of $b$, especially at low levels. The reason is straightforward. The lower $b$, the more inelastic the demand for money with respect to inflation at each rate of inflation. The more inelastic the demand, the less the erosion in the base of the tax produced by raising the rate and hence the higher the rate that will yield maximum revenue. As $b$ goes to zero, that rate goes to infinity. I have included so low a value of $b$ as 2 precisely to bring out this point. My impression is that this is well below the value that actually prevails in most developing countries.

A second striking result is that allowance for growth clearly makes a substantial difference. For values of the parameters that are not at all implausible, it turns out that zero inflation or even a negative rate of inflation may yield a developing country more revenue from its monopoly of money issuance than a substantial rate of inflation.

## Contrast with Experience

There is no getting away from this conclusion in terms of our analysis. Yet this conclusion seems to run directly counter to casual observation of the behavior of developing countries. How can we explain the observed propensity of developing countries to inflate at rates that seem, on this analysis, higher than the rate that would maximize the government's revenue?

The main explanation, I believe, is shortness of time perspective—the difference between the immediate and longer-run effects of behavior. The preceding analysis is for the long run, for steady states. Start with any steady state, whatever it may be. An increase in the rate of monetary expansion will always increase revenue from the issuance of money for a time. Prices will not react instantaneously to the increased issue; and even when the rate of inflation reacts, real balances will not adjust instantaneously to the new rate of inflation. For a time, therefore, the government's revenue from money creation will be higher on both counts. The part which is a tax on existing real balances will be higher, and so will the amount that people add to their real balances. But this is temporary. As people adjust, they will reduce real balances. The tax revenue may still be higher but it will be offset by lower revenue and perhaps a negative revenue from the adjustment of real balances (velocity will rise or fall less rapidly than earlier). Given time for people to adjust to the new rate of inflation, the end result is certain to be a smaller revenue than was obtained immediately after the beginning of more rapid monetary expansion.

In a sense, the situation is an unstable political equilibrium. Governments tend to look little farther than the next election. If that election is close, an increase in the rate of monetary expansion is sure to provide the government with more revenue. The negative effects on revenue, let alone on more fundamental economic and social matters, will come later.

That, I believe, is the fundamental explanation why governments so often inflate at a higher rate than the rate that would yield the maximum revenue over a considerable period.

## Maximum Revenue and Optimum Monetary Growth

This analysis has so far been concerned entirely with government revenue from money issue. It has not taken into account the welfare loss from inflation—the loss to the holders of money because they are induced to hold less cash than they otherwise would (Bailey 1956). This loss can be interpreted as the cost of collecting the government revenue.

From this point of view, the optimum rate of monetary growth is that which would produce a rate of price decline roughly equal to "the" real interest rate (Friedman 1969, chap. 1). From equation (3) this is given by

$$g_M(\text{optimum quantity of money}) = g_N - \rho + \eta_{my} g_y , \qquad (14)$$

where $\rho$ is "the" real interest rate. Comparison with equation (13) shows that, at least for the special case for which equation (13) is valid, the revenue-maximizing and the optimum monetary growth rates are completely independent: the first depends only on $b$; the second does not depend on $b$ at all. The two could, by accident, coincide; or either could be larger. However, any revenue-maximizing monetary growth rate that produces inflation is of course larger than the optimum growth rate.

One special case of equation (14) may be of some interest. Suppose that $\rho = g_y$, for which there is some basis in experience and theory. In that case,

$$g_M(\text{optimum quantity of money}) = g_N + (\eta_{my} - 1)g_y . \qquad (15)$$

Prices decline at a rate equal to the rise in per capita output, which means that factor prices are constant. If, in addition, $\eta_{my} = 1$, the required monetary growth rate is equal to the rate of population growth. Alternatively, if $\eta_{my} = 2$, which is more reasonable for countries in an early stage of development, the required monetary growth rate is equal to $g_N + g_y$ or to the rate of growth of real output. Both of these are monetary rules that have frequently been suggested on very different grounds.

## Conclusion

It has always seemed natural to assimilate a government monopoly of fiat money issue with a private monopoly of a product produced at zero cost—Cournot's mineral spring. Yet we have just demonstrated that this is not correct. The owner of the spring will maximize his proceeds by charging a price at which the price elasticity of demand is unity. And this will be true whether the economy is stationary or growing. If it is growing, the only effect will be that his income will also grow, but each year he will maximize his revenue by pricing at the unit-elastic point on his demand curve. Why is this not also true for the government which can issue money at a zero marginal cost? Why will it not also maximize its revenue by pricing at the unit-elastic point on its demand curve, regardless of whether there is growth or not?

The answer is because there are two different prices that are relevant to money issue—one, the goods and services that are given up to acquire a dollar; the other, the number of cents per dollar per year that the money holder must spend to keep his real balances constant. Only the latter is analogous to the price of the mineral water. If the water were priced at zero, the proprietor of the spring would get no revenue regardless of growth. On the other hand, if the annual direct (not alternative) cost of holding money were zero, that is, prices were constant, the issuer of money would still get revenue if there were growth, because he would be able to acquire goods and services with the additional pieces of paper he prints.

The closest private analogy I have been able to construct is with a monopolistic producer and servicer of computers or other durable equipment who sells the equipment but charges an annual fee for servicing it. The higher the service fee, the lower the demand for the equipment at any sales price of equipment. It would pay the monopolist to charge a fee lower than that which maximized revenue from servicing a fixed number of machines in order to raise the number that could be sold each year at a given price.[7] But even this analogy is not complete because the monopolist has two prices to juggle: his service charge affects the demand for new machines but he can determine the price at which he will sell them. In the money case, the rate of inflation simultaneously determines the total sales price, as it were, of the new equipment that is added to the stock, and also the service charge. The money monopolist has only one variable (the rate of monetary growth) to juggle.

The distinction between the money monopolist and the mineral spring monopolist is perhaps the point of chief theoretical interest in this note. The point of chief practical importance is the strong presumption that money monopolists have not only imposed welfare loss on their communities by their actual inflationary policy but have obtained less

[7] In general, it will pay him to charge for service at a price equal to marginal cost, since that will maximize the value of the machine to the user, and to obtain his monopoly return entirely from the price he charges for the machine. The "in general" is necessary because of the possibility of using the two-price system as a vehicle for price discrimination.

revenue from their monopoly than they would have obtained by less inflation.

## References

Bailey, Martin J. "The Welfare Cost of Inflationary Finance." *J.P.E.* 64 (April 1956): 93–110.

Cagan, Phillip. "The Monetary Dynamics of Hyper-Inflation." In *Studies in the Quantity Theory of Money*, edited by Milton Friedman. Chicago: Univ. Chicago Press, 1956.

Campbell, Colin D. "The Velocity of Money and the Rate of Inflation: Recent Experiences in South Korea and Brazil." In *Varieties of Monetary Experiences*, edited by David Meiselman. Chicago: Univ. Chicago Press, 1970.

Deaver, John V. "The Chilean Inflation and the Demand for Money." In *Varieties of Monetary Experiences*, edited by David Meiselman. Chicago: Univ. Chicago Press, 1970.

Diz, Adolfo Cesar. "Money and Prices in Argentina, 1953–62." In *Varieties of Monetary Experiences*, edited by David Meiselman. Chicago: Univ. Chicago Press, 1970.

Friedman, Milton. "Discussion of the Inflationary Gap." In *Essays in Positive Economics*. Chicago: Univ. Chicago Press, 1953.

———. *The Optimum Quantity of Money and Other Essays*. Chicago: Aldine, 1969.

Johnson, Harry G. "A Survey of Theories of Inflation." In *Essays in Monetary Economics*. London: Allen & Unwin, 1967.

Klein, Benjamin. "The Payment of Interest on Commercial Bank Deposits and the Price of Money: A Study of the Demand for Money." Ph.D. thesis, Univ. Chicago, 1970.

Marty, Alvin L. "Growth and the Welfare Cost of Inflationary Finance." *J.P.E.* 75 (February 1967): 71–76.

Meiselman, David, ed. *Varieties of Monetary Experience*. Chicago: Univ. Chicago Press, 1970.

Mundell, Robert A. "Growth, Stability, and Inflationary Finance." *J.P.E.* 73 (April 1965): 97–109.

Perlman, Morris. "International Differences in Liquid Assets Portfolios." In *Varieties of Monetary Experiences*, edited by David Meiselman. Chicago: Univ. Chicago Press, 1970.