Zero Inflation and the Friedman Rule: A Welfare Comparison

Alexander L. Wolman

A distinct trend in recent years has been for central banks to emphasize low and stable inflation as a primary goal. In many cases zero inflation—or price stability—is promoted as the ultimate long-run goal (Federal Reserve Bank of Kansas City 1996). Economic theory also stresses the benefits of low inflation. However, in contrast to the current fashion among central banks, one of the most famous—and robust—results in monetary theory is that the optimal rate of inflation is *negative*: in many economic models in which money plays a role, welfare is maximized when the inflation rate is low enough so that the nominal interest rate is zero. Central bankers are certainly aware of this result, yet they never seriously advocate a long-run policy of deflation (negative inflation).

How much welfare is lost from a zero inflation policy as opposed to an optimal deflation policy? As shown below, the shape of the economy's money demand function with respect to nominal interest rates holds the key to answering the question. Lucas (1994) argues for a specification where real balances increase toward infinity as the nominal interest rate approaches zero. He finds that zero inflation is not much of an improvement over moderate inflation but that optimal deflation offers sizable benefits. The analysis in this article supports a different conclusion: reducing inflation from a moderate level to zero entails substantial welfare benefits, and the additional benefit achieved by optimal deflation is small. My analysis is based on estimating a general

This article is based on the third essay in my 1996 doctoral dissertation at the University of Virginia. I would like to thank Robert King, my dissertation advisor, for his support. Thanks also to Michael Dotsey, Robert Hetzel, Andreas Hornstein, and Thomas Humphrey for their comments. The views expressed here are the author's and not necessarily those of the Federal Reserve Bank of Richmond or the Federal Reserve System.

money demand function that nests the one preferred by Lucas. The estimates imply a *satiation* level of real balances, which proves to be important for the comparison of zero inflation and optimal deflation.[1]

The original analysis of the relationship between money demand and the welfare cost of inflation is credited to Bailey (1956). I review both Bailey's analysis and that of Friedman (1969), whose "Friedman rule" is the famous result previously mentioned. I then describe informally Lucas's (1994) recent work on quantifying the costs of deviating from the Friedman rule. Whereas Lucas's work is guided by inventory theory, my own estimates follow from a broader interpretation of the transactions-time approach to money demand. I use these estimates for welfare analysis similar to Lucas's. Although the analysis suggests that the Friedman rule may not offer much of a benefit in comparison to zero inflation, it does not explain why central banks do not choose to pursue deflation. I thus point out several channels absent from my analysis through which inflation may have welfare effects. These additional channels may help to explain why central banks seem content to shoot for zero inflation.

1. MONEY DEMAND AND THE WELFARE COST OF INFLATION

Bailey (1956) showed how a money demand relationship could be used to derive estimates of the welfare cost of inflation. He assumed a money demand function that gave real balances (M/P, where M is the nominal quantity of money and P is the price level) as a function of the nominal interest rate (R) and made a consumer surplus argument: just as the area under the demand curve for any good measures the total private benefits of consuming that good, so the area under a money demand curve represents the private benefit of holding money. At a nominal interest rate of 5 percent, since people are willingly giving up 5 cents per year per dollar of money held, the marginal benefit of holding the last dollar must be 5 cents per year. Similarly, at a nominal interest rate of zero, people are not giving up any interest payments to hold money, so the marginal benefit of holding the last dollar must be zero. At a social optimum, the marginal benefit to society of holding money should equal the marginal cost to society of producing money. With the reasonable simplifying assumption that the cost to society of producing money is zero, the optimal nominal interest rate is zero.[2] In a steady state the nominal interest rate is approximately equal to the real interest rate plus the inflation rate, so optimal policy, commonly known as the Friedman rule, involves deflation at a rate equal to the real interest rate.

[1] Chadha, Haldane, and Janssen (1997) have performed an analysis similar to this article using U.K. data. They emphasize a distinction between short-run and long-run money demand.

[2] Lacker (1996) reports manufacturing and operating costs for coin and currency of approximately 0.2 percent of face value.

With a nominal interest rate of zero as the optimal policy, it is possible to measure the cost of any inflation rate for a particular money demand function. Simply measure the area under the inverse money demand curve between the real balances corresponding to the Friedman rule and the real balances corresponding to the nominal interest rate in question.[3] That is, add up all of the marginal benefits that are foregone by following a suboptimal policy; those marginal benefits are measured by the nominal interest rate (the inverse money demand function) at each level of real balances.[4] At this point the term "cost of inflation" may seem misleading; according to the theory sketched above, it would be more appropriate to use the term "cost of positive nominal interest rates." Since the former term is so widely used, however, I will stick with it.

Particular theories of money may imply more complicated money demand relationships than the one assumed by Bailey; for example, the analysis in Section 3 will involve consumption and the real wage as arguments in the money demand function. However, it is still the case that the Friedman rule is optimal, and holding consumption and the real wage constant, the area under the inverse money demand curve still provides an approximate measure of the direct cost of inflation.[5]

While the optimality of the Friedman rule holds as long as real balances are a decreasing function of the nominal interest rate (subject to the caveats in Section 5), the welfare costs of inflation can vary with the money demand function in two ways. First, the overall benefit of reducing inflation from, say, 10 percent to the Friedman rule can vary. Second, the apportionment of that benefit may vary, in the following sense. According to one money demand function, reducing inflation from 10 percent to zero may generate 99 percent of the total welfare benefit, with the remaining reduction to the Friedman rule adding essentially nothing. Another function could reverse this; reducing inflation from 10 percent to zero might generate only 1 percent of the total welfare benefit, with the remaining reduction to the Friedman rule being crucial for generating any significant benefits. This article is concerned mainly with the latter issue.

Lucas (1994) contrasts the welfare implications of two particular money demand functions, both of which specify the ratio of real balances to real

[3] The standard money demand curve expresses real balances as a function of nominal interest rates, whereas the inverse money demand curve inverts this relationship to express nominal interest rates as a function of real balances.

[4] This measure of the cost of inflation does not take into account indirect effects of inflation, as will be explained in Section 4. I thus refer to the area under the money demand curve as a measure of the direct cost of inflation.

[5] If the money demand relationship involves variables other than the nominal interest rate, the area under the inverse money demand curve ($R(m)$) only approximates the direct cost of inflation, because these other variables will generally vary across different values of the nominal interest rate.

consumption as a function of the nominal interest rate. The ratio of real balances to consumption is used because the money demand functions discussed here are assumed to apply to long-run data, and in the long run real balances move roughly one for one with consumption.[6] In the first specification, semi-log, there is a fixed relationship between the change in the nominal interest rate and the percentage change in the real balances to consumption ratio. That is, if the nominal interest rate rises from zero to 1 percent, the percent decrease in real balances/consumption is the same as if the nominal interest rate rises from 5 percent to 6 percent. In the second specification, log-log, there is a fixed relationship between the *percentage* change in the nominal interest rate and the percentage change in the real balances to consumption ratio. Thus an increase in the nominal interest rate from zero to 1 percent will cause a much larger percentage drop in real balances/consumption than an increase in the nominal interest rate from 5 percent to 6 percent. Note that if the log-log relationship is taken literally, the ratio of real balances to consumption must be infinite when the nominal interest rate is zero.

How do the two specifications compare in terms of welfare? With the log-log function, a slight increase in the nominal interest rate near zero generates a tremendous decline in the ratio of real balances to consumption. Using Bailey's (1956) reasoning, there must be a significant welfare cost of deviating just slightly from the Friedman rule. The semi-log specification generates smaller costs of slight deviations from the Friedman rule but roughly the same benefits of reducing inflation from, say, 5 percent to zero. Lucas argues that for the United States, the log-log specification fits the data more closely than the semi-log specification.[7] Most of the benefits to reducing inflation would then accrue only if the inflation rate were made negative, as it would need to be in order to achieve the Friedman rule. In his own words, "log-log demand implies a substantial gain in moving from zero inflation to the Friedman optimal deflation rate needed to bring nominal interest rates to zero, while under semi-log demand this gain is trivial" (Lucas 1994, p. 5).

Is log-log demand an accurate characterization of the data? Lucas argues that it is more accurate than semi-log demand, but is it reasonable to restrict the search to those two alternatives? Answering these questions requires one to be explicit about a model of money demand.

[6] Lucas refers to the ratio of real balances to income. In his model there is no investment, so consumption equals income. The model I will use does have investment, and the appropriate ratio will be real balances to consumption rather than real balances to income.

[7] Lucas (1994, p. 3) plots semi-log and log-log functions for various interest semi-elasticities and elasticities and concludes that "the semi-log function . . . provides a description of the data that is much inferior to the log-log curve."

2. THE TRANSACTIONS-TIME APPROACH TO MONEY DEMAND

Economists have developed a wide range of models of money, none of them entirely satisfactory. The models that are most appealing in terms of their microfoundations—that is, their descriptions of the obstacles that individuals overcome by holding money—tend to be ill-suited to quantification (e.g., estimating the welfare cost of inflation in the United States). An example is the search-theoretic class of models developed by Kiyotaki and Wright (1989).[8] On the other hand, those models that *are* easiest to quantify do not convincingly describe the obstacles that cause individuals to hold money. Examples include the money-in-the-utility function and cash-in-advance approaches (Sidrauski 1967 and Lucas and Stokey 1983, respectively). A middle ground is the transactions-time approach, developed by McCallum (1983) and McCallum and Goodfriend (1987). Their fundamental assumption is that consumption requires time spent shopping (or transacting), and transactions time may be decreased by holding a greater quantity of real balances. The analysis in this article will be conducted in the transactions-time framework.

Denoting transactions time in period t by h_t, and the transactions-time function by $h(c, m)$, the assumptions that transactions time is increasing in consumption and decreasing in real balances mean that $\partial h/\partial c > 0$ and $\partial h/\partial m < 0$. I make the further assumption that the function is homogeneous of degree zero in c and m: if c and m increase or decrease by the same percentage, then transactions time is unchanged. It follows that only the ratio of m to c matters: $h_t = h(m_t/c_t)$. Lucas (1994) shows that the transactions-time approach can be explicitly linked to earlier inventory-theoretic models of money demand developed by Baumol (1952) and Tobin (1956). The simplest inventory-theoretic model corresponds to the transactions-time technology,

$$h(m_t/c_t) = \kappa \cdot (m_t/c_t)^{-1}, \tag{1}$$

where κ can be interpreted as a fixed cost of replenishing money holdings.[9] More complicated inventory-theoretic approaches can be shown to imply similar $h(.)$ functions, with the difference being that m/c would be raised to some power less than -1:

$$h(m_t, c_t) = \kappa \cdot (m_t/c_t)^{-1/\gamma}, \gamma \in (0, 1). \tag{2}$$

See Lucas (1994).

[8] This is not to rule out the possibility that in the future, search-based models will be useful for quantitative exercises.

[9] While McCallum and Goodfriend interpreted $h()$ in terms of shopping time, Lucas interpreted it as going-to-the-bank time.

The inventory-theoretic interpretation imposes strong restrictions on the form of the transactions-time technology and hence, as I will describe below, on the form of the money demand function. Specifically, for the transactions-time technology, it implies that no matter how high the ratio of real balances to consumption, there is still some additional benefit to increasing that ratio further. Lucas (1994, p. 16) defends this implication as follows: “Managing an inventory always requires *some* time, and a larger average stock must always reduce this time requirement, no matter how small it is.” One cannot argue with this statement, according to a narrow interpretation of what it means to manage an inventory. However, holding a higher inventory of real balances also requires increased resources to protect the inventory, a point made by Friedman (1969, p. 17), who described a shopkeeper hiring guards to “protect his cash hoard.”

Given an arbitrary transactions-time technology, the associated money demand function can be derived by specifying some additional features of the economic environment. Assume that individuals face a budget constraint,

$$P_t c_t + M_t + \frac{B_t}{1+R_t} = M_{t-1} + B_{t-1} + P_t w_t n_t + D_t, \tag{3}$$

and a time constraint,

$$n_t + l_t + h_t = 1, \tag{4}$$

where P_t is the price level, M_t is nominal money balances ($m_t p_t$), B_t is holdings of one-period nominal zero-coupon bonds maturing at $t+1$, R_t is the interest rate on bonds, w_t is the real wage, n_t is the fraction of time spent working, D_t is dividend payments from firms, l_t is the fraction of time spent as leisure, and h_t is the fraction of time spent carrying out transactions. In a given period, individuals’ sources of funds are the money balances with which they enter the period, the bonds they redeem, the wage income they earn, and the dividends they receive from firms. These sources fund current consumption and money balances and bonds to carry over into the next period.

Deriving the money demand function requires knowing what it means for an individual to hold an optimal quantity of real balances. Optimal behavior involves balancing marginal benefit and marginal cost. What are the marginal benefit and marginal cost of holding money? From Section 1, the marginal cost of an additional dollar is the interest foregone in the next period (R_t); the marginal benefit of an additional dollar is the decrease in transactions time that it brings about. This decrease in transactions time is $-h'(m_t/c_t)\cdot\frac{1}{P_t\cdot c_t}$, and the extra time can be spent in the labor market earning the nominal wage ($P_t \cdot w_t$). Since marginal cost is measured as of the subsequent period, marginal benefit needs to be adjusted correspondingly: current period labor earnings can be invested in the bond market, so their value tomorrow is $\left(-P_t \cdot w_t \cdot (1+R_t)\cdot \frac{1}{P_t\cdot c_t}\cdot h'(m_t/c_t)\right)$.

Equating marginal cost and marginal benefit implies

$$-h'(m_t/c_t) = \frac{R_t}{1+R_t} \cdot \frac{c_t}{w_t}, \tag{5}$$

which can be used to confirm the Friedman rule result: at a nominal interest rate of zero, money holdings are chosen so that the marginal benefit of an additional unit of money is zero.

Under the inventory-theoretic interpretation, as mentioned earlier, the marginal benefit of an additional unit of money is never zero. Combining (5) with the specification in (2), the strictly positive marginal benefit of additional real balances corresponds to infinite real balances at the Friedman rule ($R = 0$):

$$\frac{\kappa}{\gamma} \cdot (m_t/c_t)^{-1-1/\gamma} = \frac{R_t}{1+R_t} \cdot \frac{c_t}{w_t}, \gamma \in (0, 1]. \tag{6}$$

The inventory-theoretic approach has appeal, but the implication that real balances would be infinite at the Friedman rule is extreme and argues for considering transactions-time technologies that do not share that implication. If real balances are finite at the Friedman rule, there is some quantity of real balances at which the marginal benefit of holding an additional unit of real balances is zero. That level of real balances—if it exists—will be referred to as the *satiation* level. A key proposition, namely that the welfare gains from low nominal interest rates are concentrated near the Friedman rule, depends crucially on the assumption of no satiation level; the log-log money demand function does not have satiation, whereas the semi-log function does.

The log-log function is roughly consistent with inventory theory: assuming that c and w are constant, and noting that $R_t/(1 + R_t) \approx R_t$, (6) yields a nearly linear relationship between the log of real balances and the log of the nominal interest rate. In contrast, the semi-log function is inconsistent with inventory theory, as it posits a linear relationship between the log of real balances and the *level* of the nominal interest rate. Thus Lucas's purely empirical argument favoring the log-log specification over semi-log is strengthened by his theoretical argument favoring the inventory approach. However, inventory-theoretic models do not offer the only alternative to semi-log money demand. And the fact that the inventory approach implies infinite real balances at a zero nominal interest rate suggests searching across a wider class of models. In the next section, I present estimates of a money demand function that allows for satiation and is consistent with the basic assumptions of the transactions-time model. This function nests nonsatiation (log-log) as a special case, but for many parameter values it is not consistent with inventory theory.

3. ESTIMATES OF A GENERAL MONEY DEMAND FUNCTION

From (5), in order for a transactions-time function to be consistent with satiation, it must be that for some positive value of m/c, further increases in that ratio do not decrease transactions time. In (6), under the inventory approach, transactions time is always decreasing in m/c, so subtracting a constant from the left-hand side of (6) will yield a technology consistent with satiation. That is, if $h'(m_t/c_t) = \phi - (\kappa/\gamma) \cdot (m_t/c_t)^{-1-1/\gamma}$, with $\phi \geq 0$, then the implied transactions-time technology allows for satiation. Since it will be convenient below to specify the parameters in a slightly different way, I define $\nu \equiv -\gamma/(1+\gamma)$, and $A \equiv (\kappa/\gamma)^{-\gamma/(1+\gamma)}$, so that $h'(m_t/c_t) = \phi - A^{-1/\nu} \cdot (m_t/c_t)^{1/\nu}$, with $\nu < 0,\ A > 0$. The technology can be found by integrating the previous expression:

$$\begin{gathered} h(m_t/c_t) = \phi \cdot (m_t/c_t) - \frac{\nu}{1+\nu} A^{-1/\nu} \cdot (m_t/c_t)^{\frac{1+\nu}{\nu}} + \Omega,\ \text{for } m_t/c_t < A \cdot \phi^{\nu}, \\ h(m_t/c_t) = \Omega,\ \text{for } m_t/c_t \geq A \cdot \phi^{\nu}, \end{gathered} \tag{7}$$

where Ω is a nonnegative constant that represents the minimum possible transactions time. This function is decreasing in m_t/c_t as long as m_t/c_t is less than $A \cdot \phi^{\nu}$, and the satiation level of real balances is given by $(m/c)_s = A \cdot \phi^{\nu}$. If $\phi = 0$, then there is no satiation level, and the function is consistent with inventory theory. The implicit money demand function is given by

$$A^{-1/\nu} \cdot (m_t/c_t)^{1/\nu} - \phi = \frac{R_t}{1+R_t} \cdot \frac{c_t}{w_t}, \tag{8}$$

which can be rewritten to yield an explicit money demand function:

$$m_t/c_t = A \cdot \left(\frac{R_t}{1+R_t} \cdot \frac{c_t}{w_t} + \phi \right)^{\nu}. \tag{9}$$

My strategy now is to estimate A, ϕ, and ν using (9) and to test the hypothesis that there is no satiation level of real balances ($\phi = 0$). The theory as presented thus far suggests that (9) should hold exactly. Of course it does not; I choose to model the error term as additive, but the estimation results do not change significantly if the error is assumed to be multiplicative. The data, which are from the United States for the period 1915 to 1992, are described in the appendix.

Although four separate variables enter (9), for estimation purposes it is simplest to define the two composite variables, $y_t \equiv m_t/c_t$ and $x_t \equiv [R_t/(1+R_t)] \times [c_t/w_t]$. Then the estimation equation is

$$y_t = A \cdot (x_t + \phi)^{\nu} + \varepsilon_t. \tag{10}$$

Figure 1 displays a plot of y_t versus x_t. Estimates of A, ν, and ϕ are found by solving the following nonlinear least squares (NLS) problem:[10]

$$\min_{\widehat{A},\widehat{\nu},\widehat{\phi}} \sum_{t=1}^{T} \left(y_t - \widehat{A} \cdot (x_t + \widehat{\phi})^{\widehat{\nu}} \right)^2 . \tag{11}$$

In general, the NLS estimates are consistent and asymptotically normal, as shown by Amemiya (1985, pp. 127–35); here I do not make any distributional assumptions about ε_t, the residual. Confidence intervals for the parameters were generated by bootstrapping, which allows one to construct a sampling distribution without any distributional assumptions and without relying on the accuracy of linear approximations.[11]

Table 1 contains estimates for A, ν, and ϕ, along with centered 95 percent confidence intervals. Although the estimated value of ϕ is close to zero, the implied satiation level of m/c is fairly low, 2.674. Following Amemiya (1985, p. 136), I construct a t-test of the nonsatiation hypothesis ($\phi = 0$). The test statistic is 26.99, meaning that nonsatiation is overwhelmingly rejected. Using the sampling distribution for the parameters A, ϕ, and ν, Figure 2 plots the implied sampling distribution for the satiation level of m/c. According to the sampling distribution, 90 percent of the probability mass for the satiation ratio lies below a value of 5. However, the right-hand tail of the distribution is fat; the x-axis would need to go all the way to 46,000 to encompass 97.5 percent of the probability mass, meaning that the satiation level is imprecisely estimated. This imprecision follows from the properties of the data: the lowest nominal interest rate in the sample is 0.7 percent, and for the observations with the lowest nominal interest rates, there is substantial variation in the ratio of real balances to consumption.[12] The solid line in Figure 1 shows the fitted values.

[10] The presence of consumption in the numerator of x and the denominator of y can cause the NLS estimator to be biased, as it may induce a correlation between the residual (ε_t) and x. More generally, if the residual represents a shock to the transactions-time technology, then in general equilibrium such a correlation would arise even without consumption on both sides of the estimation equation. I have investigated these problems by estimating with instrumental variables using the generalized method of moments (GMM). The GMM estimates are highly sensitive to the choice of instruments, so I report only the NLS results.

[11] The bootstrapping approach involves three steps. The first step is to produce the NLS estimates. The second step is to fit an AR model to the NLS residuals, producing a new set of disturbances, $\widehat{e_t}$, that are approximately white noise (an AR(2) was fit to $\widehat{\varepsilon_t}$ to produce $\widehat{e_t}$). The final step is to draw randomly with replacement from the $\widehat{e_t}$, producing N new vectors, $\widetilde{y_t}$, each of size T. For each of those new samples the parameters are estimated by nonlinear least squares. The $\widetilde{y_t}$ are generated by combining the x_t data and the random draws of $\widehat{e_t}$ with the initial parameter estimates.

[12] Working in a different money demand framework, Mulligan and Sala-i-Martin (1996) have developed a method of estimating the behavior of money demand near zero nominal interest rates. Their fundamental insight is that if there is a fixed cost of holding nonmonetary assets, the behavior of individuals who hold only monetary assets at positive nominal interest rates can yield information about aggregate money demand at a nominal interest rate of zero.

Figure 1 Data and Predicted Values

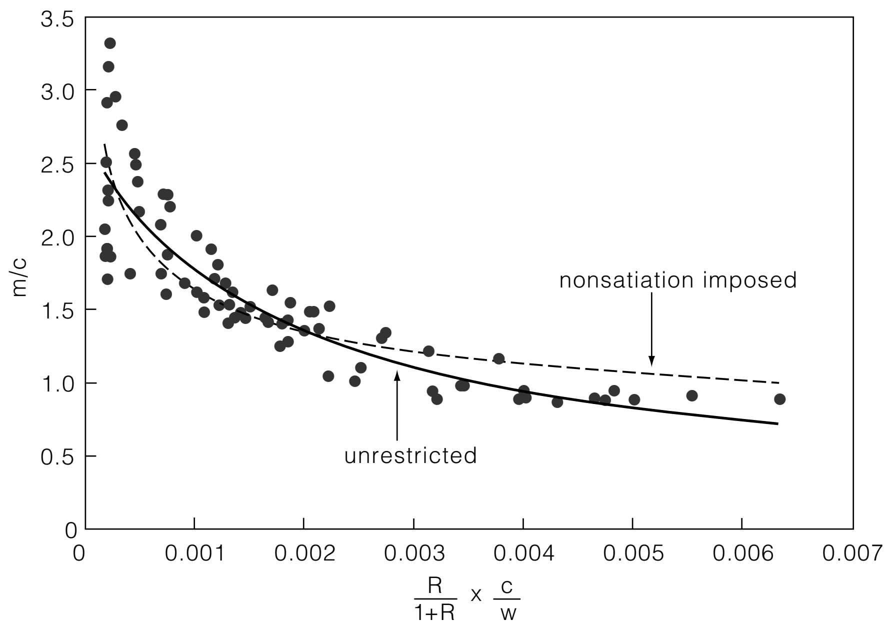


Table 1 Unrestricted Estimates and 95 Percent Confidence Intervals

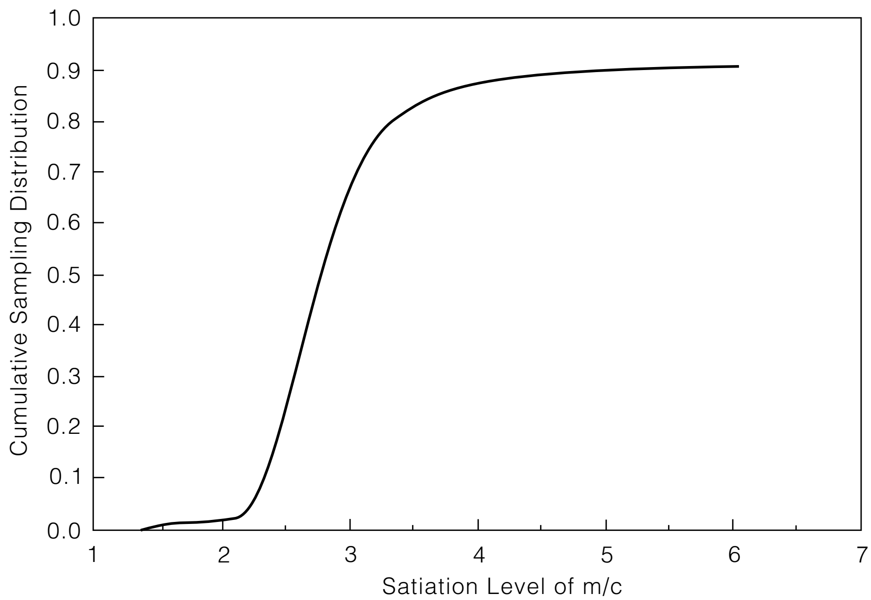
$\widehat{A}$	$\widehat{\nu}$	$\widehat{\phi}$
0.01702	-0.7695	0.001399
$(6.1 \times 10^{-7}, 0.421)$	$(-3.31, -0.20335)$	$(2.5 \times 10^{-19}, 0.0131)$

For comparison purposes, I also estimated A and ν under the nonsatiation restriction. Table 2 contains the estimates, and the dashed line in Figure 1 shows the fitted values when nonsatiation is imposed. With money demand estimates in hand, we can now look at their implications for the welfare cost of inflation.

Table 2 Restricted Estimates: Nonsatiation

$\widehat{A}$	$\widehat{\nu}$
0.2526	-0.2699

Figure 2 Distribution of Estimated Satiation Level



4. WELFARE ANALYSIS

By specifying a general equilibrium model, I can use the above estimates of the transactions-time technology to compute the exact welfare cost of inflation. I use a standard real business cycle model, as in Prescott (1986) or King, Plosser, and Rebelo (1988), augmented by the transactions-time money demand specification to answer the following question: in a world of constant 5 percent inflation, how much income would an individual willingly forfeit (or require) in order to live in a world with some lower (or higher) constant inflation rate?[13]

The economy consists of a representative individual who chooses consumption and money balances and leisure to maximize lifetime utility:

$$E_t \sum_{j=0}^{\infty} \beta^j u(c_{t+j}, l_{t+j}),$$

[13] For the purpose of computing this welfare measure, I define *full income* as the sum of consumption and $w \cdot l$, where w is the real wage and l is leisure. The computation holds the real wage constant at its benchmark level. That is, what amount of additional full income at the old real wage would give the individual the same utility as the decrease in inflation under consideration?

where $u(c, l) = \ln(c) + \psi \ln(l)$. This maximization is subject to the budget constraint (3), the transactions-time technology, and the time constraint (4). Optimal choices of consumption, leisure, bond holdings, and money holdings imply

$$u_c(c_t, l_t) = \lambda_t \cdot P_t \cdot \left(1 + w_t h'(\frac{c_t}{m_t})(\frac{1}{m_t})\right), \tag{12}$$

$$u_l(c_t, l_t) = \lambda_t \cdot w_t \cdot P_t, \tag{13}$$

and

$$1 + R_t = E_t \frac{\lambda_t}{\beta \lambda_{t+1}}, \tag{14}$$

as well as the money demand relationship (5). In these expressions λ_t is the shadow price of nominal wealth—the multiplier on (3). Since consumption requires a time expenditure, there is a wedge between the marginal utility of consumption and the marginal utility of wealth in (12). That wedge, $\lambda_t \cdot w_t \cdot P_t \cdot h'(\frac{c_t}{m_t}) \cdot (\frac{1}{m_t})$, is the value in utility terms of the marginal transactions time associated with an additional unit of consumption. The efficiency condition for leisure, (13), sets the marginal utility of leisure equal to the marginal utility of foregone earnings, and the efficiency condition for bond holding, (14), describes the equivalence between having \$1 of wealth today and \$$(1 + R)$ of wealth tomorrow. An additional equation defines transactions time as (7).

Firms produce the economy's single good using capital, which they own, and labor, which they hire on a period-by-period basis, according to a constant returns to scale production function,

$$y_t = a_t f(k_t, g^t n_t), \tag{15}$$

where y_t is output, a_t is a random productivity factor, k_t is the capital stock, and g is the exogenous growth rate of labor-augmenting technical progress. In a steady state, the exogenous technical progress will mean that output, consumption, real balances, the capital stock, investment, and the real wage will also grow at rate g. Capital accumulates according to

$$k_{t+1} = k_t \cdot (1 - \delta) + i_t, \tag{16}$$

where i_t is investment and δ is the depreciation rate. Since firms own the capital stock, they earn rents in equilibrium; those rents are paid out as dividends to individuals, who own the firms. Firms maximize the expected discounted stream of future profits—all of which are paid out as dividends—where the discount rate for period $t + j$ is the consumer's marginal rate of substitution between a dollar of wealth in periods t and $t + j$:

$$V_t = \text{Max } E_t \sum_{j=0}^{\infty} \beta^j \cdot \frac{\lambda_{t+j}}{\lambda_t} \cdot \left(P_t \cdot a_t \cdot f(k_t, g^t n_t) - w_t \cdot P_t \cdot n_t - P_t \cdot i_t\right).$$

This maximization is subject to (15) and (16). Thus the firm's first-order condition with respect to next period's capital stock is

$$P_t = \beta \cdot E_t \frac{\lambda_{t+1}}{\lambda_t} \cdot \left(P_{t+1} \cdot a_{t+1} \cdot f_k(k_{t+1}, g^{t+1} n_{t+1}) + P_{t+1} \cdot (1 - \delta)\right). \tag{17}$$

According to (17), the decrease in current-period profit that results from a marginal increase in investment should be exactly offset by the increase in future profits associated with a higher capital stock next period. Optimal choice of labor input implies that the real wage equals the marginal product of labor:

$$a_t \frac{\partial f(k_t, g^t n_t)}{\partial n_t} = w_t. \tag{18}$$

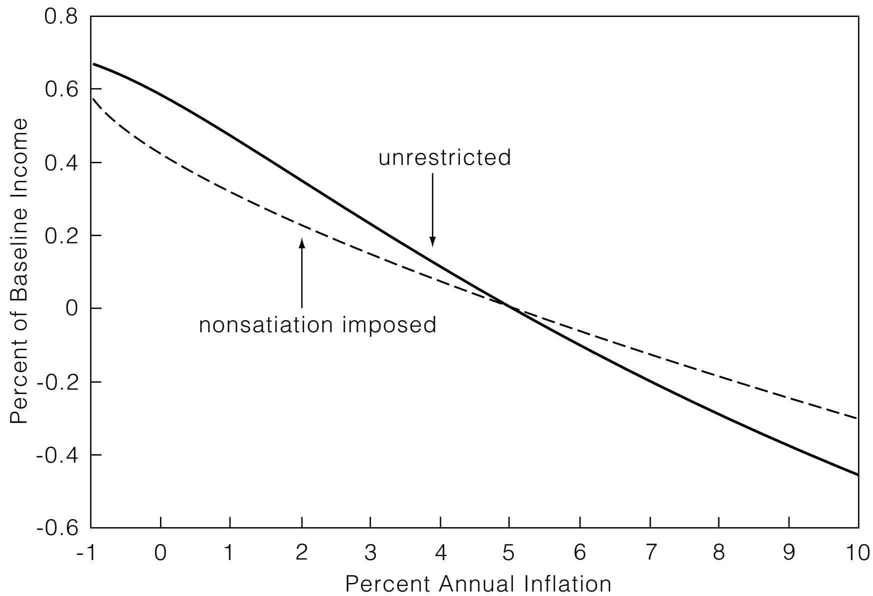
This completes the description of the model economy.

The benchmark economy has 5 percent annual inflation, and the other parameters are chosen as follows. The real growth rate (g) is 3 percent annually. The production function is Cobb-Douglas with labor's share equal to 2/3, and the depreciation rate, δ, is 0.025. The preference parameters ψ and β are set so that the real interest rate is 6.5 percent annually and steady-state hours worked are 20 percent of the time endowment.[14] All of the above values are within the normal range chosen in the real business cycle literature. Given the estimated parameters of the transactions-time technology, the constant of integration (Ω) is chosen so that steady-state transactions time is 2 percent of the time endowment, consistent with the data presented by Andreyenkov, Patrushev, and Robinson (1989). The values of the parameters ψ, β, and Ω, as well as the remaining endogenous variables, are found by solving for a deterministic steady state of the system of equations given by (12)–(18), (4), (7), and (9). To compute the welfare measure, the inflation rate alone is varied, and the new steady state is computed at each desired inflation rate.

Figure 3 plots the quantity of full income (defined in footnote 13), as a percentage of its benchmark level, that individuals would be willing to forego (would require) to live in a lower (higher) inflation world. The solid line represents the unrestricted estimated money demand specification, and the dashed line represents the restricted estimates that impose nonsatiation. Both specifications imply that if inflation were reduced to the Friedman rule from a 5 percent annual rate, for individuals in the model economy it would be as if their full

[14] The real interest rate, r, is equal to $(1 + R)/(1 + \pi)$, where π is the inflation rate. As in King and Wolman (1996), I assume that the risk-free real interest rate relevant for calculating the opportunity cost of money holding is 1 percent annually, whereas the real rate that implicitly enters (15) and (18) is 6.5 percent. The risk premium is not modeled explicitly. In practice this means that there are two real interest rates in the model, but since both of them are "known," no equations or unknowns are added to the steady-state computation. See below for a discussion of the implications of this ad hoc approach.

Figure 3 Welfare Compared to Baseline of 5 Percent Inflation



income had increased by about 0.6 percent.[15] However, the apportionment of these benefits differs in the two cases. Under nonsatiation, less than 3/4 of this benefit can be achieved with zero inflation, whereas the unrestricted money demand specification implies that almost 9/10 of the benefit can be achieved with zero inflation. While I have argued that nonsatiation is an implausible assumption, one should keep in mind that the satiation level was imprecisely estimated. To the extent that one believes the actual satiation level is higher than it was estimated in Section 3, there would be higher costs associated with zero inflation relative to the Friedman rule than are indicated by the solid line.

While the results are not sensitive to small changes in most of the model's other parameters, they are sensitive to the underlying real interest rate. I have assumed that the real return on capital is 6.5 percent annually and that the risk-free real rate of return relevant for measuring the opportunity cost of holding money is 1 percent. Since uncertainty is not explicitly incorporated into the model, this assumption is ad hoc. The assumption is made because in the United States these have been the average real returns on equity and Treasury bills, respectively. Ideally, one would explicitly model the banking system and

[15] The magnitude of these welfare benefits is similar to the magnitudes reported in Lucas (1994).

thus endogenize the spread between risky and (nominally) riskless returns. The assumption of a 1 percent riskless real rate is important for the magnitude of the welfare cost of inflation. If that rate were instead assumed to be 6.5 percent, then the Friedman rule would involve a 6.5 percent deflation rate instead of a 1 percent deflation rate. From Figure 3, this would imply a significantly larger benefit to achieving the Friedman rule. However, the behavior of money demand at the Friedman rule—that is, whether or not there is satiation—would remain important regardless of the assumed value for the real rate.

There is a long tradition, discussed earlier, of measuring the cost of inflation by the area under a money demand curve. Here that calculation would have yielded curves almost identical in shape to those in Figure 3. However, the area calculation would describe time saving only, without accounting for the effect on welfare of changes in consumption. In the case of the estimates with satiation, there is roughly a 1 percent difference in the level of consumption between the Friedman rule and 5 percent inflation. I take this difference in consumption into account in Figure 3. In general, the area under the money demand curve may misstate the welfare cost of inflation because it measures only the direct effect of increases in real balances; here the direct effect is the decrease in transactions time and the indirect effects are summarized by the increase in consumption. This distinction is especially important in Dotsey and Ireland (1996), where the (endogenous) growth rate of the economy is indirectly affected by the inflation rate. The direct effect on money demand is dwarfed by the indirect effect on growth in their model. An additional reason for preferring exact welfare calculations is that the area under the money demand curve does not take into account agents' preferences and thus cannot actually be interpreted in terms of welfare.

5. OTHER EFFECTS OF INFLATION

The above analysis compares different rates of steady inflation in a model where the only welfare effects of inflation work through the demand for money. This narrow focus was chosen to highlight the importance of assumptions about the behavior of money demand at low nominal interest rates. However, in more general models, the quantitative results involving money demand may vary. Furthermore, there may be welfare effects of inflation unrelated to the demand for money. In this section, I briefly discuss some ways in which analysis of the welfare effects of inflation differs in more general models. The references I provide are meant to serve as entry points to what in each case are extensive literatures.

Much of the literature on macroeconomic models with money has involved nominal rigidities, such as sticky prices. In contrast, the model in this article has flexible prices. Sticky prices lead to effects of steady inflation that work through other channels in addition to money demand. Models with sticky prices usually involve imperfect competition, and inflation can affect the magnitude of the distortion from imperfect competition. In King and Wolman (1996), for example, the markup of price over marginal cost—which is a distortion—varies with inflation because firms incorporate into their pricing decisions the possibility that the price they choose will remain fixed for several periods. While some have suggested that inflation can have beneficial effects on the markup (Rotemberg 1996; Benabou 1992), King and Wolman (1996) find the opposite effect, as firms choose a high markup when they set price to compensate for the deterioration that will be caused by inflation. Whether that result generalizes to a wider class of models is an open question.

A literature beginning with Phelps (1973) extends the type of analysis performed in this article by incorporating distortionary taxes. Inflation, or more properly, money creation, is a source of revenue (seigniorage) for the government. Implicitly, my analysis has assumed that this revenue can be replaced by a lump sum tax, which does not distort individual decisions. If lump sum taxes are unavailable, so that seigniorage must be replaced by a distortionary tax such as an income tax, then the optimal rate of inflation in principle could be higher than that corresponding to the Friedman rule; there would be a welfare benefit to inflation counteracting the welfare cost associated with money demand. Recent work by Chari, Christiano, and Kehoe (1996) and Correia and Teles (1997), among others, suggests that this benefit is small enough that the Friedman rule remains optimal with distortionary taxes for a wide range of money demand specifications. With satiation, however, distortionary taxes would probably make the optimal nominal interest rate positive, because with satiation the marginal welfare cost of inflation is zero at the Friedman rule.

Feldstein (1997) has emphasized another way in which inflation interacts with public finance, namely the costs of inflation that result from a nonindexed tax code. With a nonindexed tax code, inflation raises the effective tax rate on both individuals and businesses. Feldstein argues that these tax-related distortions alone cost the U.S. economy about 0.8 percent of GDP per year.

Finally, the steady-state analysis in this article leaves open the question of transitional effects of a significant decrease in the inflation rate. These transitional effects would be small in the model used here. However, models with sticky prices or other nominal rigidities may imply significant welfare costs of a transition to lower inflation, with the costs depending on such factors as how credible the disinflation is. Friedman himself stressed transitional issues: "Any decided change in the trend of prices would involve significant frictional distortion in employment and production" (1969, p. 45). This topic is currently being studied intensively; see Ball (1994a,b) and Ireland (1995) for examples of recent work. It is important to note, however, that in contrast to a one-time cost of lowering the inflation rate, the benefit of low inflation emphasized in this article accrues year after year.

6. CONCLUSIONS

At positive nominal interest rates, individuals incur an opportunity cost by holding money instead of interest-bearing securities. Since the social cost of producing money is nearly zero, there is a divergence between the private and social costs of holding money when nominal interest rates are positive. Individuals choose to equate the marginal benefit of holding money with the private cost, so positive nominal interest rates generate an inefficiency. Policymakers, by setting the nominal interest rate at zero, and so equating private and social costs, can eliminate this inefficiency. In models where there are no other distortions, it follows that this same monetary policy is optimal from a welfare perspective. Lucas (1994) has argued that the form of the money demand function implies significant welfare losses at even very low nominal interest rates. His conclusion results from his assumption that individuals do not become satiated with real balances as the nominal rate declines toward zero. Equivalently, the marginal benefit of holding real balances is positive no matter how high are individuals' money holdings.

I have estimated the money demand function implied by a general transactions-time technology and found evidence that the marginal benefit of holding real balances declines to zero at a nominal interest rate of zero. In other words, individuals can become satiated with real balances. My conclusions regarding satiation, however, are vulnerable to the criticism that zero nominal interest rates have never occurred. Nonetheless, my results imply that the welfare cost of low nominal interest rates is small. Most of the benefits from reducing inflation below, say, 5 percent can be achieved with price stability (zero inflation), and those benefits are significant. In my model a reduction in inflation from 5 percent to zero is equivalent to an increase in consumption of 0.6 percent of output. This result helps reconcile the optimality of zero nominal interest rates with the tendency of central banks to emphasize zero inflation. Still, one wonders why central banks do not simply advocate the optimal policy. Probably the explanation involves factors such as transitional costs of disinflation. This point aside, it is easier to understand why central banks would advocate sub-optimal policy if that policy is close to being optimal.

APPENDIX

The data used to estimate (9) are annual, from 1915 to 1992. The nominal interest rate is the yield on commercial paper from the National Bureau of Economic Research (NBER) database (1915–1946) and Citibase (1947–1992).

I use nominal data for consumption, the wage rate, and the money supply; taking ratios causes the price indexes to cancel. The consumption series consists of three spliced series. From 1915 to 1929, I combine personal consumption expenditures per capita in 1929 dollars, with the deflator for the same. The former is series A25 from Kendrick, reproduced in the U.S. Commerce Department's *Long-Term Economic Growth* (*LTEG*). The latter is series B64 from *LTEG*. Both are annual series. From 1930 to 1945, I combine personal consumption expenditures per capita in 1958 dollars, with the deflator for the same. The former is series A26 and the latter is series B65, both from *LTEG*, and both annual. Finally, from 1946 to 1992, I use personal consumption expenditures in current dollars, divided by population. The former is series GC, from Citibase; it is in billions of dollars and is seasonally adjusted quarterly data, which I average to create annual data. The latter is PAN (Citibase 1946–1991), with data for 1992 estimated by extrapolating the average rates of change from 1990 to 1991; population is in thousands.

As mentioned above, I use nominal wage data. Also, since the raw wage data is hourly, I multiply by the number of hours in a quarter (2,184) to get a quarterly wage. From 1915 to 1946, I "reflate" total compensation per hour at work for manufacturing production workers, using the CPI. The former is series B70 from *LTEG*; it is in 1957 dollars. The latter is m04045 from the NBER database. From 1947 to 1992, I use average hourly earnings of production workers in manufacturing, in current dollars. This is series LEHM from Citibase. Finally, since the relevant wage variables from a theoretical perspective are after-tax wages, I multiply wages by the average marginal tax rates provided by Barro and Sahasakul (1983) and updated through 1992 in the manner they describe.[16]

For money, from 1915 to 1970 I use the M1 series from Friedman and Schwartz (1963) and the Federal Reserve, which is reproduced as series B109 and B110 in *LTEG*. From 1970 to 1992 I use FM1 from Citibase. Both series are in billions of dollars and are deflated by the POPM population measure mentioned above. Prior to 1946, that population measure is the annual series in the Bureau of the Census's *Historical Statistics* (Series A–6–8, p. 8).

[16] The conclusions reached above are unchanged if before-tax wage rates are used.

REFERENCES

Amemiya, Takeshi. *Advanced Econometrics.* Cambridge, Mass.: Harvard University Press, 1985.

Andreyenkov, Vladimir G., Vasily D. Patrushev, and John P. Robinson. *Rhythm of Everyday Life: How Soviet and American Citizens use Time.* Boulder, Colo.: Westview Press, 1989.

Bailey, Martin J. “The Welfare Cost of Inflationary Finance,” *Journal of Political Economy*, vol. 64 (April 1956), pp. 93–110.

Ball, Laurence. “Credible Disinflation with Staggered Price-Setting,” *American Economic Review*, vol. 84 (March 1994a), pp. 282–89.

__________. “What Determines the Sacrifice Ratio,” in N. Gregory Mankiw, ed., *Monetary Policy.* Chicago: The University of Chicago Press, 1994b.

Barro, Robert J., and Chaipat Sahasakul. “Measuring the Average Marginal Tax Rate from the Individual Income Tax,” *Journal of Business*, vol. 56 (October 1983), pp. 419–52.

Baumol, William J. “The Transactions Demand for Cash: An Inventory-Theoretic Approach,” *Quarterly Journal of Economics*, vol. 66 (November 1952), pp. 545–66.

Benabou, Roland. “Inflation and Markups: Theories and Evidence from the Retail Trade Sector,” *European Economic Review*, vol. 36 (April 1992), pp. 566–74.

Bureau of the Census. *Historical Statistics of the United States, Colonial Times to 1970*. Washington: U.S. Department of Commerce, 1975.

Chadha, Jagjit S., Andrew G. Haldane, and Norbert G. J. Janssen. “Shoe-Leather Costs Reconsidered.” Manuscript. Bank of England, January 1997.

Chari, V. V., Lawrence J. Christiano, and Patrick J. Kehoe. “Optimality of the Friedman Rule in Economies with Distorting Taxes,” *Journal of Monetary Economics*, vol. 37 (April 1996), pp. 203–23.

Correia, Isabel, and Pedro Teles. “The Optimal Inflation Tax,” Discussion Paper 123. Institute for Empirical Macroeconomics, Federal Reserve Bank of Minneapolis, August 1997.

Dotsey, Michael, and Peter N. Ireland. “The Welfare Cost of Inflation in General Equilibrium,” *Journal of Monetary Economics*, vol. 37 (February 1996), pp. 29–47.

Federal Reserve Bank of Kansas City. *Achieving Price Stability.* Kansas City: Federal Reserve Bank of Kansas City, 1996.

Feldstein, Martin. "The Costs and Benefits of Going from Low Inflation to Price Stability," in Christina D. Romer and David H. Romer, eds., *Monetary Policy*. Chicago: University of Chicago Press, 1997.

Friedman, Milton. "The Optimum Quantity of Money," in *The Optimum Quantity of Money, and Other Essays.* Chicago: Aldine Publishing Company, 1969.

__________, and Anna J. Schwartz. *A Monetary History of the United States 1867–1960.* Princeton: Princeton University Press, 1963.

Ireland, Peter N. "Optimal Disinflationary Paths," *Journal of Economic Dynamics and Control,* vol. 19 (November 1995), pp. 1429–48.

King, Robert G., Charles I. Plosser, and Sergio T. Rebelo. "Production, Growth and Business Cycles: I. The Basic Neoclassical Model," *Journal of Monetary Economics,* vol. 21 (March/May 1988), pp. 195–232.

King, Robert G., and Alexander L. Wolman. "Inflation Targeting in a St. Louis Model of the 21st Century," Federal Reserve Bank of St. Louis *Review,* vol. 78 (May/June 1996), pp. 83–107.

Kiyotaki, Nobuhiro, and Randall Wright. "On Money as a Medium of Exchange," *Journal of Political Economy,* vol. 97 (August 1989), pp. 927–54.

Lacker, Jeffrey M. "Stored Value Cards: Costly Private Substitutes for Government Currency," Federal Reserve Bank of Richmond *Economic Quarterly*, vol. 82 (Summer 1996), pp. 1–25.

Lucas, Robert E., Jr. "On the Welfare Cost of Inflation," Working Paper 394. Stanford University: Center for Economic Policy Research, 1994.

__________, and Nancy L. Stokey. "Optimal Fiscal and Monetary Policy in an Economy without Capital," *Journal of Monetary Economics*, vol. 12 (July 1983), pp. 55–93.

McCallum, Bennett T. "The Role of Overlapping-Generations Models in Monetary Economics," *Carnegie-Rochester Conference Series on Public Policy,* vol. 18 (Spring 1983), pp. 9–44.

__________, and Marvin S. Goodfriend. "Demand for Money: Theoretical Studies," in *The New Palgrave: A Dictionary of Economics*. London: Macmillan Press, 1987, reprinted in Federal Reserve Bank of Richmond *Economic Review*, vol. 74 (January/February 1988), pp. 16–24.

Mulligan, Casey B., and Xavier X. Sala-i-Martin. "Adoption of Financial Technologies: Implications for Money Demand and Monetary Policy," *NBER* Working Paper 5504. March 1996.

Phelps, Edmund S. "Inflation in the Theory of Public Finance," *Swedish Journal of Economics*, vol. 75 (March 1973), pp. 67–82.

Prescott, Edward C. “Theory Ahead of Business Cycle Measurement,” Federal Reserve Bank of Minneapolis *Quarterly Review,* vol. 10 (Fall 1986), pp. 9–22.

Rotemberg, Julio J. “Commentary,” Federal Reserve Bank of St. Louis *Review*, vol. 78 (May/June 1996), pp. 108–11.

Sidrauski, Miguel. “Rational Choice and Patterns of Growth in a Monetary Economy,” *American Economic Review*, vol. 57 (May 1967), pp. 534 – 44.

Tobin, James. “The Interest-Elasticity of Transactions Demand for Cash,” *Review of Economics and Statistics*, vol. 38 (August 1956), pp. 241– 47.

U.S. Department of Commerce. *Long-Term Economic Growth*, *1860–1970.* Washington: Bureau of Economic Analysis, 1973.