Working Paper Series

A Framework for the Analysis of Moderate Inflations

WP 97-04

Marvin Goodfriend
Federal Reserve Bank of Richmond

THE FEDERAL RESERVE BANK OF RICHMOND
RICHMOND ▪ BALTIMORE ▪ CHARLOTTE

Working Paper 97-4

A Framework for the Analysis of Moderate Inflations

Marvin Goodfriend*

Federal Reserve Bank of Richmond, P.O. Box 27622, Richmond, VA 23261
(First Version: March 1995)

Abstract

Optimal monetary policy is studied in a model with no contractual restrictions or physical costs of changing prices. Nevertheless, the price level is sticky in a range of markup indeterminacy, and inflation occurs only when employment presses against capacity. Under full information, the monetary authority can exploit price level stickiness to minimize the markup and keep employment at a constrained optimum without inflation. Under uncertainty, negative aggregate demand shocks produce real contractions and positive shocks raise the price level. The monetary authority can raise the likelihood that aggregate demand will maximize employment, but at the cost of higher expected inflation.

Key Words: Optimal Monetary Policy, Inflation, Unemployment

JEL classification: E3, E4, E5

---

*The paper was prepared for the March 1995 Swiss National Bank Conference on "Rules versus Discretion in Monetary Policy." It is forthcoming in the Journal of Monetary Economics. The paper has benefitted from presentations at the Bank of England, the Federal Reserve Board, the Federal Reserve Research Committee on Macroeconomics, the Norwegian School of Management, the University of Helsinki, the University of South Carolina, the University of Vienna, and the Wharton School at the University of Pennsylvania. I would like to thank Philippe Bacchetta, Mark Bils, Clive Briault, Satyajit Chatterjee, Neil Ericsson, Robert King, Jeff Lacker, Bennett McCallum, John McDermott, Tom Sargent, and Alan Stockman for valuable comments. The views expressed here are mine alone and do not necessarily reflect those of the Federal Reserve Bank of Richmond or the Federal Reserve System.

1. Introduction

This paper proposes a new framework within which to understand, interpret, and evaluate monetary policy that is consistent with some common experiences and perceptions of central bankers, and can potentially serve as a conceptual basis for policy analysis. The model embodies a simple capacity constraint that relates inflation to excessive levels of employment, and it presents the policymaker with a Phillips curve tradeoff between unemployment and inflation.

The model is neoclassical in spirit in that firms never face contractual restrictions or physical costs of changing prices. Yet the model is Keynesian in the sense that the price level is sticky for some states of the economy and configuration of shocks. Under perfect information about states and shocks, the monetary authority has the power to keep the economy at a constrained optimum. However, in contrast to the standard Keynesian model there is no benefit to inflation under full information. The Phillips curve tradeoff arises only as a consequence of the fact that the monetary authority is imperfectly informed about the current state of the economy.

At the heart of the model is a market structure in which goods are produced by a large number of monopolistically competitive firms facing demand curves that are kinked because customers are imperfectly informed about individual firm price changes. The resulting discontinuity in marginal revenue curves means that there is a range of markup indeterminacy and a corresponding range in which firms are satisfied to keep their prices constant in response to shifts in marginal revenue and marginal cost.

Two pricing practices are assumed to accompany this market structure. Together, these constitute an equilibrium selection device for the economy. Pricing Practice 1 says that the price level adjusts if and only if the representative firm wishes to change its relative price to bring its markup into an acceptable range. This first practice is a natural one if firms care only about relative prices. Pricing Practice 2 says that firms change prices by the minimum amount necessary to bring the markup into an acceptable range. This second practice selects a natural focal point for firms to coordinate on when changing prices.

The market structure, together with the pricing practices, implies three pricing regimes. Prices are sustained at the previous period's level if the "incipient markup"—the markup calculated for current variables at last period's prices—lies in the range of markup indeterminacy. When that is the case, the incipient markup is sustained in equilibrium, and changes in the money stock exert strong effects on employment. On the other hand, when the incipient markup lies below the minimum, firms raise prices until that minimum is attained. Likewise, firms lower their prices whenever the incipient markup lies above the acceptable maximum.

Equilibrium employment and output are inversely related to the markup in this model because the markup acts like a tax that drives a wedge between the price of output and its marginal cost of production. The model implies that there can be inflation only if employment presses against a "capacity constraint", and deflation only if there is a sufficiently large "output gap."

The monetary authority maximizes the utility of the representative agent by attempting to minimize the two distortions in the model: the markup and the nominal interest rate. A suitably deflationary money growth rate can sustain a zero nominal interest rate. But the monetary authority must maximize the markup distortion in order to eliminate the nominal interest rate distortion.

A monetary authority mainly interested in maximizing employment and output would minimize the markup. Under full information, it could sustain that minimum exactly, without inflation, by exploiting the sticky price level within the range of markup indeterminacy. Since inflation yields no benefits under full information, the monetary authority would also maintain price stability in order to support the lowest nominal interest rate consistent with maximum employment.

A Phillips curve tradeoff emerges only as a result of the fact that the monetary authority is imperfectly informed and cannot manage the markup exactly. The tradeoff emerges because firms react asymmetrically to the incipient markup at the boundary of the range of indeterminacy. If the incipient markup comes in below the acceptable minimum, firms raise prices; if it comes in above, firms do not change prices, the incipient markup is sustained, and employment and output fall. Hence, uncertainty creates both inflation and unemployment risk for a monetary authority targeting maximum employment and output.

The monetary authority can influence the balance of risks by its choice of money stock relative to the previous period's price level. Choosing a higher money stock raises the probability that the incipient markup will lie below the minimum acceptable to firms. That raises expected inflation—but it also increases the probability that the actual markup will be minimized, and employment and output maximized. Choosing a lower money stock reduces expected inflation, but raises expected unemployment.

The paper proceeds as follows. In Section 2, the basic macroeconomic model is presented for an exogenously given markup in order to illustrate core model mechanics. We focus on the effect of the markup and the nominal interest rate distortions on equilibrium work effort and transaction time allocations because the two distortions are at the heart of the analysis of monetary policy in Section 4. Before leaving Section 2, we show that the steady state welfare cost of inflation as a percent of GDP is proportional to the markup. We also compare the welfare cost of inflation in the model to earlier estimates.

Section 3 introduces the monopolistically competitive market structure and the pricing practices that serve as the equilibrium selection device. It then describes the implied pricing regimes that underlie the endogenous behavior of the markup and employment. It explains the dependence of the markup on aggregate demand, and shows how the behavior of the markup governs inflation.

Optimal monetary policy is characterized in Section 4. Section 4.1 fixes ideas by discussing optimal policy in a nonstochastic steady state. There are two contenders for welfare maximizing policy in this case: a deflationary Friedman Rule and a Zero Inflation Rule. We compare them using the welfare cost of inflation formula presented in Section 2.3. In Section 4.2, we characterize the optimal rule when the monetary authority is incompletely informed. We show how the introduction of uncertainty presents the monetary authority with a Phillips curve tradeoff that causes it to depart from zero inflation. Section 4.3 argues that the optimal rule is sustainable in the model even in the absence of a commitment technology.

Section 5 briefly addresses some aspects of the model in more detail, and suggests some desirable extensions. A conclusion follows.

2. The Core Real Business Cycle Model

The core of the framework is a monopolistically competitive real business cycle model in which capital is not present. The model is one in which employment and output vary because of fluctuations in the markup of the price of output over its marginal cost of production. The core macromodel closely parallel's the standard setup of Blanchard and Kiyotaki (1987). A real business cycle interpretation is adopted here because it simplifies the discussion of endogenous markup variation later on. We abstract from capital because it is not central to the inflation and unemployment issues that are the focus of the paper.

The real business cycle model is specified as follows. Representative agent utility depends on consumption (C) and leisure (L):

$$(1) \qquad \sum_{t=0}^{\infty}(1+\rho)^{-t}[(1-\phi)\log C_t + \phi \log L_t]$$

Money is assumed to buy goods according to a transations technology that relates holdings of real money balances (M/P) and the fraction of "shopping time" devoted to transacting (S) to the spending flow that the representative agent carries out:

(2) $$C_t = k\ (M/P)_t\ S_t,$$

where k is a constant. This constraint is a specialization of the McCallum-Goodfriend (1987) shopping-time technology proposed and utilized in Lucas (1993, 1994). We assume that money may be acquired at the beginning of the period in which it yields transaction services. However, money balances that yield transaction services in period t must be carried into period t+1.

The production function for nonstorable output is:

(3) $$C_t = X_t\ N_t^{\alpha}$$

where N_t is hours worked (employment), X_t is a productivity coefficient, and $0<\alpha<1$. There is also a time constraint:

(4) $$1 = L_t + N_t + S_t$$

Agents begin period t with M_{t-1} units of money carried over from t-1, which is augmented by a lump sum transfer H_t. They earn a real wage (W/P) per hour of work effort supplied to firms, and they own firms and receive all profits, i.e., the excess of revenue over wage payments. There is also a one-period nominal bond that is purchased at price $1/(1+R_{t-1})$ in period t-1 and is redeemed for one unit of money in period t. The number of bonds purchased in t-1 and redeemed in t is B_{t-1}. So the representative agent's resource constraint in real terms is:

$$(5) \qquad M_{t-1}/P_t + H_t/P_t + B_{t-1}/P_t + (W/P)_t N_t + X_t\,\hat{N}_t^{\alpha} - (W/P)_t\hat{N}_t - C_t$$

$$= (M/P)_t + (B/P)_t/(1+R_t)$$

where R_t is the net interest rate on a one-period nominal bond carried from period t into period t+1, and $\hat{N}_t$ is the average economy-wide fraction of time worked.[1]

The representative agent maximizes (1) subject to (2), (4), and (5). Forming the Lagrangian with (5), and using (2) and (4) to substitute out for C_t and L_t, the FOCs with respect to S_t, N_t, and M_t imply:

$$(6) \qquad R_t\,(M/P)_t = (W/P)_t\,S_t$$

(7) $\quad 1 - N_t - S_t = [\phi/(1-\phi)][C_t/(W/P)_t][1 + ((W/P)_t/k\ (M/P)_t)]$

Condition (6) equates the marginal opportunity cost of transaction services from another dollar to that from another minute of transaction time.[2] Condition (7) equates the marginal utility of leisure to the marginal utility of work effort, net of transaction cost.

Define the (gross) markup, μ, as the ratio of price to marginal cost. Take the markup as exogenous for now. The markup will be endogenized when we specify the market structure and the pricing practices in Section 3.

Expressed directly as price over marginal cost using (3), the markup is:

$$\mu_t = P_t/[W_t/\alpha X_t N_t^{\alpha-1}] \tag{8}$$

Rewriting (8), we can express the real wage in terms of the markup:

$$(W/P)_t = \alpha X_t N_t^{\alpha-1}/\mu_t \tag{9}$$

Expressions (8) and (9) make clear that when the markup pushes price above marginal cost, it also pushes the real wage below the marginal product of labor.

Solve for equilibrium employment (N) and transaction time (S) allocations as follows. First, use (3) and (9) to yield:

$$(W/P)_t/C_t = \alpha/\mu_t N_t \tag{10}$$

Next, substitute (2) and (10) into (7) and derive:

$$S_t = 1 - \phi - [\phi((\mu_t/\alpha) \ - 1) + 1]N_t \tag{11}$$

where $\mu/\alpha > 1$. Finally, use (2), (6), and (10) to arrive at:

$$S_t^2 = [R_t/\alpha k] \ \mu_t N_t \tag{12}$$

Equations (11) and (12) can be solved for N_t and S_t as functions of μ_t and R_t. These dependencies are summarized as follows:[3]

$$N_t = N(\underset{-}{\mu_t}, \underset{-}{R_t}) \tag{13}$$

(14) $$S_t = S(\underset{+}{\mu_t}, \underset{+}{R_t})$$

The equilibrium demand for money is derived from (2) and (6) as:

(15) $$(M/P)_t = \sqrt{(W/P)_t C_t / kR_t}$$

Apart from the usual dependence on the transaction scale variable C_t and the nominal interest rate, money demand is also positively related to the real wage. The reason is that the real wage measures the consumption opportunity cost of shopping time, so a higher real wage induces agents to substitute real money balances for shopping time in the transaction technology.

In what follows it is useful to write money demand by using (10) to eliminate $(W/P)_t$ in (15):

(16) $$(M/P)_t = C_t \sqrt{\alpha / k\mu_t N_t R_t}$$

2.1. The Mechanics of the Markup

The effect of the markup on equilibrium employment is straightforward. When the markup is unity, price equals marginal cost, the real wage equals the marginal product of labor, and equilibrium employment is such that the private marginal rate of substitution of consumption for leisure equals the marginal product of labor.

A markup in excess of unity pushes the real wage below the marginal product of labor, forces the private marginal rate of substitution of consumption for leisure below the marginal product of labor, and reduces equilibrium employment. The elasticity of employment with respect to the markup is greater in absolute value the smaller the diminishing returns to labor and the larger the elasticity of labor supply with respect to the real wage.

A markup in excess of unity can be interpreted either as a tax imposed by firms on work effort or one imposed on output, whose proceeds firms distribute as profits. In effect, the imposition of a markup is a tax and transfer fiscal policy administered by firms—one that has distortionary substitution effects but no wealth effects. Equilibrium employment and consumption vary inversely with the markup because there is only the negative substitution effect.

2.2. Steady State Growth

The model has a steady-state growth path in which time allocations are constant and consumption grows at an exogenous trend rate of productivity growth, g. The nominal interest rate in the steady state is therefore given approximately by:

$$\text{(17)} \qquad \mathrm{R} = \rho + g + E \log\left(\mathrm{P}_{t+1}/P_t\right)$$

where the last term denotes steady state expected inflation. By money demand function (16), money growth governs steady-state inflation, and the level of the nominal interest rate. Expressions (13) and (14) determine steady state time allocations for an exogenously given markup μ and the nominal rate R consistent with steady state money growth. Production function (3) and the productivity growth trend then determine the consumption growth path; the money demand function and the money supply path determine a path for the price level. Although steady state output growth is given exogenously by g, the level of the path for output varies with μ and R through their effects on employment according to (13).

2.3. The Steady State Welfare Cost of Inflation

This section derives a formula for the welfare cost of inflation and shows the cost to be proportional to the markup. By (12), S is time wasted from the social point of view due to the fact that the nominal interest rate is not zero. The fraction of time, S, spent economizing on money holding is, therefore, a direct measure of the welfare cost of inflation. To measure the welfare cost as a percentage of GDP, value S at the social opportunity cost of shopping time—the marginal product of labor—and divide by GDP. Using (9) express the steady state welfare cost as a percentage of GDP: 100 x S (W/P) μ/C; and using (10), write it as 100 x $\alpha S/N$. Substituting for S with (12) yields the welfare cost formula:

$$\Psi = 100 \text{ x } (\alpha/N)\ \sqrt{R\mu N/\alpha k} \tag{18}$$

In order to evaluate Ψ, we need to calibrate the constant k. To do so use (6) and (10) and eliminate S in (12) to express k as:

$$k = (\alpha/\mu N)/[R\ [(M/P)/C]^2]. \tag{19}$$

To facilitate the comparison below to the welfare cost formula presented in Lucas

(1993), calibrate k using his 1990 observations for the inverse M1 velocity and the short-term interest rate in the U.S., 0.15 and 0.075, respectively. Doing so yields $k = 593\ (\alpha/\mu N)$.

Substituting for k in (18) yields the welfare cost as a percentage of GDP expressed as a function of R and μ:

$$\Psi(\mathrm{R},\, \mu) = 100 \text{ x } \mu(0.041)\sqrt{R} \tag{20}$$

Welfare cost function (20) is exactly Lucas's (1993) welfare cost of inflation formula for transaction technology (2) when the markup is unity.

To see why the welfare cost varies directly with the markup, recall that $\Psi =$ 100 x $\alpha S/N$. As discussed above, a higher markup reduces employment because it amounts to a tax on work effort. A higher markup raises S as follows. First, a higher markup lowers the real wage. By (6), the lower opportunity cost of shopping time causes agents to substitute money for time in transactions. Since the percentage decline in C is less than the percentage decline in W/P, S must rise to satisfy transactions constraint (2).

Evidence regarding the size of the average markup in the U.S. economy is reported, for instance, in Hall (1988), and Basu and Fernald (1994). Although

Hall reports markups ranging from 1.8 to 3.8 for the seven one-digit industries he considers, modifications of his work by Basu and Fernald and others show much smaller markups. If the economy were characterized by markups at the low end of Hall's estimates, the welfare cost of inflation would be nearly twice as high as if the markup were unity. More reasonable estimates ranging, say, below 1.5 still raise the cost of inflation significantly. In any case, economists who believe in large markups should recognize that their view multiplies the welfare cost of inflation accordingly.

To get an idea of the welfare gain to bringing inflation down from 4 percent per year to zero, consider that according to Ibbotson (1994) the average inflation adjusted yield on Treasury bills between 1926 and 1993 (excluding the '42 and '47 wartime peg) was 0.8 percent per year. Formula (20), thus, roughly yields $\Psi(0.05, \mu) = 0.9\mu$ at 4 percent inflation and $\Psi(0.01, \mu) = 0.4\mu$ at zero inflation. Even at a markup of unity, the welfare gain to eliminating this moderate inflation is half a percent of GDP per year. At a markup of 2, the gain is a full percent of GDP per year. The gain to pursuing deflation sufficient to bring the nominal interest rate to zero is another 0.4 percent of GDP if the markup is unity, and 0.8 percent of GDP if the markup is 2. The estimates are overstated somewhat to the extent that deposits pay interest.

At any rate, these gains are very large compared to those implied by Bailey's (1956) welfare cost formula. Technically, the reason is that the welfare cost rises with the square root of the nominal interest rate in (20), but with the square of the nominal interest rate according to Bailey. The difference arises because a shopping-time technology like (2) implies that real balances increase without bound as the nominal interest rate approaches zero. Lucas (1994) defends this implication by pointing out that managing an inventory always takes some time so that a larger average stock must always reduce the time requirement. Moreover, he argues that the log-log money demand function implied by the shopping-time technology fits the U.S. data at low interest rates much better than does the semi-log form implicit in Bailey's work. At any rate, the point here is that a markup in excess of unity has the potential to further increase the welfare cost of inflation.

3. Market Structure, Pricing Practices, and the Markup

This section describes the market structure underlying the behavior of the markup. The main purpose is to show how equilibrium variations in the markup are sustained endogenously in the model. Second, it is to illustrate the dependence of the markup on aggregate demand, opening the way for the analysis of monetary policy in Section 4. Third, it is to specify how firm pricing practices depend on the markup and how the behavior of the markup governs inflation.

3.1. Monopolistic Competition and the Range of Markup Indeterminacy

Aggregate output is assumed to consist of a large, fixed number of differentiated products, each produced by a single firm that acts as a monopolistic competitor. The demand for each good depends negatively on its relative product price and positively on aggregate demand. Firms produce output with technology (3).

Each firm maximizes profit by choosing a product price in excess of the marginal cost of production in order to exploit its market power. Hence, firms willingly sell as much as demand will allow at their profit maximizing relative product price P_i/P, where P_i is a firm's nominal product price and P is the price level. We assume symmetry among goods and firms so that profit maximizing relative product prices are all unity in equilibrium. This means that aggregate output can be thought of as a single composite good, and further, that output and employment are proportional to aggregate demand.

We assume, in addition, that the representative firm faces not only a downward sloping demand curve, but a kinked demand. Stiglitz (1987) has shown that a kink results at the common relative product price in a model of sequential search by consumers if there are increasing marginal search costs and a large number of competitors.[4] A firm's demand is relatively elastic above the common relative price because its customers see the price increase immediately and can expect to find a lower price elsewhere. But a firm's demand is relatively inelastic below the common relative price because new customers that the low price would potentially attract are initially unaware of the price reduction. In short, a firm gains fewer customers when it lowers its price than it loses when it raises its price—giving rise to a kink. The existence of a kink at the existing price, in turn, makes that the profit maximizing price.

Stiglitz discusses two cases that support a kink, one with price dispersion and search in equilibrium, and another with no price dispersion and no search. The application in this paper explores the case in which all firms charge a common price and there is no search.

The kink in the demand curve implies a discontinuity in the marginal revenue curve. This means that a firm will not change its relative product price in response to changes in aggregate demand, real wages, or productivity, as long as marginal cost cuts through the gap in the marginal revenue curve. Therefore, the kink creates a range in which the representative firm is indifferent to changes in its markup. Only when marginal cost cuts above or below the gap in the marginal revenue curve does the representative firm react by raising or lowering its product price to bring marginal revenue into equality with marginal cost for profit maximization. In what follows, we assume that the bounds on the range of markup indeterminacy are fixed at $\overline{\mu}$ and $\underline{\mu}$, respectively.

The existence of a range in which the representative firm allows its markup to vary implies a corresponding range of real equilibria for the economy.[5] Subject to the constraint that the markup lies in the range of indeterminacy, firms are always willing to hire more labor to accommodate increased demand because the real wage is below the marginal product of labor. An increase in aggregate employment raises the marginal cost of production by both increasing the wage and lowering the marginal product of labor. Nevertheless, the representative firm accepts the decline in its markup without passing along the cost increase in its product price as long as the markup remains above $\underline{\mu}$.

In effect, this thought experiment runs (13) in reverse—determining the markup as a function of the employment necessary to satisfy aggregate demand. The labor-market-clearing real wage is exactly the one that yields the markup required to support the equilibrium level of output that satisfies aggregate demand.

The upshot is that changes in aggregate demand bring forth accommodating changes in aggregate supply—as long as the markup remains within the range of indeterminacy. This transmission mechanism opens the door for monetary policy to influence employment and output through aggregate demand management.

It is worth pointing out that anecdotal evidence from customer markets is consistent with the view that underlying cost increases are absorbed and decreases are enjoyed by firms within limits without being passed through to product prices. In other words, firms producing differentiated products do appear to allow their markups to vary inversely with their costs over some range before changing their prices.[6]

3.2. Equilibrium Selection Device: The Pricing Practices

This paper explores the implications of one particular equilibrium selection device in the form of two pricing practices that are assumed to accompany the market structure. Firm pricing practices are premised on three features of the model. The first is that firms do not face contractual restrictions or physical costs of changing prices. Second, firms observe all the components of their markup when setting product prices. According to (8), this means that a firm knows the nominal wage (W), which it takes as given in the labor market, as well as its productivity (X), and its desired employment (N), when setting its nominal product price (P_i). Firms also know they are all alike (except for producing differentiated products) and that relative product prices are all unity in equilibrium.

The first pricing practice is a natural one if firms care only about relative prices:

> Pricing Practice 1—The representative firm taking other firms' prices as given changes its product price if and only if its markup lies outside the range of indeterminacy.

The second pricing practice selects a natural focal point for firms to coordinate on when changing prices:

> Pricing Practice 2—Firms change prices the minimum necessary to bring the equilibrium markup into the range of indeterminacy.

3.3. The Pricing Regimes

Pricing practices 1 and 2 imply three pricing regimes. To illustrate these, consider a simple money demand function that determines real aggregate demand:

$$\log \mathrm{C}_t^D = \log \mathrm{M}_t - \log P_t + \log V_t \tag{21}$$

where V_t is velocity. Assume the monetary authority determines M_t. To keep matters simple, (21) ignores the interest sensitivity of money demand.

Let aggregate supply depend on the markup, μ_t, and productivity, X_t:

$$(22) \qquad \log C_t^S = \log X_t - \log \mu_t.$$

We characterize the three pricing regimes in terms of an incipient markup in period t—the markup calculated for period t variables at the period t-1 price level. The incipient markup (μ_t^I) is determined by equating aggregate demand and supply in (21) and (22) and setting $P_t = P_{t-1}$:

$$(23) \qquad \log \mu_t^I = \log X_t + \log P_{t-1} - \log M_t - \log V_t$$

The pricing practices imply that when the incipient markup falls in the range of markup indeterminacy ($\underline{\mu}<\mu_t^I<\overline{\mu}$) the representative firm does not wish to change its product price. In this case, both the previous period's price level and the incipient markup are sustained in equilibrium. The money stock and velocity shocks exert powerful effects on the markup and aggregate supply in this case. This is the Stable Price Level Regime:

$$(24) \qquad \underline{\mu} < \mu_t^I < \overline{\mu}: \ \mu_t = \mu_t^I \text{ and } P_t = P_{t-1}$$

On the other hand, when aggregate demand is strong enough to push the incipient markup below the minimum acceptable to firms ($\mu_t^I < \underline{\mu}$), then firms raise their product prices until the minimum acceptable markup is attained. This is the <u>Inflation Regime</u>:

$$(25) \qquad \mu_t^I < \underline{\mu}: \ \ \mu_t = \underline{\mu} \text{ and } \log \mathrm{P}_t = \log \underline{\mu} + \log M_t + \log V_t - \log X_t$$

Money and velocity affect the price level in the Inflation Regime, but they do not affect aggregate supply.

Finally, if aggregate demand is too weak to push the incipient markup below the maximum acceptable to firms ($\overline{\mu} < \mu_t^I$), then firms cut product prices until $\overline{\mu}$ is attained. This is the <u>Deflation Regime</u>:

$$(26) \qquad \overline{\mu} < \mu_t^I: \ \ \mu_t = \overline{\mu} \text{ and } \log P_t = \log \overline{\mu} + \log M_t + \log V_t - \log X_t.$$

Before proceeding to analyze optimal monetary policy when firms price according to practices 1 and 2, it is worth pointing out that the model is entirely compatible with other pricing practices that could supplement or substitute for the ones assumed above. In fact, firms could coordinate on other pricing practices that would select very different equilibria. Ultimately, one would have to choose from among the various possible pricing practices on the basis of their empirical implications.

One supplemental pricing practice, for instance, could have firms adjust prices proportionately with the money stock. If every firm believed that others would price that way, then each would have an incentive to price that way too, and the supplemental pricing practice would select an equilibrium in which real money balances were invariant to changes in the money stock. The price level could still change according to pricing practices 1 and 2, so real balances could still move around. And changes in money growth and expected inflation could still affect real output through the nominal interest rate in the money demand function. But the equilibrium with the supplemental pricing practice would differ radically from the one without, because now any inflation rate could be consistent with any markup and employment in the range of indeterminacy.

4. Optimal Monetary Policy

In this section we discuss the money supply rule that an optimizing monetary authority would follow in the environment presented in Sections 2 and 3. The monetary authority's policy problem is to choose a rule to maximize representative agent utility (1), subject to transaction technology (2), production function (3), and time constraint (4). The monetary authority must also respect implementation constraints (6) and (7) that reflect representative agent optimization, as well as the pricing practices and regimes that reflect firm profit maximization.

To fix ideas we first discuss optimal policy in a nonstochastic steady state. Next, we discuss optimal policy in a stochastic setting assuming the monetary authority is incompletely informed about the current state of the economy. We do the analysis in both cases assuming that the monetary authority has a technology enabling it to commit to a rule. We close the section, however, by arguing that the optimal money supply rule in this model is sustained by reputational forces, even in the absence of a precommitment mechanism.

4.1. Optimal Policy in a Nonstochastic Steady State

The easiest way to determine the optimal money supply rule in a nonstochastic steady state is to work backward from the welfare maximizing combination of markup and nominal interest rate, since by (13) and (14) these determine the other equilibrium allocations.

If it were possible, the monetary authority would clearly like to set R to zero by deflating prices at $\rho + g$, while minimizing the markup distortion at $\mu = \underline{\mu}$. According to the pricing regimes described in Section 3.3, however, firms cannot be induced to deflate prices unless the markup is at the top of the range of indeterminacy.

This creates the possibility that a rule that minimizes the markup might dominate one that pursues deflation. Since firms never let their markup fall below the minimum $\underline{\mu}$, there are no benefits to inflationary monetary policy when the monetary authority is fully informed. The best alternative to deflation is zero inflation with $\mu = \underline{\mu}$.

Thus, we have two contenders for welfare maximizing policy in the nonstochastic steady state, which we denote as follows:

The Friedman Rule:[7] $\quad$ R $= 0$ and $\mu = \overline{\mu}$

The Zero Inflation Rule: R $= \rho + g$ and $\mu = \underline{\mu}$

We saw in Section 2.3 that the welfare gain as a percentage of GDP of going from zero inflation to the Friedman Rule was 0.4μ percent of GDP. Some tedious algebra yields a formula for the deadweight cost as a percentage of GDP of a one percent increase in the markup in the neighborhood of steady state values for μ and N:

$$\text{(27)} \qquad \Delta \log \text{GDP} = [[\alpha(1-N)+N]/[1+(\mu N/\alpha(1-N))]](\mu-1)\Delta \log \mu$$

According to (27), the deadweight cost of an incremental increase in the markup would be negligible for a steady state markup near unity. But calibrated at $\alpha = 2/3$, N $= 1/3$, and, say, $\mu = 2$, a one percent increase in the markup would yield a welfare loss of about $1/3$ percent of GDP per year. In this case, for example, the Zero Inflation Rule would dominate the Friedman Rule if and only if $\overline{\mu}$ exceeded $\underline{\mu}$ by about three percent. The prevailing view widely shared by central bankers and financial market participants is against the Friedman Rule. That view is potentially justifiable in this model if the range of markup indeterminacy is wide enough and the steady state markup is high enough.

The money supply rule that supports a stable price level and $\mu = \underline{\mu}$ is found by using (13) to eliminate N in (16) to yield:

$$\text{(28)} \qquad (\text{M/P})_t = X_t\sqrt{\alpha[N(\mu,R)]^{2\alpha-1}/k\mu R}$$

In (28) it is easy to see that the nominal money stock must grow at the rate of productivty growth, g, in order to maintain stable prices. With R $= \rho + g$ and P given by history, (28) may be solved for the level of the money growth path that supports $\mu = \underline{\mu}$.

4.2. Optimal Policy When the Monetary Authority Is Incompletely Informed

We characterize optimal monetary policy when the monetary authority is incompletely informed about the current state of the economy due to a one-period data-processing lag. In this case, the monetary authority sets the period t money stock conditional only on lagged information (I_{t-1}) and on the period t *nominal* interest rate (R_t). There are two serially uncorrelated sources of uncertainty, a productivity shock and a velocity shock. Because the monetary authority conditions on only one indicator, R_t, it can no longer manage the markup exactly. The best it can do is target the conditional mean of the markup by its choice of the money stock relative to the previous period's price level.

The idea is to see how the introduction of uncertainty affects the optimal monetary policy relative to a full information steady state in which the monetary authority would follow the Zero Inflation Rule described above. We work with a log-linearized system in the neighborhood of $\mu = \underline{\mu}$ and $R = \rho + g$, first deriving the marginal rate of substitution of R_t for $\underset{t}{E}\left[\log \mu_t \mid I_{t-1}, R_t\right]$, then characterizing the marginal rate of transformation, and finally comparing the two.

Working from the FOCs in Section 2, with some effort we can write the marginal rate of substitution of R_t for $\underset{t}{E} \log \mu_t$ at steady state values for μ, R, L, and N as:

$$\text{(29)} \qquad \text{MRS } (\mathrm{R}_t \text{ for } \underset{t}{E} \log \mu_t) = -(200L\sqrt{R}/\mu)[\alpha L(1+(\alpha(1-N)/\mu N)) - N]$$

To characterize the marginal rate of transformation of R_t for $\underset{t}{E} \log \mu_t$, consider the effect of uncertainty when the monetary authority continues to choose M_t to target the conditional mean of the incipient markup at the minimum acceptable to firms. The introduction of uncertainty spreads the probability mass of the conditional distribution of log μ_t^I above and below log $\underline{\mu}$. Because $\underline{\mu}$ is the boundary of the Stable Price Level Regime and the Inflation Regime, firms respond asymmetrically to the incipient markup above and below $\underline{\mu}$.[8] When μ^I falls below $\underline{\mu}$, firms raise prices until the minimum acceptable markup is established; when μ^I falls above $\underline{\mu}$ (but below $\overline{\mu}$) firms sustain both the previous period's prices and the incipient markup. So uncertainty in the neighborhood of the minimum markup raises both expected inflation, $\underset{t}{E} \log P_t/P_{t-1}$, and the expected markup, $\underset{t}{E} \log \mu_t$.

Formally, we can write:

$$\underset{t}{E}\log(\mu_t/\underline{\mu}) = F[\underset{t}{E}\log(\mu_t^I/\underline{\mu}),\ \underset{t}{Var}\ \mu_t^I] \tag{30}$$

$$\underset{t}{E}\log(P_t/P_{t-1}) = G[\underset{t}{E}\log(\mu_t^I/\underline{\mu}),\ \underset{t}{Var}\ \mu_t^I] \tag{31}$$

where F_1, $F_2 > 0$, $G_1 < 0$, $G_2 > 0$.

A positive $\underset{t}{Var}\ \mu_t^I$ presents the monetary authority with a tradeoff between the expected markup and expected inflation that it can exploit by its choice of $\underset{t}{E}\log(\mu_t^I/\underline{\mu})$. Lowering the mean of the conditional distribution of the incipient markup relative to $\underline{\mu}$, shifts the distribution further into the region where μ_t^I falls below $\underline{\mu}$ and out of the region where μ_t^I exceeds $\underline{\mu}$. Hence, by lowering its target for $\underset{t}{E}\log(\mu_t^I/\underline{\mu})$ the monetary authority lowers the expected markup and raises expected inflation.

The tradeoff between expected inflation and the expected markup implies one between expected inflation and expected unemployment. Expected unemployment is governed by the expected markup approximately according to: $\log N(\underline{\mu}) - \underset{t}{E}\log N_t = \gamma_{N\mu}\underset{t}{E}\log(\mu_t/\underline{\mu})$, where $\gamma_{N\mu} \equiv 2L/(1+L+N)$.

The Phillips curve tradeoff per se is between expected current inflation and the expected current markup or employment. We still need to translate it into one between the nominal interest rate and the expected current markup in order to characterize the marginal rate of transformation between the latter two. This is straightforward once we recognize that exploiting the Phillips curve to tolerate higher current expected inflation in order to reduce current expected unemployment translates into higher expected future inflation, and a higher average nominal interest rate according to (17).

We are finally in a position to compare the marginal rate of transformation (MRT) with the marginal rate of substitution (MRS) of ER for $E \log \mu$ in the neighborhood of the minimum markup and zero inflation. Consider first the MRT. When expected inflation is near zero, very little probability mass of the conditional distribution of the incipient markup falls below $\underline{\mu}$. If the conditional distribution is shaped like a bell with thin tails, the monetary authority can then target a smaller conditional mean of the incipient markup (with a commensurate reduction of the expected equilibrium markup and unemployment) without moving much probability mass of the incipient markup below $\underline{\mu}$, that is, without raising expected inflation and the nominal interest rate much. Thus, the marginal cost in terms of a higher interest rate of reducing unemployment is close to zero in the neighborhood of zero inflation.

On the other hand, by (29) the MRS is bounded away from zero in the neighborhood of zero inflation, which means that it is worth accepting some increase in the nominal rate for an incremental reduction in unemployment. In other words, a monetary authority that is otherwise inclined to maintain price stability under full information will optimally pursue some inflation in order to reduce unemployment somewhat under incomplete information.

With some algebra, the optimal money supply rule in this situation can be shown to support a price level generating process that is difference-stationary in logs with an inflationary trend and serially uncorrelated departures from trend. In particular, both the money stock and the price level optimally exhibit base drift.

4.3. Monetary Policy without Commitment

The world's monetary authorities, including the Federal Reserve, make monetary policy on a discretionary basis, that is, in the absence of a technology or institutional mechanism that commits policy to a rule. Optimal policy without commitment may not coincide with that under a rule. Hence, the relevance of the view of monetary policy advanced in Section 4.2 would be enhanced if it could be supported in a discretionary equilibrium.

In the language of Chari, Kehoe, and Prescott (1990) the outcome without commitment requires that policy actions be sequentially rational. In a sequentially rational equilibrium, the money supply must maximize the monetary authority's objective function at each date, given that private agents behave optimally. Likewise, private agent optimality requires that they forecast future monetary policies that are sequentially rational for the monetary authority. A sequence of monetary policy rules, time allocations, and prices that satisfy these conditions is a time consistent or sustainable equilibrium.

To verify that the optimal money supply rule is sustainable in this model without commitment, first consider the full information case. Assume that the Zero Inflation Rule is optimal. In this case, because there are no contractional or physical restrictions on changing prices and wages, firms would react to surprise money growth by raising their prices in order to keep the markup from falling below $\underline{\mu}$. Wages would rise too, neutralizing any effect on real variables. Since there is no benefit to a positive money growth surprise, and none to a negative surprise, the Zero Inflation Rule in the model is clearly sustainable under full information.

Now consider the incomplete information case. Here the optimal money supply rule dictates that E log μ_t exceeds log $\underline{\mu}$, so that a positive deviation of M from the rule reduces the expected markup distortion. Hence, the monetary authority has an incentive to deviate from the rule under incomplete information. Suppose that $\log \mu_t^I$ is normally distributed within finite upper and lower bounds. In that case, the benefit from a positive deviation would be exhausted once surprise money growth put the upper bound of the $\log \mu_t^I$ distribution at $\log \underline{\mu}$, since that would put the markup at its minimum with certainty. Accompanying the high money growth would be an expected rate of inflation in excess of that associated with the optimal money supply rule.

Without commitment, then, the monetary authority would be tempted to depart from the optimal rule in order to assure that μ would always equal $\underline{\mu}$. However, if the monetary authority were to attempt a money growth surprise, it could expect private agents to react by raising expected inflation immediately.[9] In that case, the monetary authority's deviation would create an immediate social cost associated with a higher nominal interest rate, as agents respond by immediately substituting shopping time for real money balances. A monetary authority that understands that its deviation would trigger an immediate corresponding increase in the nominal interest rate will abide by the optimal rule because, in that case, the choice among deviations merely reproduces the choice among rules.

Thus, as in Barro and Gordon (1983), the potential loss of reputation—or credibility—motivates the monetary authority to follow the rule here. Reputation alone was not able to completely sustain their ideal rule because the punishment from cheating occurred with a lag in their model. Here, however, if the cost of any deviation is borne simultaneously with the benefit, reputation alone is capable of fully enforcing the optimal rule.[10]

5. Discussion

This section addresses some aspects of the model in more detail and suggests some extensions.

5.1. A Smooth Phillips Curve

As it now stands, the model does not generate inflation and unemployment realizations that lie smoothly along a curve such as that presented by Phillips (1958). According to the pricing regimes specified in Section 3.3, the economy is in the Inflation Regime only when the realized markup is at the minimum acceptable to firms ($\underline{\mu}$), that is, only when unemployment is at its minimum. Moreover, the price level is stable as long as unemployment is above its minimum. Thus, the model generates an L-shaped Phillips curve (ignoring the Deflation Regime).

This was not a problem for the policy analysis in Section 4 because the incompletely informed monetary authority had to choose between expectations of inflation and unemployment, and the tradeoff in expectations does exhibit smoothness. In a more realistic model firms' markups could depend on relative and aggregate shocks, so that there would be a dispersion of incipient markups around the economy-wide average.[11] This way, realized aggregate inflation and unemployment that obtain in a given period could vary smoothly with the fraction of firms whose incipient markups fall below $\underline{\mu}$. Needless to say, this elaboration would greatly complicate the model because it would introduce price dispersion and search in equilibrium.

5.2. Inflation Scares[12]

A model such as this that would determine inflation at a tangency between a social marginal rate of transformation and a social marginal rate of substitution provides a natural framework within which to study inflation scares: fluctuations in private agents' inflation expectations. Although inflation expectations are constant in the economy studied in Section 4.2, expected inflation could vary in a more realistic model. Moreover, when private agents have more information than the monetary authority, variations in expected inflation could constitute another source of shock to the economy from the monetary authority's point of view.

Forecastable changes in the variances of the underlying productivity or velocity shocks would be one way to introduce variable inflation expectations. In a more elaborate model, forecastable cyclical movements in employment would do the same. Moreover, changes in fiscal policy involving government purchases, transfers, and taxes could shift the social marginal rate of substitution between the nominal interest distortion and the markup distortion. And, fiscal policies that affect the private marginal return to work effort could shift the Phillips curve tradeoff.

5.3. A Long-Run Phillips Curve Tradeoff

In a recent paper, King and Watson (1994) provide empirical support for the existence of a long-run Phillips curve tradeoff between unemployment and inflation in postwar U.S. data. For a Keynesian identification in the early sample period, they find that a 1 percent increase in inflation is associated with a 1.3 percentage point decline in the unemployment rate; the estimate is cut roughly in half in the latter part of the sample period. A Rational Expectations Monetarist identification yields a long-run tradeoff of about 1 to 0.5 in the earlier period, and 1 to 0.3 in the latter.

The model offered in this paper is consistent with the evidence of a permanent Phillips curve tradeoff under rational expectations and optimization on the part of private agents, firms, and the monetary authority. The model predicts the extent of a tradeoff to depend on the conditional forecast variance of the incipient markup. The model also predicts the tradeoff to worsen with higher marginal income taxes, increasingly generous unemployment compensation, or more liberal welfare programs. The reasons is that such distortions operate much as the markup does, by driving a wedge between labor's marginal product and its marginal compensation. Since each level of employment in the model is supported by a specific gap between the marginal product of labor and the real wage, it takes a higher before tax real wage, or equivalently, a smaller markup, for the economy to support a given level of employment in the presence of an income tax. An income tax worsens the Phillips curve tradeoff that arises under uncertainty because it associates a given mean incipient markup and expected inflation with higher average unemployment.

6. Conclusion

The model was offered as a conceptual framework for understanding and analyzing moderately inflationary monetary policy. It deliberately embodies Keynesian price level inflexibility to study the cost of a moderate inflations in a way that addresses the concerns of central bankers. The result is a transmission mechanism that allows monetary policy to influence aggregate supply by managing the "markup tax" on employment.

The monetary authority faces a Phillips curve tradeoff that arises because it is imperfectly informed about the state of the economy. The tradeoff emerges because firms react asymmetrically to the incipient markup at the boundary of the range of indeterminacy. If the markup comes in below the acceptable minimum, firms raise prices, if it comes in above, firms do not change prices, the markup is sustained, and employment and output fall.

Since the actual markup cannot fall below the minimum acceptable to firms, policymaker uncertainty raises the average markup and lowers the levels of employment and output at which the economy can be expected to operate. The monetary authority can compensate with monetary policy for the negative effect of its ignorance on economic performance. Expansionary policy raises expected employment by improving the chances that aggregate demand will be sufficient to minimize the markup; but it does so at the cost of higher expected inflation. The optimizing monetary authority chooses an inflation (or deflation) rate to minimize the overall deadweight cost due to the nominal interest rate and the markup distortions.

Sufficiently inflationary monetary policy can keep the realized markup approximately equal to the minimum acceptable to firms in the model, avoiding Keynesian unemployment entirely. In fact, at high inflation rates the model behaves like a noncompetitive real business cycle model in which money affects real variables only through expected inflation. Nevertheless, such highly inflationary policy is not optimal. The reason is that the Phillips curve tradeoff becomes very steep at low levels of unemployment. The marginal welfare cost, in terms of a higher nominal interest rate that must be tolerated to reduce unemployment, exceeds the marginal benefit before Keynesian unemployment is eliminated entirely.

We showed that the welfare cost of inflation may very well be much higher than is commonly supposed if a shopping time technology underlies the demand for money and the average markup in the economy is significant. Making use of the fact that Keynesian unemployment in the model is due entirely to the markup distortion, we also calculated the welfare cost of unemployment. It is an open question, in terms of welfare, how much inflation the monetary authority should be willing to tolerate to reduce unemployment. In fact, it is entirely possible that the deflationary Friedman Rule is optimal in this model, in spite of the model's Keynesian features.

FOOTNOTES

1. The population size is fixed and normalized to unity.

2. R_t in equation (6) is an approximation for $R_t/(1 + R_t)$.

3. Employment and shopping time are invariant to productivity growth (X) because the latter exerts exactly offsetting substitution and wealth effect on employment. And money demand is proportionate to the transactions scale variable, which means that shopping time is invariant to productivity growth by (2) and (3).

4. Stiglitz (1984) also discusses monopolistic competition with a kinked demand curve. Such a market structure has been employed by Woglom (1982) to demonstrate the possibility of underemployment with rational expectations. It has also been used by Ball and Romer (1990), in combination with small frictions in nominal adjustment, as a source of real rigidity in a model designed to explore the nonneutrality of money.

5. Woodford (1991) stresses the importance of multiple real equilibria for understanding business cycles in a model with a market structure closely related to the one studied here.

6. Bils (1987) reports that markups in two-digit manufacturing data for the U.S. decline on average by 3.3 percent with a 10 percent expansion. Rotemberg and Woodford (1991, 1992) discuss the attractiveness of countercyclical markups for macroeconomics, and survey alternative theories of endogenous markups. Carlton (1989) also surveys evidence of the cyclicality of markups.

7. This policy prescription is associated with Milton Friedman (1969).

8. Assume that the shocks are small enough that the incipient markup never falls above $\overline{\mu}$ and the economy is never in the Deflation Regime.

9. Since private agents observe all the variables in the monetary authority's rule, they immediately observe any deviation from that rule, and calculate the implied increase in inflation that a deviation implies.

10. In a different context, Grossman and Van Huyck (1986) also show that an optimal rule can be sustained without commitment if the loss of reputation from deviations is immediate. Ireland (1994) contains a thorough analysis of sustainable monetary policies.

11. Such an elaboration would be along the lines of the imperfect information model in Ball and Romer (1990).

12. See Goodfriend (1993).

REFERENCES

Ball, Laurence, and David Romer, 1990, Real rigidities and the nonneutrality of money, *Review of Economic Studies* 57, 183-203.

Bailey, Martin, 1956, The welfare cost of inflationary finance, *Journal of Political Economy* 64, 93-110.

Barro, Robert, and David Gordon, 1983, Rules, discretion, and reputation in monetary policy, *Journal of Monetary Economics* 12, 101-21.

Basu, Susanto, and John G. Fernald, 1994, Constant returns and small markups in U.S. manufacturing, (Board of Governors of the Federal Reserve System, International Finance Discussion Paper, Washington, DC).

Bils, Mark, 1987, The cyclical behavior of marginal cost and price, *American Economic Review* 77, 838-55.

Blanchard, Olivier and Nobuhiro Kiyotaki, 1987, Monopolistic competition and the effects of aggregate demand, *American Economic Review* 77, 647-666.

Carlton, Dennis, 1989, The theory and the facts of how markets clear: Is industrial organization valuable for understanding macroeconomics? In: R. Schmalensee and R.D. Willig, eds., *Handbook of Industrial Organization*, Volume 1 (North Holland, Amsterdam) 907-946.

Chari, V.V., Patrick Kehoe, and Edward Prescott, 1989, Time consistency and policy, In: Robert Barro, ed., *Modern Business Cycle Theory* (Harvard University Press, Cambridge, MA).

Friedman, Milton, 1969, *The optimum quantity of money and other essays*, (Chicago, Aldine Publishing Co.).

Goodfriend, Marvin, 1993, Interest rate policy and the inflation scare problem: 1979-1992, Federal Reserve Bank of Richmond *Economic Quarterly*, 1-24.

Grossman, Herschel I., and John B. Van Huyck, 1986, Seigniorage, inflation, and reputation, *Journal of Monetary Economics* 18, 21-31.

Hall, Robert, 1988, The relation between price and marginal cost in U.S. industry, *Journal of Political Economy* 96, 921-947.

Ibbotson, Roger, 1994, *Stocks, bonds, bills, and inflation 1994 Yearbook*, (Chicago, Ibbotson Associates).

Ireland, Peter, 1994, Sustainable monetary policies, (Richmond, Federal Reserve Bank of Richmond).

King, Robert G., and Mark W. Watson, 1994, The post-war U.S. phillips curve: A revisionist econometric history, (Charlottesville, University of Virginia), forthcoming in the Carnegie-Rochester Conference Series on Public Policy.

Lucas, R.E., Jr., 1993, 1994, On the welfare cost of inflation, (Chicago, University of Chicago).

McCallum, Bennett T., and Marvin Goodfriend, 1987, Demand for money: Theoretical studies, In: John Eatwell, Murray Milgate, and Peter Newman, eds., *The New Palgrave: A Dictionary of Economics*, (Stockton Press, London: Macmillan; NewYork) 775-81.

Phillips, A.W., 1958, The relation between unemployment and the rate of change of money wage rates in the United Kingdom, 1861-1957, *Economica* 25, 283-300.

Rotemberg, Julio, and Michael Woodford, 1992, Oligopolistic pricing and the effects of aggregate demand on economic activity, *Journal of Political Economy* 100, 1153-1207.

_______, 1991, Markups and the business cycle, In: O.J. Blanchard and S. Fischer, eds., *NBER Macroeconomics Annual 1991*, (MIT Press, Cambridge) 63-129.

Stiglitz, Joseph E., 1987, Competition and the number of firms in a market: Are duopolies more competitive than atomistic markets? *Journal of Political Economy* 95, 1041-61.

_______, 1984, Price rigidities and market structure, *American Economic Review* 74, 350-355.

Woglom, Geoffrey, 1982, Underemployment equilibrium with rational expectations, *Quarterly Journal of Economics,* 89-107.

Woodford, Michael, 1991, Self-fulfilling expectations and fluctuations in aggregate demand, In: Mankiw and Romer, eds., *New Keynesian economics* (MIT Press, Cambridge, MA) 77-112.