The Effects of Inflation on the Number of Firms and Firm Size

A typical money and growth model generally incorporates an implicit assumption that the number of firms (or the set of goods available) is fixed. This paper attempts to investigate the implications of relaxing this assumption in a monopolistically competitive model with endogenous markup. It is found that among other effects, inflation reduces the number of firms and each firm’s size; moreover, due to this new channel, inflation induces secondary effects. One direct implication is that the welfare costs of inflation in our framework are substantially higher than those documented in existing models with standard features. Our findings suggest that it is the lessening of competition that appears to be the primary driving force.

ONE OF THE CENTRAL ISSUES in the literature on money and growth, in the years following the classic papers of Tobin (1965), Sidrauski (1967) and Brock (1974), has been the analysis of the effects of anticipated inflation. In an early work, Dornbusch and Frenkel (1973) provided a comprehensive survey and insightful discussion of the issue. More recently, these effects have been investigated in a variety of general equilibrium models; see, for example, Stockman (1981), Lucas (1982), Abel (1985), Lucas and Stokey (1987), Cooley and Hansen (1989), Gomme (1993), Ireland (1994), Marquis and Reffett (1994), and Jones and Manuelli (1995). This literature has not only identified the mechanisms, from different perspectives, through which changes in monetary policy might affect the real sector of an economy, but also obtained estimates on the quantitative effects of these changes.

All these studies, however, start from an implicit assumption that the number of firms (or the set of goods available) in the economy is fixed. As a result, policy interventions do not affect this variable. Recent theoretical work, in particular in the analysis of growth and international trade, shows that the assumption that the set of

The authors thank two anonymous referees of this journal for helpful and constructive comments. They have also benefitted from conversations with Mike Wickens as well as comments from conference participants of the 1996 Royal Economic Society Annual Meeting (UK).

YANGRU WU is associate professor of finance and economics in the Faculty of Management at Rutgers University. E-mail: yangruwu@andromeda.rutgers.edu JUNXI ZHANG is associate professor of economics at the University of Hong Kong.

Journal of Money, Credit, and Banking, Vol. 33, No. 2 (May 2001, Part 1)
Copyright 2001 by The Ohio State University
goods never changes is rather restrictive; moreover, a seemingly small change in such an assumption can have very important positive and normative implications (for example, see Romer 1994). This paper reexamines the effects of inflation by departing from the above assumption and introducing market power.

We present a simple general equilibrium cash-in-advance model with monopolistic competition and increasing returns to scale to analyze the implications once the aforementioned assumption is relaxed. Following Gali (1995), in this paper we assume that competition among a continuum of monopolistically competitive intermediate goods producers intensifies with the level of economic development and the associated range of intermediate products available. In contrast to conventional real models with monopolistic competition, for example, Spence (1976) and Dixit and Stiglitz (1977), our model endogenizes each firm’s markup, in a sense that entry and exit of firms can affect incumbent firms’ price-cost margins. Specifically, the presence of market power drives wedges between factor shares and their respective elasticities of production, and it is this channel through which inflation generates effects on markups and thus on the number of firms in the intermediate goods sector and each firm’s size.

To visualize this channel, it is important to note that in a cash-in-advance economy of this kind, changes in the rate of monetary expansion, by changing the rate of inflation, will ultimately affect the opportunity cost of holding money, and real allocative effects will result from individuals’ responses to this increased cost. We refer to these effects as primary effects.

One interesting feature is that our model also generates secondary effects. These effects occur through a decline in the number of firms. Exit of firms lessens the competition and raises incumbent firms’ markups, which in turn lowers the marginal return of investment. This lessening of competition induces further decline in output, capital stock, and other real variables. These secondary effects produce, unambiguously, larger magnitudes in our model than those in conventional models.

We then examine the welfare costs of inflation in our monopolistically competitive model with endogenous markup. It is conceivable that the presence of secondary effects along with the usual primary effects can lead to higher welfare costs than those documented in existing studies with standard features, and this result is con-

---

1. Pagano (1990) also allows equilibrium markups to be negatively related to the number of firms in a discrete-time model. As will become clear below, our results hinge on the presence of such a relationship, but not on the particular mechanism chosen to generate it. Any other choice of technologies, preferences, and market structure that preserved the property would generate similar outcomes. For example, Smith and Venables (1988), Yang and Heijdra (1993), and d’Aspremont et al. (1996) develop models with Bertrand competition and free entry, in which price elasticity of demand at firm level (evaluated at symmetric equilibria) is proportional to the number of firms in the industry. Gali and Zilibotti (1995) and van de Klundert and Smulders (1997) obtain the same relationship in models with Cournot competition and free entry.

2. Following Gali (1995), equilibrium markup is assumed to be a function of the number of firms in the intermediate good sector [also see equation (22)]. Such a functional form is chosen only for illustration rather than reflection of realism and it is only intended to capture the notion that greater competition diminishes the firm’s ability to mark up the product price. It is known from the empirical industrial organization literature that there is little evidence of falling markups beyond three firms. To address this finding, one has to employ a slightly different functional form.
firmed in a number of numerical experiments. In addition, to understand and evaluate the impact of the endogenous markup, we isolate the effects of monopolistic competition from those of the endogenous markup. It is found that endogenizing markup is the primary factor for the higher welfare cost in our framework. To summarize, the main finding of our study seems to suggest that conventional theoretical perspectives lead to substantial underestimates of the magnitudes of the real effects and the welfare costs of inflation.\(^3\)

The remainder of the paper is organized as follows. In section 1, we cast the basic model and define a perfect foresight competitive equilibrium. Section 2 characterizes the steady state and determines equilibrium values of the real variables. Section 3 examines the effects of inflation, while section 4 investigates the welfare costs of inflation in the model. Finally, section 5 offers some concluding remarks.

1. A BASIC MODEL OF INFLATION AND ENDOGENOUS MARKUPS

A. The Economic Environment

Consider an economy inhabited by a large number of infinitely lived identical individuals who possess perfect foresight. Utility at any moment \( t \) is derived from per capita consumption, \( c_t \), and per capita leisure, \( \ell_t \), or equivalently per capita work effort, \( h_t \). Preferences are given by the following utility function,

\[
\int_0^\infty e^{-\rho t} u(c_t, h_t) dt,
\]

where \( \rho \) is a constant, positive rate of time preference. Function \( u \) is assumed to be strictly concave and twice continuously differentiable, satisfying \( u_c > 0, u_h < 0, u_{cc} < 0, u_{ch} = 0, u_{hh} < 0 \), and the Inada conditions.

Individuals are assumed to be endowed with one unit of time and supply labor to a firm that produces the goods. Individuals are also engaged in accumulating capital, \( k_t \), which they rent to the firm. In addition, they carry over real money balances, denoted by \( m_t \), and these balances are augmented with a lump-sum transfer in real term equal to \( \tau_t \) from the government. The government finances this transfer to individuals through the creation of money, facing the per capita budget constraint

\[
\tau_t = gm_t,
\]

where \( g \) is the constant growth rate of money supply.

\(^3\) It is interesting to note that policymakers appear to believe that the costs of inflation are large while academics find that the costs of inflation are fairly small. For the past two decades or so, policymakers in many countries have tightened monetary policies, set formal inflation targets, and taken various steps for fear of runaway inflation.
Individuals are required to finance their purchases of the consumption good through the previously acquired cash balances. That is, a typical individual faces a cash-in-advance (CIA) constraint of the form

$$c_t \leq m_t + \tau_t.$$  

(3)

A representative individual receives factor payments in real terms for capital and labor respectively, $r, k_t, w, h_t$, where $r$ and $w$ are the rental price of capital and the wage rate, all in terms of the final output good to be specified below. Along with any unspent cash balances, the individual allocates her earnings between purchases of the investment good, $i_t$, and the accumulation of real cash balances to carry forward, $\dot{m}_t$, leading to the dynamic budget constraint

$$c_t + i_t + \dot{m}_t \leq w_t h_t + r_t k_t + \tau_t - \pi_t m_t,$$  

(4)

where $\pi_t$ is the rate of inflation. Investment is undertaken to augment the capital stock owned by the individual, with the capital stock obeying the following law of motion:

$$k_{t+1} = i_t - \delta k_t,$$  

(5)

where $\delta \in [0, 1]$ is the depreciation rate of capital.

On the production side, there exists a continuum of monopolistically competitive intermediate goods producers at time $t$, indexed by $j \in [0, n_t]$, where $n_t$ is the range of intermediate goods or the number of firms. Final output, produced in a competitive sector, is given by

$$Y_t = \left[n_t^{-1+1/(\mu(n_t))} \int_0^{n_t} Y(j)^{1/(\mu(n_t))} dj \right]^\mu(n_t).$$  

(6)

where $Y_t$ is final output, $Y(j)$ is the quantity of intermediate goods used, and $\mu(n_t)$ measures the degree of monopoly power in the markets for intermediate products. Throughout this paper, capital letters are used to distinguish per capita variables that a competitive individual takes as parametric from individual-specific variables that are chosen by her, although in equilibrium they will be the same. Following the standard practice in the monopolistic competition literature, in (6), the presence of the loading term $n_t^{-1+1/(\mu(n_t))}$ is included to avoid a “taste for variety.” Moreover, the $\mu(n_t)$ function is assumed to satisfy that $\mu'(n_t) < 0$, $\lim_{n_t \to 0} \mu(n_t) = \bar{\mu} \in (1, \infty)$, and $\lim_{n_t \to \infty} \mu(n_t) = 1$. Intuitively, these properties imply that the smaller the number of firms in the industry, the less intense the competition, and then the higher the $\mu(n)$ and the monopoly profits. In what follows, $\mu(n_t)$ corresponds to the optimal markup.

Let $P_t(j)$ be the relative price of the $j$th intermediate good in terms of the final good at time $t$, the profit of a final good producer can now be written as
\[ \Pi_i = Y_i - \int_0^1 P_i(j)Y_i(j)dj. \]  

(7)

Maximizing the profit function (7) subject to the constraint (6) yields the demand function for a typical intermediate good:

\[ Y_i(j) = P_i(j)^{-\xi(n_i)}(Y_i n_i). \]  

(8)

where \( \xi(n_i) = \mu(n_i)/[\mu(n_i) - 1] \) is the elasticity of substitution among intermediate goods. Evidently, this elasticity is no longer constant in this paper, which extends the popular models of Spence (1976) and Dixit and Stiglitz (1977) in the literature on monopolistic competition.

The production technology available to an intermediate commodity producer is assumed to exhibit the Cobb-Douglas form:

\[ Y_i(j) = F(K_i(j), H_i(j)) = AK_i(j)^\alpha H_i(j)^{1-\alpha}, \]  

(9)

where \( \alpha \) is the capital elasticity of production. By making use of (8) and (9), we can obtain the reduced form of the profit function for the \( j \)th intermediate good producer:

\[ \Pi_i(j) = (Y_i n_i)^{1/\xi(n_i)} A^{1/\mu(n_i)} K_i(j)^{\alpha/\mu(n_i)} H_i(j)^{(1-\alpha)/\mu(n_i)} - w_i H_i(j) \]
\[ - r_i K_i(j) - P_i(j)\phi, \]  

(10)

where \( \phi \) represents the fixed cost. In other words, an amount \( \phi \) of the intermediate good is immediately used up for administration purposes to keep production going, and this is independent of how much output is produced. This fixed cost implies that there exist increasing returns to scale, although capital and labor elasticities sum to one.

The first-order conditions for the intermediate good producer’s problem yield the following functions for the wage rate and the rental rate of capital:

\[ w_i = \frac{1-\alpha}{\mu(n_i)} \frac{P_i(j)Y_i(j)}{H_i(j)}, \]  

(11a)

\[ r_i = \frac{\alpha}{\mu(n_i)} \frac{P_i(j)Y_i(j)}{K_i(j)}. \]  

(11b)

It should be noted that in equations (11a) and (11b), \( (1-\alpha)/\mu(n_i) \) and \( \alpha/\mu(n_i) \) correspond to the labor and capital shares, respectively. By the assumption that \( \mu(n_i) > 1 \) for \( 0 < n_i < \infty \), factor shares are less than their respective elasticities of production and sum to less than one; moreover, these shares vary with the number of firms in the industry or the firms’ markups, which is in stark contrast to the conventional models with monopolistic competition.
To simplify the subsequent analysis, we assume that all firms are symmetrical. This implies that in equilibrium all firms employ the same amounts of inputs, produce the same quantities and set the same prices: $K_j(n) = K/n$, $H_j(n) = H/n$, $Y_j(n) = Y/n$, and $P_j(n) = 1$, where the latter comes from the fact that the final good sector is competitive. Finally, free entry drives profits of intermediate goods producers to zero, leading to

$$\frac{1}{\mu(n)} = 1 - \frac{\phi n_i}{AK_t^a H_t^{1-a}}. \quad (12)$$

B. Perfect Foresight Equilibrium

The representative consumer is assumed to behave competitively, taking all prices as given. Her problem is to choose a path $\{c^*, h^*, k^*, m^*, l^*, v^*, \ldots \}$ to maximize (1) subject to (3)–(5). The current-value Hamiltonian for this optimal problem is (superfluous time index is dropped hereafter)

$$H = u(c, h) + \lambda_1 (wh + rk + \tau - \pi m - c - i) + \lambda_2 (i - \delta k) + \lambda_3 (\tau + m - c),$$

where $\lambda_1$ and $\lambda_2$ are the shadow prices used to value increments to real balances and capital respectively, and $\lambda_3$ is the Lagrange multiplier associated with the CIA constraint (3). The first-order conditions for this problem are thus:

$$u_t(c, h) = \lambda_1 + \lambda_3, \quad (13a)$$

$$\lambda_1 = \lambda_2, \quad (13b)$$

and

$$-u_t(c, h) = w \lambda_1. \quad (13c)$$

On the margin, equations (13a) and (13b) state that goods must be equally valuable in their two uses—consumption and capital accumulation, while equation (13c) states that time must be equally valuable in its two uses—leisure and production.

The rates of changes of the prices $\lambda_1$ and $\lambda_2$ are given by

$$\dot{\lambda}_1 = \rho \lambda_1 + \pi \lambda_1 - \lambda_3, \quad (13d)$$

$$\dot{\lambda}_2 = \rho \lambda_2 - r \lambda_1 + \delta \lambda_2, \quad (13e)$$

where standard transversality conditions are omitted. A sufficient condition for the first-order conditions to maximize (1) is that the Hamiltonian is jointly concave in $(k, m)$ after the controls $c$ and $h$ are substituted out with their maximizing values. It can be easily shown that this condition is satisfied.
equilibrium of this model is characterized by the set of the first-order conditions (13) and the pricing equations (11a)–(11b); moreover, the money market, the labor market, the capital market, and the final good market all clear: \( h = H, k = K, \) and

\[
c + \dot{k} = wh + rk - \delta k, \tag{14a}
\]
\[
m = \tau - \pi m. \tag{14b}
\]

The subsequent analysis focuses exclusively on a circumstance under which the CIA constraint holds with equality, or equivalently \( \lambda_3 > 0. \) This condition requires the growth rate of money supply, \( g, \) to exceed the negative value of the discount rate of time preferences, \( \rho. \)

2. STEADY-STATE ANALYSIS

In this section, we solve for the steady-state equilibria of this model. In order to find the steady states, we need to reduce the dimensionality of the system by expressing the three multipliers in terms of the control variables \( c \) and \( h. \) Since \( \lambda_1 = \lambda_2 \) and \( \lambda_3 = \lambda_2 = 0, \) we derive an equation immediately from (13e):

\[
r = \rho + \delta. \tag{15}
\]

Then substituting (13b) into (13d) and (13e) yields

\[
\frac{\lambda_3}{\lambda_1} = r + \pi - \delta, \tag{16}
\]

where the inflation rate is constant in the steady states because the growth rate of money supply, \( g, \) is assumed to be constant. Notice that \( r + \pi \) is the nominal rate of interest, so a sufficient condition for the CIA constraint to be binding (that is, \( \lambda_3 > 0 \)) is that this nominal rate exceeds the depreciation rate of capital. Alternatively, from (13d), it follows that \( \pi > -\rho, \) or the inflation rate exceeds the negative value of the discount rate. Combining (13a), (13c), (15), and (16), we obtain

\[
\frac{-u_h(c^*, h^*)}{u_c(c^*, h^*)} = \frac{w}{1 + \rho + \pi}, \tag{17}
\]

where an asterisk denotes the steady-state value of that variable. Consequently, after the substitution of \( w \) and \( r \) from (11a) and (11b), we end up with a reduced system consisting of four equations, (12), (14a), (15), and (17), in four variables, \( c^*, h^*, k^*, \)

5. As usual, by Walras' law, the market-clearing conditions for money and the final good are not independent.
and $\mu(n^*)$ (or $n^*$). Once $c^*$ is determined, the equilibrium value of $m^*$ can be easily obtained from the CIA constraint; furthermore, from (2) and (14b), it follows that steady states require that $\pi = g$, implying that a percentage increase in the money supply is translated into an equal proportional increase in the inflation rate.

Next, we use a specific functional form of the utility in the general model laid out above to investigate the steady-state properties and thereby the effects of inflation on, especially, the number of firms and firm size. We take a form, standard in the real business cycle literature, which reflects indivisible labor; see, for example, Hansen (1985), Rogerson (1988), and Cooley and Hansen (1989):

$$U(c,h) = \text{Inc} - Bh,$$  \hfill (18)

where $B$ is a constant parameter. Hence, the steady states are defined by the following equations:

$$Bc^* = \frac{1}{1 + \rho + \pi \mu(n^* \alpha) h^{* \alpha - \alpha}}, \hfill (19a)$$

$$\frac{\alpha}{\mu(n^*)} Ak^{* \alpha - 1} h^{* \alpha - 1} = \rho + \delta, \hfill (19b)$$

$$c^* = \frac{Ak^{* \alpha} h^{* \alpha - 1}}{\mu(n^*)} - \delta k^*, \hfill (19c)$$

$$\frac{1}{\mu(n^*)} = 1 - \frac{\phi n^*}{Ak^{* \alpha} h^{* \alpha - 1}}. \hfill (19d)$$

Finally, straightforward algebraic manipulation of (19a)-(19d) yields a nonlinear equation in $\mu(n^*)$ or $n^*$:

$$\frac{n^* \mu(n^*)^{1/(1-\alpha)}}{\mu(n^*) - 1} = \frac{D}{1 + \rho + \pi}, \hfill (20)$$

where $D = (1 - \alpha)\alpha^{1/(1-\alpha)} A^{1/(1-\alpha)}/[\phi B(\rho + \delta)^{1/(1-\alpha)} - 1 - (1 - \alpha)\delta]$ is a constant parameter. The solution(s) of (20) can now be illustrated diagrammatically. Let $\Gamma$ and $\Theta$ be the expressions in the left- and right-hand sides of (20), respectively. Clearly, function $\Theta$ is independent of $n^*$ and thus $\mu(n^*)$. For function $\Gamma$, it can be easily shown that if $\mu(n^*) < 1/\alpha$, $\Gamma$ declines monotonically with $\mu(n^*)$. Moreover, note that

---

6. The derivation of (20) is briefly given as follows. From (19b), one obtains an expression for the capital-labor ratio, $k^*/h^*$. Next, substituting it into (19a) yields $c^*$. Then, combine $c^*$ and $k^*/h^*$ with (19c) to produce $k^*$. Finally, substituting $k^*$ and $k^*/h^*$ into (19d), equation (20) follows.
by the assumptions that \( \lim_{n \to 0} \mu(n) = \bar{\mu} \in (1, \infty) \) and \( \lim_{n \to \infty} \mu(n) = 1 \), the domain for \( \mu(n^*) \) is \([1, \bar{\mu}]\). When \( \mu(n^*) \) approaches 1, the \( \Gamma \) curve approaches \(+\infty\); while, when \( \mu(n^*) \) approaches \( \bar{\mu} \), it approaches 0. Figure 1 demonstrates that there exists a unique \( \mu(n^*) \) that solves (20). Interestingly, as long as \( \mu(n) \) belongs to the general class of monotonic and continuous functions and satisfies the above two limiting properties, the uniqueness of the solution (20) is independent of its particular form.

It should be pointed out that the condition that \( \mu(n^*) \leq 1/\alpha \) is empirically plausible. Economists generally agree that estimates of the capital elasticity of production, \( \alpha \), range from 0.25 (for example, Lucas 1988) to 0.42 (for example, Rotemberg and Woodford 1991), while those of markups, based on either the gross output measure or the value added measure, lie within the range 1.05 to 2.3 (for example, Morrison 1990, Norrbom 1993, and Roeger 1995). These numbers suggest that the above sufficient condition is empirically satisfied. Also, once the functional form of \( \mu(\cdot) \) is specified, the above condition can be further relaxed.

At the unique steady state, other endogenous variables are given by

\[
\begin{array}{c}
\Gamma \\
\theta \\
\mu \\
\bar{\mu}
\end{array}
\]

**Fig. 1. Determination of the Optimal Markup \((n^*)\)**

7. For a comprehensive review of some of the empirical evidence, see a recent paper by Schmitt-Grohe (1997, pp. 130–132).

8. In the case where the above condition does not hold, multiple equilibria and a number of complicated equilibrium dynamics may emerge; see Gali (1995). Such an investigation is beyond the scope of the present paper.
\[ k^* = \frac{D_1}{\mu(n^*)^{1/(1-\alpha)}(1 + \rho + \pi)}, \tag{21a} \]

\[ h^* = \frac{D_2}{1 + \rho + \pi}, \tag{21b} \]

\[ c^* = \frac{D_3}{\mu(n^*)^{1/(1-\alpha)}(1 + \rho + \pi)}, \tag{21c} \]

\[ Y^*_\text{net} = Y^* - n^* \phi = \frac{D_4}{\mu(n^*)^{1/(1-\alpha)}(1 + \rho + \pi)}, \tag{21d} \]

\[ Y(j)^*_\text{net} = \frac{Y^*_\text{net}}{n^*} = \frac{D_4}{n^* \mu(n^*)^{1/(1-\alpha)}(1 + \rho + \pi)} = \frac{\phi}{\mu(n^*) - 1}, \tag{21e} \]

where \( D_1 = \alpha \phi D/(\rho + \delta), D_2 = (1-\alpha)(\rho + \delta)/[B[\rho + (1-\alpha)\delta]], D_3 = \phi D[\rho + (1-\alpha)\delta]/(\rho + \delta), \) and \( D_4 = \phi D \) are constant parameters, and \( Y^*_\text{net} \) denotes output net of fixed cost. The last equality in (21c) is obtained through the zero-profit condition (19d). Note that the steady-state value of the real money balances can be obtained from the CIA constraint: \( m^* = c^*/(1 + g). \)

Four comments are in order. First, in contrast to conventional models of monopolistic competition with constant markups, our model with entry and endogenous markups suggests that all model variables except for the labor input are increasing functions of the number of firms. The positive relationship between the number of firms and the level of output per firm (firm size) as in equation (21d) may sound paradoxical, but it in fact bears good intuition. Entry of new firms intensifies competition and drives down the existing firms’ markups. Because in equilibrium, each firm is making zero profit, its output has to be raised in order to amortize the fixed cost.

Second, the independence of the labor input to the number of firms is only a special result, due to our use of the specific form of the utility function. It is therefore conceivable that once more general forms are allowed for, this independence result may no longer hold.

Third, alternatively, equation (21a) suggests that the number of firms is increasing in the per capita capital stock of the economy. In other words, positive net savings lead to entry of firms, thus increasing the variety of intermediate goods available. In this paper, it is this channel through which inflation exerts an effect on the number of firms and hence on firm size as well.

Finally, since the personal income is defined as \( w(h + rk), \) which is equal to \( Y^*_\text{net}, \) equation (21d) indicates that this variable is positively associated with the number of firms, \( n^* \). Moreover, although consumption is also increasing in \( n^* \) as in equation

---

9 In the subsequent analysis, the net output measure—output net of fixed cost—will be used without special reference.
(21c), the long-run average propensity to consume, $D_3/D_4$, is nonetheless a constant, independent of $n^*$.

3. THE EFFECTS OF INFLATION

This section examines the effects of inflation in our economy, in particular on the number of firms and firm size. We start by analyzing the effect on the steady-state markups, $\mu(n^*)$. An increase in the anticipated rate of inflation, $\pi$, resulted from an increase in the rate of money growth, shifts the $\Theta$ curve in Figure 1 downward, while keeping the $\Gamma$ curve intact. As a result, higher inflation yields higher markups; on the other hand, since $\mu'(n^*) < 0$, it reduces the number of firms, $n^*$. Immediately from equation (21c), we conclude that inflation also reduces firm size. These effects on $n^*$ and firm size are the first set of major results in this paper.

Inspection of equations (21a)–(21d) reveals that inflation produces the standard reversed Tobin (1952) effect: it leads to lower consumption, capital stock, real money balances, hours worked, and output. With indivisible labor, the negative impact on hours worked can be interpreted as unemployment. In addition, (21a) and (21b) suggest that the steady-state capital-labor ratio is also lower [alternatively, see (19b)]. However, unlike conventional models, the magnitudes of these effects in our model, in absolute terms, are higher [see equations (21a), (21c), and (21d)], except again for that on the labor input. This stems from the fact that those conventional models, based on the assumption that entry and exit of firms do not affect existing firms’ price-cost margins, fail to take into account the effect of inflation on the markups. Mathematically, direct differentiation of equations (21a), (21c), and (21d) with respect to inflation, $\pi$, yields two terms. The first term features the usual effects, that is, the primary effects, while the second features the new effects via reducing the number of firms and thus the markups for incumbent firms, that is, the secondary effects. One immediate implication of these results is that this model leads to higher welfare costs of inflation, which will be studied in the next section. To reiterate, the general equilibrium effects of the endogenous entry and exit of firms increase the welfare costs of inflation.

To examine the inflation effects described above, we choose a specific, albeit rather restrictive, functional form of $\mu(n)$. As in Gali (1995), $\mu(n)$ takes the form such that the elasticity of substitution among inputs is linear: $\xi(n) = \mu(n)/[\mu(n)-1] = \varepsilon_0 + \varepsilon_1 n$, where $\varepsilon_0 \in [1, \infty)$ and $\varepsilon_1 \in (0, \infty)$ are constants. The steady-state markups are then determined by

$$\mu(n^*) = 1 + 1/(\varepsilon_0 + \varepsilon_1 n^*) - 1.$$  

10. The positive correlation between long-run unemployment and inflation rates are known as the long-run Phillips curve, which is also obtained by Greenwood and Huffman (1987) and Cooley and Hansen (1989). The empirical evidence presented by Cooley and Hansen supports this relationship.
In order to obtain an expression for the number of firms as a function of the capital stock, using (20) we rewrite (21a) as

\[ k^* = \frac{\alpha \phi n^*}{(\rho + \delta)[\mu(n^*) - 1]} . \]  

(23)

Substituting \( \mu(n^*) \) from (22) into (23) and solving for \( n^* \), we have

\[ n^* = \frac{(1 - \epsilon_0) + \sqrt{(1 - \epsilon_0)^2 + 4 \epsilon_1 (\rho + \delta) k^* / (\alpha \phi)}}{2 \epsilon_1} . \]  

(24)

Evidently, the capital stock, \( k^* \), positively affects the number of firms, \( n^* \).

Intuitively, higher inflation induces individuals to economize on real money holdings, but since money is used in purchasing the consumption good, real purchases of this good fall with decreased money holdings. Consequently, equilibrium production decreases as well. Notice that the decline in output is more than that in consumption, due to the fact that marginal propensity to consume is a positive number smaller than one. This induces less investment and lower steady-state capital stock. Lower output creates lower derived demand for labor, and thus the steady-state work input declines. On the supply side, households partially offset the loss of utility associated with lower consumption by adjusting along the labor-leisure margin, thereby reinforcing the effect from the demand side to reduce labor effort.\(^{11}\)

To appreciate the importance of the presence of secondary effects, we note that as discussed above an increase in inflation reduces aggregate output. While the fall in aggregate output could be accommodated by an equiproportional fall in output of each of the monopolistically competitive firms, the presence of a fixed cost implies that at least part of the adjustment in the total output of the monopolistically competitive firms occurs through a decline in the number of firms, due to the zero-profit condition. The negative relationship between the markup of a survival firm and the number of firms then implies that the markup will rise, reflecting a fall in competition. It is this lessening of competition that generates larger effects of inflation as well as larger welfare costs of inflation in this paper.

4. WELFARE COSTS OF INFLATION

As discussed in the preceding section, one of the major results in this paper is that our model with monopolistic competition and endogenous markups generates higher welfare costs of inflation than related models with constant markups. We illustrate this by first evaluating the steady-state values of the key variables and then comput-

\(^{11}\) We thank an anonymous referee for pointing out this labor supply effect.
ing the welfare costs under various monetary growth rates. Our benchmark welfare measure is calculated based on the assumption that the relevant length over which the economic agents are constrained to hold money is one year. In order to make our results comparable to those in the existing literature, the parameters used in this paper are, wherever possible, chosen to be the same as those suggested in other studies in this area. Specifically, the following parameters are taken from Gali (1995): $A = 0.397$, $\delta = 0.10$, $\rho = 0.04$ and $\epsilon_0 = 1.27$. Using data from Citibase, we estimate the average annual inflation rate in the United States between 1959 and 1989 to be 4.8 percent. We then choose $\epsilon_1$ such that at the 4.8 percent annual inflation rate, the equilibrium markup, $\mu$, is equal to 1.5, a value suggested by Hornstein (1993). This leads to $\epsilon_1 = 0.138$. The parameters $\alpha$ and $B$ are taken from Cooley and Hansen (1989): $\alpha = 0.36$ and $B = 2.86$. Finally, as for the fixed cost parameter, $\phi$, we have made the arbitrary decision to set the steady-state fixed cost equal to 4 percent of total revenue when the inflation rate is equal to the average U.S. inflation rate (4.8 percent), or $\phi = 0.0025$. The above choice of parameters produces the steady-state quarterly capital to output ratio equal to 10.286, which is very close to Cooley and Hansen’s (1989) value of 10.263.

As a further confirmation, we also allow the relevant length over which individuals are constrained to hold cash to be one quarter and one month, respectively. The parameters $A$, $\delta$, $\rho$, and $\phi$ need to be scaled accordingly in these two cases. Namely, in the case of quarterly constraint, we divide each of these four parameters by four to convert them to the quarterly rates; while in the case of monthly constraint, we divide them by twelve to convert them to the monthly rates.

Equations (21a)–(21d) serve as the basic formulae for our calculation of the steady-state values. As for the welfare cost computation, to simplify the analysis, we assume that there are no other distortions in the economy, apart from the inflation tax. Under each particular inflation rate, the welfare cost is measured by the increase in consumption which is necessary for the economic agent to be as well off as she would be in the case when the CIA constraint (3) is not binding. As has been shown in section 2, when the inflation rate is less than or equal to the negative of the discount factor (that is, $\pi \leq -\rho$), the CIA constraint is not binding. Let $\hat{c}$ and $\hat{h}$ denote the steady-state values of consumption and work effort when $\pi = -\rho$, and $c^A$ and $h^A$ be the steady-state values of consumption and work effort under an alternative rate of inflation, respectively. Using the utility function (18), the welfare cost, denoted by $\Delta c$, is defined as follows:

$$\ln(\hat{c}) - \hat{B}\hat{h} = \ln(c^A + \Delta c) - B h^A.$$ (25)

We express the cost as the percentage of steady-state values of consumption and output. The way that the welfare cost is defined in (25) is fairly standard in the literature.

12. The results reported in Table 1 are not sensitive to the choice of $\phi$ and one may obtain similar results over a range of values. We explored values in which the cost lies in between 2 percent to 6 percent of total revenue and obtained similar results in all cases.
Table 1 presents the welfare costs as well as the steady-state values of some key variables. It is observed that higher inflation lessens competition by reducing the number of firms and hence raises the optimal markup. As a result, steady-state values of capital stock, work effort, output, and consumption all unambiguously decline. While the increase in leisure improves the agent’s welfare, the reduction in consumption more than offsets the leisure effect so that the total utility level of the agent drops.

A striking feature in Table 1 is that the welfare costs of various inflation rates are quite high, which is consistent across all inflation rates. For example, when the relevant length over which individuals are constrained to hold currency is a quarter, a 10 percent inflation results in a welfare cost of 1.632 percent in terms of consumption. By contrast, using conventional welfare triangles analysis, Fischer (1981) obtains the deadweight loss of a 10 percent annual inflation to be about 0.3 percent of GNP, while Lucas (1981) offers a figure of 0.45 percent. In general equilibrium real-

<table>
<thead>
<tr>
<th>TABLE 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>STEADY-STATE AND WELFARE COSTS OF ANTICIPATED INFLATION</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>Annual Inflation Rate</td>
</tr>
<tr>
<td>-4%</td>
</tr>
<tr>
<td>Annual Constraint</td>
</tr>
<tr>
<td>Steady State: Consumption</td>
</tr>
<tr>
<td>Work Effort</td>
</tr>
<tr>
<td>Capital Stock</td>
</tr>
<tr>
<td>Output</td>
</tr>
<tr>
<td>Markup (µ)</td>
</tr>
<tr>
<td>Welfare Costs: (Δc/c^4)×100</td>
</tr>
<tr>
<td>(Δc/c^4)×100</td>
</tr>
<tr>
<td>Quarterly Constraint</td>
</tr>
<tr>
<td>Steady State: Consumption</td>
</tr>
<tr>
<td>Work Effort</td>
</tr>
<tr>
<td>Capital Stock</td>
</tr>
<tr>
<td>Output</td>
</tr>
<tr>
<td>Markup (µ)</td>
</tr>
<tr>
<td>Welfare Costs: (Δc/c^4)×100</td>
</tr>
<tr>
<td>(Δc/c^4)×100</td>
</tr>
<tr>
<td>Monthly Constraint</td>
</tr>
<tr>
<td>Steady State: Consumption</td>
</tr>
<tr>
<td>Work Effort</td>
</tr>
<tr>
<td>Capital Stock</td>
</tr>
<tr>
<td>Output</td>
</tr>
<tr>
<td>Markup (µ)</td>
</tr>
<tr>
<td>Welfare Costs: (Δc/c^4)×100</td>
</tr>
<tr>
<td>(Δc/c^4)×100</td>
</tr>
</tbody>
</table>

Note: Parameters are specified as: α = 0.36, B = 2.86, ε = 1.27 and e = 0.138. The remaining parameters are selected as follows: in the case of annual constraint, A = 0.397, δ = 0.10, p = 0.04, and φ = 0.0025; in the case of quarterly constraint, we divide each of these four parameters by four to convert them to the quarterly rates, and in the case of monthly constraint, we divide them by twelve to convert them to the monthly rates.

13. It should be pointed out that these costs are measured relative to a 0 percent inflation rate, while in the dynamic general equilibrium studies cited below, they are measured relative to an optimal rate of inflation.
business-cycle models, Cooley and Hansen (1989, 1991) report a cost of about 0.520 percent. Due to the presence of secondary effects in this monopolistically competitive model with endogenous markups, the larger welfare costs result becomes quite intuitive. The heart of our results is that higher inflation lessens competition and raises the optimal markups.

To formally address this latter point, we compute the welfare costs with constant markups and compare them with those obtained with endogenous markups for various inflation rates (0 percent, 5 percent, and 10 percent). Specifically, we take $\mu = 1.5$ as given, and compute the welfare costs, while keeping all other model parameters the same as in the endogenous markup case. This experiment allows us to isolate the effects of monopolistic competition from those of endogenous markups and will shed more light on the underlying force for the relatively higher welfare cost of inflation. Since the results on the costs in terms of output are qualitatively the same as those in terms of consumption, in what follows, we only report the costs in terms of consumption. The results are reported in Panel A of Table 2. To save space, Table 2 presents only the case of a quarterly CIA constraint. For comparison, we also compute and report the costs under the Cooley and Hansen (1989) model.

Two interesting observations are worth mentioning. First, at any given inflation rate, the welfare cost under endogenous markup is significantly higher than under constant markup (about three times as large). At the 10 percent inflation rate, the welfare cost is 1.632 percent under endogenous markup while 0.528 percent under constant markup.

Second, the welfare cost estimates obtained under constant markup are quite close to those reported in the literature, such as in Cooley and Hansen (1989, 1991) and Lucas (1981). This leads us to conclude that while our model differs from the more standard models in that we have monopolistic competition and increasing returns to scale, these two elements are not the primary ingredients for the higher cost of infla-

<table>
<thead>
<tr>
<th>TABLE 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Welfare Costs of Inflation as a Percentage of Consumption with Quarterly CIA Constraint Under Alternative Specifications</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Models</th>
<th>0%</th>
<th>5%</th>
<th>10%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A. Scale Factor $\gamma = 1$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cooley-Hansen (1989)</td>
<td>0.144%</td>
<td>0.335%</td>
<td>0.520%</td>
</tr>
<tr>
<td>Constant Markup</td>
<td>0.142%</td>
<td>0.330%</td>
<td>0.528%</td>
</tr>
<tr>
<td>Endogenous Markup</td>
<td>0.457%</td>
<td>1.039%</td>
<td>1.632%</td>
</tr>
<tr>
<td>Panel B. Scale Factor $\gamma = 1.05$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant Markup</td>
<td>0.077%</td>
<td>0.185%</td>
<td>0.306%</td>
</tr>
<tr>
<td>Endogenous Markup</td>
<td>0.429%</td>
<td>0.976%</td>
<td>1.535%</td>
</tr>
<tr>
<td>Panel C. Scale Factor $\gamma = 1.10$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant Markups</td>
<td>0.009%</td>
<td>0.033%</td>
<td>0.072%</td>
</tr>
<tr>
<td>Endogenous Markups</td>
<td>0.400%</td>
<td>0.913%</td>
<td>1.438%</td>
</tr>
</tbody>
</table>
tion. The most important factor in this model setting seems to be the secondary effects through endogenous markups. This point will be made even more transparent below when an alternative formulation of increasing returns to scale is introduced.

To check the robustness of our welfare results, we also attempt an alternative formulation of increasing returns to scale. We follow Hornstein (1993) to specify each firm’s production function [as opposed to (9)] as follows:

\[ Y_i(j) = A[K_i(j)^{\alpha}H_i(j)]^{1-\alpha} \gamma, \]  

(26)

where \( \gamma > 1 \) is a scale parameter. Our objective is then to examine how changes in \( \gamma \) affect the magnitudes of inflation effects and the welfare costs of inflation. The derivation of steady-state equations under the new production function (26) is given in the appendix.

We conduct a simple comparative statics analysis to examine how the degree of noncompetitiveness affects the welfare cost of inflation in our model setting. To this end, we compare the welfare costs for two different values of \( \gamma \): \( \gamma = 1.05 \) and 1.10. Because a change in the scale factor \( \gamma \) affects the steady-state values of endogenous variables, we recalibrate the model. Specifically, we choose the discount rate \( \delta \) such that the steady-state quarterly capital to output ratio is the same as that under the basic specification (10.286), and the fixed cost parameter \( \phi \) is again selected such that the steady-state fixed cost is equal to 4 percent of total revenue when the economy has a 4.8 percent annual inflation rate. The values of other parameters remain the same as in our basic model. We obtain \( \delta = 0.107 \) and \( \phi = 0.00195 \) when \( \gamma = 1.05 \); and \( \delta = 0.114 \) and \( \phi = 0.0015 \) when \( \gamma = 1.10 \).

Panels B and C of Table 2 report the welfare costs of three inflation rates using the two respective scale factors. Interestingly, we find that the welfare costs are quite insensitive to the value of \( \gamma \). A higher value of \( \gamma \) slightly lowers the welfare cost. For example, at 10 percent inflation and with endogenous markup, the costs are 1.632 percent for \( \gamma = 1.00 \), 1.535 percent for \( \gamma = 1.05 \) and 1.438 percent for \( \gamma = 1.10 \). Like in the basic case, at any given inflation rate, the welfare cost under endogenous markup is significantly higher than under constant markup in these alternative specifications.

In summary, we find that the effect of the value of \( \gamma \) on the magnitude of the welfare cost is quite small. This observation provides further support to our claim for the basic model that the presence of secondary effects through endogenous markups may be the primary factor responsible for the relatively higher welfare cost of inflation in our model economy.

5. CONCLUDING REMARKS

Standard monetary theory usually incorporates the assumption that the number of firms (or the number of goods available) is fixed. Under this hypothesis, not only are the effects of anticipated inflation on macroaggregates (that is, primary effects) small but the welfare costs of inflation are small as well. The objective of this paper has
been to examine the implications of relaxing such an assumption. To this end, we introduced noncompetitive elements with endogenous markups and increasing returns to scale, two important characteristics emphasized in macroeconomic modeling for the last decade or so, into a money and growth model. It was found that among other effects, inflation reduces the number of firms and each firm's size; moreover, it generates secondary effects, thereby leading to substantially higher welfare costs of inflation.

Our analysis is subject to several qualifications which call for further research. For the sake of brevity, we just outline three of them. First, for simplicity, this model abstracts from transitional dynamics by assuming that the economy jumps immediately to the steady state. Consequently, it leaves an open question whether (and how) inflation affects the speed at which the economy converges to the steady state. It would be interesting to examine the above issues in more detail that allows for transitional effects. Second, this model also abstracts from long-run growth consideration. A recent paper by Wu and Zhang (1998) extends Lucas' (1988) endogenous growth model to a monetary economy and finds that the welfare costs of inflation are considerably higher than those in a conventional model without endogenous growth. 14

Thus, in a more general model with endogenous growth of this kind, one could analyze the effect of inflation on the rate of economic growth in the context of monopolistic competition and reinvestigate the issues studied in this paper. Of course, the welfare and growth rate results will depend crucially on the mechanism of endogenous growth. Finally, another interesting extension will be the addition of a banking sector and working capital financing constraints on firms. This probably will also increase the welfare costs of inflation substantially.

APPENDIX

This appendix provides the steady-state equations for the alternative formulation of increasing returns to scale in section 4 in the text. Under the new production function (26), the factor pricing equations (11a) and (11b) become

\begin{align}
  w_i &= \frac{\gamma(1-\alpha) P_i(j) Y_i(j)}{\mu(n_i) H_i(j)}, \quad \text{(A.1a)} \\
  r_i &= \frac{\gamma \alpha P_i(j) Y_i(j)}{\mu(n_i) K_i(j)}. \quad \text{(A.1b)}
\end{align}

These alterations lead to the following zero profit condition as in (12):

\[
\frac{\gamma}{\mu(n)} = 1 - \frac{\phi n_i^\gamma}{\alpha K^\alpha H^\gamma(1-\alpha)}, \tag{A.2}
\]

and the equation for the determination of the equilibrium value \(\mu(n^*)\)—equation (20)—is replaced by

\[
\frac{n^* \gamma(1+\alpha(\gamma-1)/(1-\alpha))}{\mu(n^*)^{1/(1-\alpha)}} = \frac{\bar{D}}{(1+\rho+\pi)\gamma(1-\alpha)/(1-\alpha)}, \tag{A.3}
\]

where \(\bar{D} = \{(1-\alpha)\gamma(1-\alpha)\gamma(1-\alpha)\lambda A/\phi(1-\alpha)B^\gamma(1-\alpha)(\rho+\delta)(1-\alpha)/(1-\alpha)\lambda\}^{1/(1-\alpha)}\) is a constant parameter.

Correspondingly, the steady-state equations are given by

\[k^* = \frac{\bar{D}_1}{n^* \gamma(1-\alpha)/(1-\alpha)} \frac{\mu(n^*)^{1/(1-\alpha)}}{(1+\rho+\pi)\gamma(1-\alpha)/(1-\alpha)}, \tag{A.4a}\]

\[h^* = \frac{\bar{D}_2}{1+\rho+\pi}, \tag{A.4b}\]

\[c^* = \frac{\bar{D}_3}{n^* \gamma(1-\alpha)/(1-\alpha)} \frac{\mu(n^*)^{1/(1-\alpha)}}{(1+\rho+\pi)\gamma(1-\alpha)/(1-\alpha)}, \tag{A.4c}\]

\[Y_{net}^* = \frac{\bar{D}_4}{n^* \gamma(1-\alpha)/(1-\alpha)} \frac{\mu(n^*)^{1/(1-\alpha)}}{(1+\rho+\pi)\gamma(1-\alpha)/(1-\alpha)}, \tag{A.4d}\]

\[Y_{net}(j)^* = \frac{Y_{net}^*}{n^*} = \frac{\bar{D}_4}{n^* \gamma(1-\alpha)/(1-\alpha)} \frac{\mu(n^*)^{1/(1-\alpha)}}{(1+\rho+\pi)\gamma(1-\alpha)/(1-\alpha)} = \frac{\phi}{\mu(n^*) - \gamma}, \tag{A.4e}\]

where \(\bar{D}_1 = \gamma \alpha \phi D_1 (\rho+\delta), \bar{D}_2 = \gamma (1-\alpha)(\rho+\delta)/[B(\rho+1-\alpha)(\delta)], \bar{D}_3 = \phi \bar{D}[\rho+(1-\alpha)(\delta)](\rho+\delta), \) and \(\bar{D}_4 = \phi \bar{D}\) are constant parameters. It is interesting to note that when \(\gamma = 1\), all these equations collapse to their counterparts in section 4 in the text.

Since the choice of the value of \(\gamma\) is crucial for the welfare costs, we briefly explain how it is taken from the mixed existing empirical results. A number of recent
papers have attempted to estimate the extent of returns to scale for actual economies. The standard technique in these studies is the simple linear regression. Among others, Hall (1990) and Caballero and Lyons (1992) found that there were quantitatively significant increasing returns to scale in U.S. manufacturing. Oulton (1996) applied the method of Caballero and Lyons to industry-level data for U.K. manufacturing, and no evidence was found for increasing returns which were internal to the industry but there was a small external effect at the national level. Benaroch (1997) also used the same method to both national and provincial data for Canadian manufacturing and concluded that in Canada there are increasing returns that were primarily national rather than internal to the province or the industry.

In contrast, a recent paper by Burnside (1996) argued that these simple linear regressions may have been misleading for one of three reasons. First, most regression-based evidence is obtained by imposing cross-industry equality restrictions on parameters. Second, the choice of sample and instrument in general produces different results regarding the estimates of the extent of increasing returns. Third, correcting for cyclical variations in the utilization of capital has a significant impact on point estimates of returns to scale. Burnside then presented a richer set of alternative regressions, including considerations of sample and instrument as well as energy correction, and suggested that returns appear to be constant for the U.S. manufacturing industry.

The above studies seem to indicate that increasing returns to scale, if any, will be relatively weak in actual economies. Thus, the parameter value of $\gamma$ is empirically small and close to 1. In light of this, we compute the welfare cost for two small values of $\gamma$: $\gamma = 1.05$ and 1.10.

LITERATURE CITED


