DYNAMIC GENERAL EQUILIBRIUM
MODELS WITH IMPERFECTLY
COMPETITIVE PRODUCT MARKETS

Julio J. Rotemberg
Michael Woodford

DYNAMIC GENERAL EQUILIBRIUM
MODELS WITH IMPERFECTLY
COMPETITIVE PRODUCT MARKETS

ABSTRACT

This paper discusses the consequences of introducing imperfectly competitive product markets into an otherwise standard neoclassical growth model. We pay particular attention to the consequences of imperfect competition for the explanation of fluctuations in aggregate economic activity. Market structures considered include monopolistic competition, the "customer market" model of Phelps and Winter, and the implicit collusion model of Rotemberg and Saloner. Empirical evidence relevant to the numerical calibration of imperfectly competitive models is reviewed. The paper then analyzes the effects of imperfect competition upon the economy's response to several kinds of real shocks, including technology shocks, shocks to the level of government purchases, and shocks that change individual producers' degree of market power. It also discusses the role of imperfect competition in allowing for fluctuations due solely to self-fulfilling expectations.

Julio J. Rotemberg
Alfred P. Sloan School of Management
Room E52-432
Massachusetts Institute of Technology
Cambridge, MA 02139
and NBER

Michael Woodford
Department of Economics
University of Chicago
1126 East 59th Street
Chicago, IL 60637
and NBER

In this paper we discuss the consequences of introducing imperfectly competitive product markets in otherwise standard neoclassical growth models. We are particularly interested in the effects of imperfect competition on the way the economy responds at business cycle frequencies to various shocks. The literature on equilibrium modeling of aggregate fluctuations has mainly assumed perfectly competitive firms. While that represents an obvious starting point for analysis, we argue that there are important reasons for allowing competition to be imperfect.

One reason is that imperfect competition makes equilibrium possible in the presence of increasing returns technologies. [1] This increased flexibility in the specification of technology may be of great importance in modeling fluctuations, in particular for understanding cyclical variations in productivity. Empirical evidence on increasing returns is discussed in section 3 below. We argue below that not only is imperfect competition necessary if one wishes to assume increasing returns, but that conversely if market power is important one is virtually required to postulate increasing returns, in order to account for the absence of significant pure profits in economies like that of the U.S. Increasing returns may also make possible new sources of aggregate fluctuations – in particular, fluctuations due solely to self-fulfilling expectations.

Allowing for market power (and hence prices higher than marginal cost) and increasing returns may have important consequences for the interpretation of business cycles, although we do not argue for a particular theory of the cycle in this paper. In the real business cycle literature, technology shocks have usually been assigned a dominant role as the source of fluctuations. But the existence of market power and increasing returns implies that the Solow residual, interpreted in the RBC literature as a measure of exogenous technology shocks, contains an endogenous component. As Hall (1988, 1990) has emphasized this endogenous component may be unrelated to true changes in technology. For example, we show below in a complete dynamic model that increases in government purchases in the presence of imperfect competition result in a positive Solow residual, as well as an increase in output and hours. Thus, if one assumes that firms are perfectly competitive when they are not, one can be led to attribute to random technical progress a fraction of the increase in output which is in fact generated by increased government purchases, or other types of shocks that raise equilibrium output other than through an effect upon production possibilities.

---

[1] Increasing returns are compatible with competitive firms if the increasing returns are external, rather than internal to the firm, as in the model of Baxter and King (1990). We believe, however, that there is much more reason to believe that internal increasing returns are important.

Various empirical puzzles regarding the behavior of Solow residuals suggest that this endogeneity may be important, as we discuss in Section 4.

Imperfect competition also changes the predicted effects of technology shocks when they occur. As shown by Hornstein (1993), the ability of increases in productivity to stimulate increased employment is greatly reduced in the case of even rather moderate degrees of market power and increasing returns. We discuss this further in section 5 below. Hence if imperfect competition is important, it seems likely that other types of shocks will have to be assigned a major role in the explanation of business cycle variations in employment.

The hypothesis of imperfect competition does not in itself point one towards any particular alternative source of shocks. However, it does introduce new potential sources of employment fluctuations as well as providing a channel through which the importance of other shocks is increased. Imperfect competition implies not only that price generally differs from marginal cost, but is also consistent with variations over time in the gap between price and marginal cost. We show that an increase in this gap results in a reduction in the level of labor input that firms demand at a given real wage. The effect upon labor demand is thus similar to an adverse technology shock, while there need not be so large an offsetting wealth effect upon labor supply. Thus changes in the markup of price over marginal cost are potentially an important determinant of employment. We show that fluctuations in these markups, of a size that is not implausibly large given data on the U.S. economy, can generate employment variations of the size observed. [2]

One view of the source of markup fluctuations is that they result from exogenous changes in market structure - for example, in the context of the model of monopolistic competition developed in Sections 1-2 below, variations in the degree of substitutability between the differentiated goods. Under this view, imperfect competition introduces a new source of shocks. But it is also possible for markups to vary endogenously in response to aggregate shocks with no direct relation to market structure. In that case, the effects of markup variation become an additional channel through which such shocks can affect aggregate activity. As we will show, this channel may be particularly important in understanding cyclical variations in employment. We discuss several theories of endogenous variation in markups in Section 7, giving particular attention to theories in which it is possible for the markup to fall when there is an increase in aggregate demand. We

---

[2] In addition to the channels discussed below, it is worth pointing out that the assumption that producers have market power in their product markets is an essential element in models of nominal price rigidity. Such price rigidity enhances the role of monetary policy shocks as a source of aggregate fluctuations. For examples of completely specified dynamic model, see Hairault and Portier (1992) and Yun (1993).

then illustrate the potential importance of allowing for endogenous markup variation by showing how these theories affect the response of equilibrium employment to changes in government purchases.

Finally, the assumption of imperfect competition can lead to equilibrium aggregate fluctuations in the absence of any shocks at all. In standard real business cycle models, there exists a unique equilibrium (which corresponds to the solution to a planning problem). This equilibrium is necessarily independent of "sunspot" variables. But the introduction of imperfect competition implies that rational expectations equilibria no longer correspond to the solution of a planning problem, and indeed there can exist several distinct equilibria. Among these equilibria there may exist "sunspot" equilibria in which fluctuations result from self-fulfilling shifts in expectations. The quantitative plausibility of this possibility has been analyzed by Benhabib and Farmer (1992), Gali (1992), and Farmer and Guo (1993).

A further aim of this paper is to show how existing empirical studies using data at various levels of aggregation can be used to obtain estimates of the departures from perfect competition and from constant returns. While our survey of this literature is far from complete, we show how existing evidence bears upon the calibration of certain of the key parameters of imperfectly competitive models.

Finally, the paper shows that incorporating imperfect competition into equilibrium business cycle theory is easy. It is true that, because the resulting allocation is not Pareto optimal, it is not possible to compute the equilibrium by considering the solution to a planning problem. However, familiar methods for the computation of dynamic general equilibrium models, that make use of an Euler equation characterization of equilibrium (as discussed in detail in the next chapter) can also be applied when markets are not perfectly competitive.

Our paper proceeds as follows. In section 1, we first develop a basic imperfectly competitive model, in which firms are monopolistic competitors and there is no intertemporal aspect to their pricing problem. We show that, under quite general conditions, firms will choose to charge a price which represents a constant markup over marginal cost. We also discuss the connections between imperfect competition and increasing returns. In section 2 we embed this model of firm behavior in a complete dynamic general equilibrium model, and discuss the way in which the resulting model generalizes a standard real business cycle model. Section 3 provides a brief overview on the microeconomic evidence on the importance of imperfect competition. We then discuss the numerical solution of the response to shocks. In section 4, numerical results are presented for

shocks to the level of government purchases while section 5 concentrates on technology shocks. These sections illustrate the importance of taking account of imperfect competition. They also show how a comparison of the U.S. data and the model's responses can be used to obtain estimates of the quantitative importance of the departures from imperfect competition. Section 6 shows that for some parameter values (involving a sufficiently large degree of imperfect competition and increasing returns), the equilibrium response to shocks becomes indeterminate. This opens up a new potential source of aggregate fluctuations, namely, fluctuations due solely to self-fulfilling expectations, that may occur even in the absence of any stochastic disturbances to economic "fundamentals".

Sections 7 and 8 then discuss models with markup variations. In section 7, we consider the consequences of exogenous variations in the degree of market power (here modeled as due to exogenous changes in the substitutability of differentiated products) in a model of the kind treated in sections 4 and 5. We show that shocks of this kind can produce an overall pattern of co-movement of aggregate variables that captures several features of observed aggregate fluctuations, and that they are in particular able to produce large movements in employment. We estimate the size of markup fluctuations that would be required to account for observed business cycle variations in employment in the U.S., and compare this with independent calculations of the degree of cyclical variation in markups in the U.S. economy. In section 8, we discuss models in which endogenous variations in markups occur in response to other kinds of shocks. We develop two models in detail, the "customer markets" model of Phelps and Winter (1970), and the model of oligopolistic collusion used in our own previous work. In this section we also illustrate the predictions of these two models regarding the effects of changes in the level of government purchases. Here we particularly emphasize the ability of shocks other than technology shocks to produce variations in labor demand. We also briefly discuss the model of Gali (1992), in which the equilibrium markup depends on the composition of aggregate demand. In the context of this model we discuss the possibility of aggregate fluctuations in the absence of any shock at all.

1   The Behavior of Monopolistically Competitive Firms

We suppose that there exists a continuum of potentially producible differentiated goods indexed by the positive real line. At any point in time, only the subset whose index runs from zero to I_t is actually produced.

These goods are bought by consumers, the government and firms. The latter buy the differentiated goods both as materials that are used in current production and in the form of investment goods that increase the capital stock available for future production. To simplify the model, and to make it comparable to the standard perfectly competitive model, we assume that all of these ultimate demanders are interested in a single "composite good". In other words, the utility of consumers, the productivity of materials inputs, and the addition to the capital of firms depends only upon the number of units of the composite good that are purchased. An agent whose purchases of individual differentiated goods are described by a vector B_t obtains Q_t units of the composite good, where Q_t is given by

$$Q_t = f_t(B_t) \tag{1.1}$$

We assume that the aggregator f_t is an increasing, concave, symmetric and homogeneous degree one function of the measure B_t. By a symmetric function we mean that its value is unchanged if one exchanges the quantities purchased of any of the individual goods, so that the value Q_t depends only upon the *distribution* of quantities purchased of individual goods, and not upon the identities of the goods purchased. The aggregator is the same for all of the purposes mentioned above; we allow, however, for variation over time, as the set of differentiated goods being produced changes. The producer of each of the differentiated goods sets a price for it; the collection of these prices describes a price vector P_t conformable with the vector of goods purchases. Consider an agent (be it a consumer, government or firm) wishing to buy G_t units of the composite good. The agent will distribute its purchases over the various differentiated goods so as to minimize the total cost $< P_t, B_t >$ of obtaining G_t. Because f_t is homogeneous of degree one, this cost-minimizing demand is equal to G_t times a homogeneous degree zero vector-valued function of the price vector:

$$B_t = G_t \Delta_t(P_t)$$

Hence for a given vector of prices P_t, all agents will choose scalar multiples of the same vector $\Delta(P_t)$. This allows us to aggregate the demands of all types to obtain

$$B_t = Q_t \Delta_t(P_t)$$

where Q_t denotes total demand for the composite good for all purposes.

Furthermore, because f_t is symmetric, the component $\Delta_t^i(P_t)$ indicating purchases of goods i must depend only upon the price p_t^i charged for that good and the overall distribution of prices charged. We will

6

be concerned here only with symmetric equilibria. We will thus consider situations where all firms (with the possible exception of firm i) charge a price of p_t at t while firm i charges p_t^i. In this case $\Delta_t^i(P_t)$ can be written in the form $D_t(p_t^i/p_t)/I_t$, where D_t is a decreasing function, the same for all i, and I_t denotes the number of goods produced at date t. D_t depends only on the ratio of the two prices because Δ is homogeneous of degree zero. (The normalization by I_t is simply for convenience.) The demand for firm i's product is then given by

$$d_t^i = \frac{Q_t}{I_t} D_t(\frac{p_t^i}{p_t}) \tag{1.2}$$

By monopolistic competition we mean that each firm i takes as given aggregate demand Q_t and the price p_t charged by other firms, and chooses its own price, p_t^i, taking into account the effect of p_t^i on its sales indicated by (1.2). [3] We are therefore interested in the properties of the demand curve D_t.

Let us assume as a normalization of f_t in each period that $f_t(M) = I_t$ for all t, where M is a vector of ones (or, in the continuum case, the uniform measure). Then since f_t is symmetric, one must have $D_t(1) = 1$ for all t. We furthermore assume that D_t is differentiable at 1, and that the value $D_t'(1) < -1$ is similarly independent of t. The latter assumption means that the degree of substitutability between different differentiated goods, evaluated in the case of equal purchases of all goods, remains the same as additional differentiated goods are added, and the common elasticity of substitution is greater than one. Finally, we assume that for each t, $D(\rho) + \rho D'(\rho)$ is a monotonically decreasing function of the relative price ρ over the entire range of relative prices for which it is positive. This implies the existence of a downward sloping marginal revenue curve for each producer of a differentiated good. The result of these assumptions is that, at a symmetric equilibrium, firms face a time-invariant elasticity of demand. It is worth stressing that we have obtained this result without having to make global assumptions on preferences, as is done for instance in Dixit and Stiglitz (1977). [4]

---

[3]This equilibrium concept is an obvious one, given that we eventually wish to assume the existence of a continuum of goods in each period, identified with an interval $[0, I_t]$ on the real line. In the continuum limit, the price charged for an individual good obviously has no effect upon any agent's intertemporal allocation of total expenditure, and so has no effect upon the equilibrium processes $\{p_t, Q_t\}$. Even in the case of a finite number of goods, we can *define* equilibrium in this way, just as one can define Walrasian equilibrium for an economy with a finite number of traders. However, the equilibrium concept only becomes compelling as a formal representation of the outcome of competition in the limit of an infinite number of traders. For rigorous development of this equilibrium concept for the finite case, see Benassy (1991, sec. 6.4) and references cited therein. It has become common in the literature in macroeconomics and international trade to assume a continuum of differentiated goods, especially when, as here, one wishes to treat the number of goods as an endogenous variable.

[4]In their model $f_t(B_t)$ is equal to $I_t^{-\frac{\sigma}{1-\sigma}} \left[\sum_{i=1}^{I_t} (b_t^i)^{1-\sigma}\right]^{\frac{1}{1-\sigma}}$ where $0 < \sigma < 1$. They thus assume a globally constant elasticity of substitution equal to $1/\sigma$. This means that $D_t(\rho)$ is equal $\rho^{-\frac{1}{\sigma}}$, so that $D_t'(1) = -1/\sigma$ for all t. We also observe that the condition of decreasing marginal revenue is satisfied for all ρ. However, a globally constant elasticity of substitution

7

We now turn to a discussion of the technology with which firms produce output. When considering an economy with imperfectly competitive firms, it no longer makes sense to assume the kind of technology specification that is standard in the real business cycle literature. First of all, it is common in that literature to assume a production function using only capital and labor inputs, ignoring produced materials. In the case of perfectly competitive firms, this involves no loss of generality, as the output measure that one is concerned with is total value added (the total product net of the value of materials inputs), and this can indeed be expressed as a function of capital and labor inputs. Suppose that the production function for good i is given by

$$q_t^i = G(K_t^i, z_t H_t^i, M_t^i) \tag{1.3}$$

where z_t is an index of labor augmenting technical progress at t, while q_t^i denotes the output, K_t^i the level of capital services, H_t^i the hours employed, and M_t^i the materials inputs of firm i in period t. In a symmetric equilibrium, the price of materials is the same as the price for output, so that value added is given by $q_t^i - M_t^i$. In an equilibrium with perfect competition, this will necessarily equal

$$F(K_t^i, z_t H_t^i) = \max_{M_t^i}[G(K_t^i, z_t H_t^i, M_t^i) - M_t^i] \tag{1.4}$$

This follows from profit maximization by price-taking firms. Thus there exists a pseudo-production function for value added with only capital and labor inputs as its arguments. Furthermore, by the envelope theorem, the derivatives of F with respect to its two arguments equal the marginal products of those two factors, that must be equated to their prices (deflated by the price of output) in equilibrium. Thus one obtains correct equilibrium conditions if one simply treats F as the true production function. This is implicitly what is being done in the real business cycle literature (as in other common growth models).

But with imperfect competition, (1.4) is no longer correct. A monopolistically competitive firm will choose its materials inputs so as to maximize

$$\frac{p_t^i}{p_t}G(K_t^i, z_t H_t^i, M_t^i) - M_t^i \tag{1.5}$$

given the relation (1.2) between its sales and its price. Because p_t^i is not independent of the quantity sold, materials inputs are not chosen as in (1.4), even though in a symmetric equilibrium $p_t^i = p_t$. If (1.3) is a

of this kind is nowhere essential to our conclusions, which depend only on the assumption that the elasticity of demand in the case of uniform prices is independent of scale.

smooth neoclassical production function, the first-order condition for choice of materials inputs implies that in a symmetric equilibrium,

$$G_M(K_t^i, z_t H_t^i, M_t^i) = [1 + D'(1)^{-1}]^{-1}$$

so that the use of materials inputs depends upon the degree of market power. Of course, we can still solve this equation for M_t^i as a function of $(K_t^i, z_t H_t^i)$, and so obtain an expression for value added as a function of those two inputs alone. But apart from the fact that this function would not represent the economy's true production possibilities, it would also change in the case of changes in the degree of market power. In general, in the presence of imperfect competition the economy will be strictly inside its production possibilities frontier (and not simply at an inefficient point on it given preferences over consumption and leisure), and the degree to which it is inside will depend upon the degree of market power. [5]

This complication can be avoided if one assumes a fixed-coefficient technology as far as materials inputs are concerned. Suppose that (1.3) takes the form

$$q_t^i = \min\Big[\frac{V(K_t^i, z_t H_t^i)}{1 - s_M}, \frac{M_t^i}{s_M}\Big]$$

Here $0 < s_M < 1$ corresponds to the share of materials costs in the value of gross output (in a symmetric equilibrium). In this case firm i will always choose materials inputs $M_t^i = s_M q_t^i$ regardless of the degree of market power, and there exists a production function for value added that is independent of the degree of market power, namely $V(K_t^i, z_t H_t^i)$. We will in fact assume a production function of this form in what follows, insofar as it seems realistic to assume that opportunities to substitute capital or labor inputs for materials are relatively small; but in so doing we neglect effects that may actually be of importance.

With imperfect competition, materials inputs matter also for another reason and they continue to matter for this reason even in the case of a fixed-coefficients production function. With imperfect competition, materials inputs affect the size of the wedge between the marginal product of labor and the wage. For a given degree of market power (i.e., a given slope $D'(1)$), and hence a given markup of price over marginal cost, this wedge is greater the larger the share of materials. Hence taking account of materials inputs is important when we wish to calibrate our model on the basis of evidence about typical degrees of market power.

---

[5] For further development of this point, see Basu (1992).

9

To see this, note that in the case of the fixed-coefficients production function, capital and labor inputs are chosen to maximize

$$\frac{p_t^i}{p_t} q_t^i - w_t H_t^i - r_t K_t^i - s_M q_t^i$$

where w_t represents the wage deflated by the price of the composite good and r_t the rental price of capital deflated in the same way, again given (1.2). (Because the present paper is concerned solely with imperfectly competitive product markets, we assume that firms are price-takers in factor markets.) In a symmetric equilibrium, the first order conditions for factor demands take the forms

$$[1 + D'(1)^{-1} - s_M] V_1(K_t^i, z_t H_t^i) = [1 - s_M] r_t \tag{1.6a}$$

$$[1 + D'(1)^{-1} - s_M] z_t V_2(K_t^i, z_t H_t^i) = [1 - s_M] w_t \tag{1.6b}$$

We assume that

$$s_M < 1 + D'(1)^{-1}$$

so that a symmetric equilibrium of this kind is possible. (We also defer discussion of second order conditions for the moment.) The wedge between marginal products and factor prices is observed to be

$$\mu = \frac{1 - s_M}{1 - s_M + D'(1)^{-1}} > 1$$

This is a monotonic function of $(1 - s_M) D'(1)$, higher when the latter is a smaller negative quantity. Hence $s_M > 0$ has an effect equivalent to making each firm's degree of market power higher.

We now consider the relation between price and marginal cost in a monopolistically competitive equilibrium. In period t, each firm's marginal cost of production (using the composite good as numeraire) is

$$(1 - s_M) \frac{w_t}{z_t V_2(K_t^i, z_t H_t^i)} + s_M$$

Comparing this with (1.6b), we observe that each firm's markup (ratio of price to marginal cost) will equal

$$\gamma = [1 + D'(1)^{-1}]^{-1} > 1 \tag{1.7}$$

Note that in this simple model, the markup is a constant, regardless of any changes that may occur in equilibrium output due to technology shocks or for other reasons. This conclusion depends crucially on the homogeneity of the aggregator function f as well as upon the assumption of monopolistically competitive

10

behavior on the part of firms. Without homogeneity, each firm's markup can depend on the level of output itself. But abandoning homogeneity also makes the aggregation of different buyers' demands more complicated. For this reason, our entire analysis is conducted with a homogeneous f. We do consider departures from monopolistic competition in section 8 which result in endogenous variations in markups.

If $s_M > 0$, the inefficiency wedge μ is larger than the individual firm's markup γ, although the two are closely related. The former can be thought of as the ratio between the price of value added (the difference between the price of output and the cost of materials) and its marginal cost. This is larger than the ratio of price to marginal cost γ because firms mark up their materials inputs as well. For a given value of s_M, μ is a monotonically increasing function of γ, given by

$$\mu = \frac{(1 - s_M)\gamma}{1 - s_M \gamma} \tag{1.8}$$

This relation is important when we consider below the effects of variations in market power. (Recall that s_M is a parameter of the production technology, rather than an endogenous variable.) It is also important for understanding the relation between different measures of the degree of market power found in the empirical literature, as we discuss further in section 3.

We now discuss the relationship between the inefficiency wedge introduced by market power and the degree of returns to scale. We mentioned earlier that imperfect competition makes equilibrium possible even in the case of an increasing returns technology, so that we have more flexibility in the specification of V. In fact, once we have assumed market power, assuming increasing returns is not only possible but also more reasonable. In the case of a constant returns production function (V homogeneous of degree one), Euler's theorem together with (1.6a)-(1.6b) implies that

$$q_t^i - M_t^i = \mu(r_t K_t^i + w_t H_t^i) \tag{1.9}$$

Hence $\mu > 1$ implies pure profits (the value of output exceeds the sum of materials costs, capital costs, and labor costs). Yet studies of U.S. industry generally find evidence of little if any pure profits on average. (We discuss this further in section 3.) Hence if market power is significant (as evidence discussed below suggests), we must conclude that there do not exist constant returns.

Let us use as a measure of returns to scale in the production of value added the quantity

$$\eta_t^i = \frac{V_1(K_t^i, z_t H_t^i)K_t^i + V_2(K_t^i, z_t H_t^i)z_t H_t^i}{V(K_t^i, z_t H_t^i)} \tag{1.10}$$

11

Here increasing, decreasing or constant returns correspond to η_t^i greater than, less than, or equal to one. Note that this is a purely local measure that may vary over time as production varies; in the case that V is homogeneous of some degree, then η^i is a constant and corresponds to that degree of homogeneity. Note also that this is a measure of the *short run* returns to scale associated with changes in the factors employed while the number of differentiated goods being produced remains fixed; it has no implication regarding the *long run* returns to scale in a growing economy with growth in the number of types of goods being produced as well as in the total quantity of factors employed. Finally, note that this is a measure of increasing returns in the production of *value added* rather than of gross output. A standard (local) measure of increasing returns in the production of gross output would instead be the ratio of average to marginal cost for firm i, *i.e.*,

$$
\begin{aligned}
\rho_t^i &= \frac{r_t K_t^i + w_t H_t^i + M_t^i}{[(1-s_M)\frac{w_t}{z_t V_2(K_t^i, z_t H_t^i)} + s_M]q_t^i} \\
&= \frac{\eta_t^i \mu^{-1}(1-s_M) + s_M}{\mu^{-1}(1-s_M) + s_M}
\end{aligned}
$$

Note that if $s_M > 0$, the measure η_t^i will be larger than ρ_t^i, though the two coincide when there are no intermediate inputs. [6]

Abandoning the assumption of short run constant returns ($\eta = 1$), we find instead of (1.9) the more general result

$$
\eta_t^i(q_t^i - M_t^i) = \mu(r_t K_t^i + w_t H_t^i)
$$

Hence zero pure profits on average are consistent with $\mu > 1$ if the average returns to scale are $\eta = \mu > 1$. Thus increasing returns (in the sense just explained) are required. Note that $\eta > 1$ also means $\rho > 1$, so that there are also increasing returns in the production of gross output, of a magnitude

$$
\rho_t^i = [\mu^{-1}(1-s_M) + s_M]^{-1} = \gamma
$$

The intuition is simple. With increasing returns, average cost exceeds marginal cost so that price can be equal to average cost and profits be zero even if price exceeds marginal cost. It is important that these increasing returns be internal to the firm, rather than due to externalities of the kind postulated by Baxter and King (1990). Even if there are also external returns to scale the firm will make positive profits unless its own average costs exceed its own marginal cost.

12

The existence of increasing returns of this size is furthermore not merely fortuitous, but follows from economic principles, as Chamberlin (1933) pointed out. If in any period $\eta_t^i \neq \mu$, then there are non-zero pure profits. This is consistent with profit-maximization, assuming that the number of differentiated goods being produced, and the identities of their producers, cannot change. (Negative short-run profits are possible if it is assumed that it is not possible to simply shut down at zero cost, perhaps because fixed costs have been paid in advance.) But it does not make sense that such a situation should persist. Recall that we have assumed that the aggregator f_t depends upon the number of goods sold I_t in such a way that the elasticity of demand for each good $-D'(1)$ remains the same after I_t changes. With this assumption, an entrepreneur that produces an additional product earns the same profits on the new product as the profits earned on all existing products. The reason is that his optimal markup is the same as that of all existing firms and their optimal markup does not change either. Hence sustained positive profits in the production of existing goods give entrepreneurs a reason to introduce new goods. Sustained negative profits should correspondingly eventually lead to exit of some producers. Hence it is reasonable to assume that in the long run, profits return to the level zero, due to adjustment in the number of produced goods I_t.

A simple specification for the production function (1.3) that makes this possible is

$$q_t^i = \min\left[\frac{F(K_t^i, z_t H_t^i) - \Phi}{1 - s_M}, \frac{M_t^i}{s_M}\right] \tag{1.11}$$

where F is assumed to be homogeneous degree one, and $\Phi > 0$ indicates the presence of fixed costs. [7] This is a production function in which marginal cost is independent of scale, but average cost is decreasing due to the existence of the fixed costs; it generalizes the specification that is standard in the equilibrium business cycle literature in a way that involves the introduction of only one new parameter. In the case of production function specification (1.11), the index of increasing returns is given by

$$\eta_t^i = \frac{Y_t^i + \Phi}{Y_t^i}$$

again using Euler's theorem, where Y_t^i denotes value added in industry i in period t. Profits become zero in the long run if Y_t^i in each industry tends toward a particular value, namely $\Phi/(\mu - 1)$. This in turn can be brought about by having the right number of differentiated goods I_t relative to the aggregate quantity of

---

[7] Alternative specifications are possible. In particular, one can assume that the fixed costs take the form of some fixed amount of labor, or capital. The current specification assumes that both labor and capital can be used as fixed costs and that the proportions in which they are employed for this purpose depend on factor prices.

capital and labor inputs used in production; specifically, we require that in the long run

$$\frac{I_t}{F(K_t, z_t H_t)} = \frac{\mu - 1}{\mu \Phi} \tag{1.12}$$

where K_t and H_t denote aggregate quantities of capital and labor inputs respectively. This condition requires that the steady state number of differentiated goods grows at the same rate as the capital stock, the quantity of effective labor inputs and aggregate output. [8]

Let us now collect our equations describing aggregates in a symmetric equilibrium. Aggregate value added is determined by aggregate factor inputs through the relation

$$Y_t = F(K_t, z_t H_t) - \Phi I_t \tag{1.13}$$

(This is of course the output concept for which we have aggregate data.) Aggregate factor demands are related to factor prices through the relations

$$F_1(K_t, z_t H_t) = \mu r_t \tag{1.14a}$$

$$z_t F_2(K_t, z_t H_t) = \mu w_t \tag{1.14b}$$

where the inefficiency wedge μ is determined by (1.7)-(1.8). Because our interest in this paper is in the short run effects of shocks, we will treat both of the processes $\{z_t\}$ and $\{I_t\}$ as exogenous, even though we recognize that in the long run macroeconomic conditions may affect both technical progress (for reasons stressed in the literature on endogenous growth) and the number of differentiated products that are produced (for reasons sketched above). Our specification of these exogenous processes, however, will be such as to imply that our equilibria will involve only transitory deviations from a scale of operations at which (1.12) holds.

These equations jointly describe the production side of our model. Note that in the case that $\mu = 1$ and $\Phi = 0$ these are the equilibrium conditions of a standard real business cycle model. Hence our specification nests a standard competitive model as a limiting case.

---

[8] In this formulation, the output of each firm stays constant in the steady state, the entire growth in output is accounted for by growth in firms. An alternative formulation that preserves a constant steady state markup and degree of increasing returns is to assume as we did in Rotemberg and Woodford (1992) that

$$q_t^i = \min\left[\frac{F(K_t^i, z_t H_t^i) - \Phi_t}{1 - s_M}, \frac{M_t^i}{s_M} \right]$$

and that Φ_t has the same steady state growth rate as z_t and hence the same growth rate as K_t and Q_1. In this case, an equilibrium with a constant steady state markup has no growth in the number of firms. The entire growth in output is reflected in growth in each firm's output. One could probably also construct intermediate models with constant steady state markups where the number of firms as well as output per firm grow over time.

14

2 A Complete Dynamic Equilibrium Model with Monopolistic Competition

The economy also contains a large number of identical infinite-lived households. At time t, the representative household seeks to maximize

$$E_t \left\{ \sum_{\tau=0}^{\infty} \beta^\tau N_{t+\tau} u(c_{t+\tau}, h_{t+\tau}) \right\} \tag{2.1}$$

where E_t takes expectations at time t, β denotes a constant positive discount factor, N_t denotes the number of members per household in period t, c_t denotes per capita consumption by the members of the household in period t, and h_t denotes per capita hours worked by members of the household in period t. By normalizing the number of households at one, we can use N_t also to represent the total population, $C_t = N_t c_t$ to denote aggregate consumption, and so on. We assume, as usual, that u is a concave function, increasing in its first argument, and decreasing in its second argument. (The class of utility functions is further specialized below.)

Rather than make assumptions about the parameters of u directly, it turns out to be more convenient to make assumptions about the Frisch demand functions for consumption and leisure. These Frisch demand functions depend on the marginal utility of wealth λ_t which is given by

$$\lambda_t \equiv u_1(c_t, h_t) \tag{2.2}$$

Assuming that households can freely sell their labor services for the real wage w_t, they must satisfy the first order condition

$$\frac{u_2(c_t, h_t)}{u_1(c_t, h_t)} = w_t = \frac{u_2(c_t, h_t)}{\lambda_t} \tag{2.3}$$

where the second equality follows from the definition (2.2). Combining (2.2) with the second equality of (2.3), we can solve for c_t and h_t as functions of λ_t and w_t which gives the Frisch demand curves

$$c_t = c(w_t, \lambda_t) \tag{2.4a}$$

$$h_t = h(w_t, \lambda_t) \tag{2.4b}$$

One advantage of using the Frisch demand functions is that the effects of all future variables (and their expectations) on current choices is captured by the marginal utility of wealth λ_t. In terms of these functions, the condition for market clearing in the labor market is

$$H_t = N_t h(w_t, \lambda_t) \tag{2.5}$$

15

while that for the product market is

$$Y_t = N_t c(w_t, \lambda_t) + [K_{t+1} - (1 - \delta)K_t] + G_t \tag{2.6}$$

where G_t represents government purchases of produced goods, and δ is the constant rate of depreciation of the capital stock, satisfying $0 < \delta \leq 1$. Equation (2.6) with Y_t representing value added is the standard GNP accounting identity, except that we do not count value added by the government sector as part of either G_t or Y_t. This equation says that one unit of the composite good at t can be used to obtain one unit of capital at $t + 1$. It follows that the purchase price of capital, just like that of materials, is equal to one. We assume that households, like firms, have access to a complete set of frictionless securities markets. As a result, the expected returns to capital must satisfy the asset-pricing equation

$$1 = \beta E_t \left\{ \left(\frac{\lambda_{t+1}}{\lambda_t} \right) \left[r_{t+1} + (1 - \delta) \right] \right\}$$

Substituting (1.14a), we obtain

$$1 = \beta E_t \left\{ \left(\frac{\lambda_{t+1}}{\lambda_t} \right) \left[\frac{F_1(K_{t+1}, z_{t+1} H_{t+1})}{\mu} + (1 - \delta) \right] \right\} \tag{2.7}$$

A rational expectations equilibrium is a set of stochastic processes for the endogenous variables $\{Y_t, K_t, H_t, w_t, \lambda_t\}$ that satisfy (1.13), (1.14b) and (2.5) - (2.7), given the exogenous processes $\{G_t, z_t, N_t, I_t\}$. [9] We analyze the response of this model to changes in z_t and government purchases using essentially the method of King, Plosser, and Rebelo (1988a). This involves restricting our attention to the case of small stationary fluctuations of the endogenous variables around a steady state growth path. Let us first consider the conditions under which stationary solutions to these equations are possible. Given the existence of trend growth in the exogenous variables, the equilibrium requires that variables such as $\{Y_t, w_t, ...\}$ exhibit trend growth as well. However, a stationary solution for transformed (detrended) variables can exist if the equilibrium conditions in terms of these transformed variables do not involve any of the trending exogenous variables (z_t or N_t as opposed to their growth rates).

As in King, Plosser and Rebelo (1988a), this requires only that there exists a $\sigma > 0$ such that $h(w, \lambda)$ is homogeneous of degree zero in $(w, \lambda^{\frac{-1}{\sigma}})$, and $c(w, \lambda)$ is homogeneous of degree one in $(w, \lambda^{\frac{-1}{\sigma}})$. Given these

---

[9] The variables Y_t, H_t, w_t, λ_t must be measurable with respect to information available at time t, while K_t must be measurable with respect to information available at time $t - 1$. Information available at time t consists of the realizations at time t or earlier of the variables G_t, z_t, N_t, I_t. In section 6, we allow the information set to also contain "sunspot" variables.

conditions, there exists an equilibrium in which the detrended endogenous state variables

$$\tilde{Y}_t = \frac{Y_t}{z_t N_t}, \qquad \tilde{K}_t = \frac{K_t}{z_{t-1} N_t}, \qquad \tilde{H}_t = \frac{H_t}{N_t}$$

$$\tilde{w}_t = \frac{w_t}{z_t}, \qquad \tilde{\lambda}_t = \lambda_t (z_t)^\sigma$$

are stationary, given stationary processes for the exogenous random variables

$$\tilde{G}_t = \frac{Y_t}{z_t N_t}, \qquad \tilde{I}_t = \frac{I_t}{z_t N_t}, \qquad \gamma_t^z = \frac{z_t}{z_{t-1}}$$

and a constant population growth rate so that $\frac{N_{t+1}}{N_t} = \gamma^N$ in all periods. The equilibrium conditions in terms of the stationary variables become

$$\tilde{Y}_t = F\left((\gamma_t^z)^{-1} \tilde{K}_t, \tilde{H}_t \right) - \Phi \tilde{I}_t \tag{2.9}$$

$$F_2\left(\frac{\tilde{K}_t}{\gamma_t^z}, \tilde{H}_t \right) = \mu \tilde{w}_t \tag{2.10}$$

$$\tilde{H}_t = H(\tilde{w}_t, \tilde{\lambda}_t) \tag{2.11}$$

$$c(\tilde{w}_t, \tilde{\lambda}_t) + [\gamma^N \tilde{K}_{t+1} - (1-\delta)(\gamma_t^z)^{-1} \tilde{K}_t] + \tilde{G}_t = \tilde{Y}_t \tag{2.12}$$

$$1 = \beta E_t \left\{ (\gamma_{t+1}^z)^{-\sigma} \left(\frac{\tilde{\lambda}_{t+1}}{\tilde{\lambda}_t} \right) \left[\frac{F_K((\gamma_{t+1}^z)^{-1} \tilde{K}_{t+1}, \tilde{H}_{t+1})}{\mu} + (1-\delta) \right] \right\} \tag{2.13}$$

These equilibrium conditions involve only the detrended state variables, and so admit a stationary solution in terms of those variables. [10]

Like King, Plosser and Rebelo, we furthermore seek to characterize such a stationary equilibrium only in the case of small fluctuations of the detrended state variables around their steady state values, i.e., the constant values that they take in a deterministic equilibrium growth path in the case that γ_t^z, $\tilde{I}_t$ and $\tilde{G}_t$ are constant. This steady state can be found by solving the 5 equations (2.9)-(2.13) for the five unknowns $\tilde{Y}$, $\tilde{H}$, $\tilde{w}$, $\tilde{\lambda}$ and $\tilde{K}$. Since these solutions vary continuously with $\tilde{I}$, we can then find a value for $\tilde{I}$ for which (1.12) holds and profits are zero. At this solution, $\tilde{I}\Phi$ equals $(\mu - 1)\tilde{Y}$.

Given a steady state, we approximate a stationary equilibrium involving small fluctuations around it by the solution to a log-linear approximation to the equilibrium conditions. This linearization uses derivatives

---

[10] Note that equilibrium conditions in terms of stationary variables can be obtained with this technique even if the model has multiple steady state equilibria.

evaluated at the steady-state values of the state variables. [11] In writing the log-linear equations, we use the notation $\hat{Y}_t$ for $\log(\tilde{Y}_t/\tilde{Y})$, $\hat{w}_t$ for $\log(\tilde{w}_t/\tilde{w})$, and so on, where the $\tilde{Y}, \tilde{w}$ denote steady state values. The log-linear approximation to the Frisch consumption demand and labor supply functions (2.4) can be written

$$\hat{C}_t = \epsilon_{Cw}\hat{w}_t + \epsilon_{C\lambda}\hat{\lambda}_t \tag{2.14a}$$

$$\hat{H}_t = \epsilon_{Hw}\hat{w}_t + \epsilon_{H\lambda}\hat{\lambda}_t \tag{2.14b}$$

where the coefficients represent the elasticities of the Frisch demands. Thus the only way in which the specification of preferences affects our equilibrium conditions is in the values implied for these elasticities, and so we calibrate the model by specifying numerical values for these elasticities directly. Our homogeneity assumptions furthermore imply that

$$\epsilon_{Hw} - \sigma\epsilon_{H\lambda} = 0 \tag{2.8a}$$

$$\epsilon_{Cw} - \sigma\epsilon_{C\lambda} = 1 \tag{2.8b}$$

$$\frac{\epsilon_{Cw}}{\epsilon_{Hw}} = \frac{1-\sigma}{\sigma}\frac{\tilde{w}\tilde{H}}{\tilde{C}} \tag{2.8c}$$

The three restrictions (2.8a), (2.8b) and (2.8c) imply that there are only two independent parameters among ϵ_{Cw}, $\epsilon_{C\lambda}$, ϵ_{Hw}, $\epsilon_{H\lambda}$, and σ. To preserve comparability with earlier studies we calibrate σ and ϵ_{Hw}. The former is the inverse of the intertemporal elasticity of consumption growth holding hours worked constant while the latter is the intertemporal elasticity of labor supply.

The log-linearized equilibrium conditions can then be written as

$$\hat{Y}_t = \mu s_K \hat{K}_t + \mu s_H \hat{H}_t - \mu s_K \gamma_t^z - (\mu - 1)\hat{I}_t \tag{2.15}$$

$$\frac{s_K}{\epsilon_{KH}}(\hat{K}_t - \hat{\gamma}_t^z - \hat{H}_t) = \hat{w}_t \tag{2.16}$$

$$\hat{H}_t = \epsilon_{Hw}\hat{w}_t + \epsilon_{H\lambda}\hat{\lambda}_t \tag{2.17}$$

$$s_C[\epsilon_{Cw}\hat{w}_t + \epsilon_{C\lambda}\hat{\lambda}_t] + s_I[\hat{\gamma}_t^z + \left(\frac{1+g}{g+\delta}\right)\hat{K}_{t+1} - \left(\frac{1-\delta}{g+\delta}\right)(\hat{K}_t - \hat{\gamma}_t^z)] + s_G\hat{G}_t = \hat{Y}_t \tag{2.18}$$

$$-\sigma\hat{\gamma}_t^z + E_t\left\{\hat{\lambda}_{t+1} - \hat{\lambda}_t + \left(\frac{r+\delta}{1+r}\right)\frac{s_H}{\epsilon_{KH}}(\hat{H}_{t+1} - \hat{K}_{t+1} + \hat{\gamma}_{t+1}^z)\right\} = 0 \tag{2.19}$$

---

[11] The method is the same as in King, Plosser and Rebelo. This can be made rigorous, and justified as an application of a generalized implicit function theorem, as shown in Woodford (1986). It should be understood that when we refer to small fluctuations around the steady state values, we have in mind stationary random variables with a sufficiently small bounded support.

18

This system may be further simplified as follows. We can solve (2.15), (2.16) and (2.17) for $\hat{H}_t$, $\hat{Y}_t$ and $\hat{w}_t$ as functions of $\hat{K}_t$ and $\hat{\lambda}_t$. Substitution of these solutions into (2.18) and (2.19) gives a system of two difference equations of the form

$$A \begin{pmatrix} E_t \hat{\lambda}_{t+1} \\ \hat{K}_{t+1} \end{pmatrix} = B \begin{pmatrix} \hat{\lambda}_t \\ \hat{K}_t \end{pmatrix} + C \begin{pmatrix} \hat{\gamma}_t^z \\ \hat{G}_t \\ \hat{I}_t \end{pmatrix} + D \begin{pmatrix} E_t \hat{\gamma}_{t+1}^z \\ E_t \hat{G}_{t+1} \\ E_t \hat{I}_{t+1} \end{pmatrix} \tag{2.20}$$

We assume that the exogenous variables $\{\gamma_t^z, \hat{G}_t, \hat{I}_t\}$ are subject to stationary fluctuations. Thus, as shown for example by Blanchard and Kahn (1980), (2.20) has a unique stationary solution if and only if the matrix A is non-singular, and the matrix $A^{-1}B$ has one eigenvalue with modulus less than one and one with modulus greater than one. Woodford (1986) shows that this is also the case in which the original nonlinear equilibrium conditions have a locally unique stationary solution. Moreover, when perturbed by exogenous shocks whose support is sufficiently small, this solution involves only small fluctuations around the steady state. For the calibrated parameter values discussed in the next section, and for all sufficiently nearby values, we find that there is indeed exactly one stable eigenvalue. We thus focus much of our analysis on this case. In section 6, we discuss alternative parameter values that lead both eigenvalues to be smaller than one.

In the case where there is only one stable eigenvalue, there is a unique equilibrium response to the shocks with which we are concerned. We can approximate this unique response by calculating the solution to the log-linear system (2.20) using the formulae of Blanchard and Kahn (1980) or Hansen and Sargent (1980). The resulting solution is a linear function of the exogenous variables. This means that we can decompose fluctuations in the state variables into the contributions from each of the shocks that affect our exogenous variables. It also means that the analysis of the effect of any one shock would not be affected by the inclusion of additional exogenous shocks.

The coefficients in the log-linear equation system (2.15)-(2.19) have been written in terms of parameters presented in Table 1. Column 2 of the table gives the formulas which, when evaluated at the steady state values of the detrended state variables, allow us to compute the value of these parameters. With the exception of population growth and the parameters related to the lack of perfect competition, the values we have assigned follow those presented in King, Plosser and Rebelo (1988a). Note that, in the case where μ is equal to one the model we have presented reduces to the standard real business cycle model. We survey some of the evidence relating to values for the average inefficiency wedge μ (for now taken to be a constant) in the next section. The same evidence is equally relevant for calibration of the steady state markup in the

19

case of the variable-markup models discussed later.

3 Evidence on the Size of Markups and Increasing Returns

Evidence on the size of markups and increasing returns comes from several distinct sources. First, there is a literature that attempts to measure the degree of increasing returns from engineering studies of the average costs of different plants. Most of this literature, which is summarized in Panzar (1989) has concerned itself with returns to scale in regulated industries. The returns to scale found in the telecommunications industry tend to be substantial, with most studies finding returns to scale, which correspond to a value of η in (1.10) of the order of 1.4. Those found for electric power generation seem to be somewhat sensitive to the exact specification. Christensen and Greene (1976) found that only half the 1970 plants were operating at a scale where marginal cost was below average cost. By contrast, Chappell and Wilder (1986) found much more substantial returns to scale when taking into account the multiplicity of outputs of many electric utilities.

These findings are of only limited relevance to our analysis. These studies seek to measure the degree of long run returns to scale, i.e.,, the rate at which average costs decline as one goes from a small plant to a larger one. However, constant returns in this sense is perfectly consistent with large gaps between short run average costs and short run marginal costs. This would happen in particular if plant size exceeds the size that minimizes costs. Firms would rationally make such capacity choices if, for instance, producing at an additional location or introducing an additional variety raises a firm's sales for any given price, as in the Chamberlinian model of monopolistic competition. This is in essence what occurs in our model. We have a fixed cost per plant so that lowest average cost would be obtained by having a single plant. In equilibrium there are several plants, all of which have the same average cost and, nonetheless, marginal cost is below average cost.

Second, there is a literature which attempts to measure the elasticity of demand facing individual products produced by particular firms. This literature is relevant because, as is clear from the derivation of (1.7), it is never profit maximizing to set the markup γ lower than one over one plus the inverse of the elasticity of demand for the product. There are many estimates of the elasticity of demand for particular products in the marketing literature. Tellis (1988) surveys this literature, and reports that the median measured price elasticity is just under 2. This suggests that the markup γ of these individual firms would equal 2 if they behaved

20

like monopolistic competitors. If our symmetric model where correct, the demand elasticity estimated in the marketing literature would correspond to the elasticity of demand faced by the typical firm. In practice, elasticities of demand probably differ across products and the elasticity of demand of those products studied in the marketing literature is probably atypically low. This is because the marketing literature focuses on the demand for branded consumer products which are more differentiated than unbranded products so that their demand is probably less price sensitive. Thus, the typical product in the economy probably has a price elasticity of demand that exceeds 2.

Finally, there is a literature which tries to obtain econometric estimates of marginal cost and, in some cases, combine them with econometric estimates of the elasticity of demand. The aim of this approach is to obtain simultaneous, independent estimates of the markup and of the degree of increasing returns. Morrison (1990) is an example of this approach. She estimates a flexible functional form cost function, using data on gross industry output and materials inputs. Her estimates of γ and of η range between 1.2 and 1.4 for 16 out of her 18 industries. One notable feature of these estimates is that her industry estimates of the ratio of average to marginal cost closely resemble her estimates of the markup itself. Thus the relation between these two parameters that we imposed through our zero profit condition appears to be validated.

Hall (1988, 1990) proposes a variant of this approach in which, essentially, equation (2.15) is estimated using instrumental variables. The share coefficients are treated as known (from measured factor payments) rather than estimated, so that only μ need be estimated; instruments are used that are believed *a priori* to be orthogonal to exogenous technological progress γ_t^i, and endogenous variation in $\hat{I}_t$ is ignored. Hall obtains large estimates of μ for many U.S. manufacturing sectors; it is over 2 for six out of seven one-digit sectors. We discuss this approach and related estimates at the end of the next section. Here we wish simply to note that one reason that Hall's estimates of the "markup ratio" are larger than those obtained by authors such as Morrison is that he is estimating μ rather than γ.

Before closing this section, it is worth discussing the basis on which we state that profits in the U.S. economy are zero, so that μ must equal η. [12] Total tangible assets minus durables in 1987 were equal to 2.25 times that years private value added. Since private investment (again excluding durables) equaled about 18% of private value added and the yearly growth rate is about 3%, it follows that the yearly depreciation

---

[12] For a more complete discussion along similar lines, with independent estimation of the degree of increasing returns, see Hall (1990).

of this capital stock must be about 5% (to obtain this one must subtract the growth rate from the ratio of investment to capital which equals 8%). The share of payments going to capital has been 25% on average. These payments can be decomposed into the product of the capital output ratio and the sum of the rate of depreciation and an implied rate of return on capital. Thus, this implied rate of return to capital has been about 6% per annum. It is the fact that this rate of return is close to the rate of return on stock market securities that leads us to say that there are no profits in the economy. If owners of capital also received some pure profits, the payments to capitalists would exceed the product of the stock market rate of return and the actual capital stock. Yet another way of making the same point is to note that, on average Tobin's q is one (Summers (1981)). This again says that the present value of payments to the owners of capital discounted at the required rate of return on equity shares equals the cost of replacing the capital stock.

In the next two sections we study the effects of having a constant markup different from one for the effects of shocks to government purchases and of technological shocks.

4 Responses to Government Purchases

The first exercise we consider is a change in government purchases G_t. To perform this exercise we must postulate a stochastic process for G_t. The reason we must do so is that the behavior of the agents in the model depends also on their expectation of future government purchases of goods. We assume that $\hat{G}_t$ is given by

$$\hat{G}_t = \rho^G \hat{G}_{t-1} + \nu_t^G \qquad\qquad |\rho^G| < 1 \qquad\qquad (4.1)$$

In the present section we assume constant growth of z_t and N_t; hence (4.1) specifies an exogenous stochastic process for G_t. For the purposes of analyzing the response of the economy to the shock ν_t^G, we assume that the number of firms I_t does not respond to the shock. In the perfectly competitive case where μ equals 1 and Φ is equal to zero this involves no loss of generality since the number of firms is indeterminate. For this case we just set I_t equal to one. In the case of imperfect competition, changes in G_t do change the profits of the existing firms. We nonetheless assume that I_t continues to grow exogenously at the same rate as $z_t N_t$ (so that $\bar{I}_t$ is constant). Abstracting from variations in the rate of entry is reasonable as an approximation because entry decisions involve relatively long lead times.

In a model with constant markups, a change in G increases output only through an increase in the supply

of hours at a given real wage. This is apparent from (1.14b). This equation is, for a given constant inefficiency wedge μ, a relationship between the wage and the marginal product of labor where the latter depends on employment, the state of technology and capital. Since technology and capital are fixed, the willingness of firms to hire labor at a given real wage does not change. However, there are two reasons emphasized by Barro (1981) why labor supply will change. The first is that the increase in government purchases makes households less wealthy. The second is that they tend to increase real interest rates. Both of these effects raise the marginal utility of wealth λ_t and thus raise H_t for any given w_t. Thus the real wage falls and employment rises.

We compute simulations of the model's response to changes in ν_t^G for the case where ρ^G is equal to .9. We chose this value of ρ^G because changes in government purchases are persistent but also have large components which are mean reverting. [13] We assume that the share of materials in total output, s_M equals one half. This is a conservative choice since value added in manufacturing is only about half of the value of gross output in manufacturing. This parameter plays no role in our perfectly competitive model but is crucial in the imperfectly competitive case. We report results for both μ equal to one and for μ equal to 1.4. (which, using (1.8) with s_M equal to 0.5 corresponds to a markup γ equal to 1.17). This markup is fairly low relative to those that have been estimated in the literature. The other parameter values are listed in Table 1; they are taken from King, Plosser and Rebelo (1988a).

Figures 1 and 2 report the percent response in hours, output and the wage to a one percent change in ν_t^G. We see in the figures that the qualitative response of hours, output and the wage is the same when the μ equals one as in our imperfectly competitive case where $\mu = 1.4$. Output and hours rise in both cases and, as was suggested by the earlier discussion, real wages decline. But, there is a change in the magnitude of these effects, for given values of the model's other parameters. We see that output rises by about 0.04% in the case of perfect competition while it increases by 0.05% when $\mu = 1.4$. Because government purchases equal 11.7% of GNP the multiplier for government purchases is less than one half in both cases but it is larger with imperfect competition. On the other hand, the percent response in hours and real wages is smaller when μ equals 1.4 than when it equals one.

The difference in the responsiveness of hours and output does cast some doubt on the accuracy of the

---

[13] In our analysis of military purchases we found these to follow a very persistent but stationary AR(2) process. For simplicity and comparability with previous literature we consider a persistent AR(1) here.

common view according to which the presence of imperfect competition, in itself magnifies the short run effect of government purchases on economic activity. [14] This is argued to reflect the presence of "aggregate demand spillovers" on consumption demand . The idea is that the increase in output induced by the rise in government purchases raises the profits of the imperfectly competitive firms and the resulting increase in income makes consumers purchase more. It is argued that these effects amplify the response of economic activity through a "multiplier" process of the Kahn-Keynes type. It is true that, in our model, a larger μ. for given values of the other parameters, magnifies the response of output. But this is solely due to the fact that imperfectly competitive firms set the wage below the marginal product of labor so that a one percent increase in hours raises output by μs_H percent rather than by s_H percent. [15] On the other hand, a larger μ reduces the response of hours.

The reason for this is easily seen. Hours become less responsive when μ is increased because the (negative) wealth effect of an increase in government purchases become smaller. For given prices (wages and interest rates), labor supply depends upon the present value of after tax income net of wages. In either case, an increase in G increases the present discounted value of taxes (and hence reduces after-tax income) by exactly the amount of the increase in G. But, potentially, changes in G have an additional effect on the present value of income. If they lead equilibrium hours and output to rise, income net of wages increases by the amount by which the increase in output exceeds the increase in the wage bill. To first order, this amount is zero when the marginal product of hours equals the real wage but it is positive if the marginal product of hours exceeds the real wage (i.e.,if firms have market power). Thus, in the case of perfect competition, a one dollar increase in G lowers the present value of income by one dollar, while in the case of imperfect competition it lowers it by less than one dollar.

In a sense the "demand spillover" literature is correct in arguing that the increased profits in the case of imperfect competition give an additional boost to household income. Where this argument errs is in supposing that the stimulative effects of government purchases result from a *positive*, rather than a *negative* effect of government purchases on household income. It also errs in supposing that the direct determinant

[14] See, e.g. Mankiw (1988), Startz (1989) and Silvestre (1993).

[15] One can also understand from (2.15) the basis of Startz's (1989) argument that this is just a short run effect. A permanent increase in G would permanently increase profits, requiring an eventual increase in the number of firms I. In a comparison of steady state equilibria (in which I is assumed to adjust so as to keep profits equal to zero), the percentage increase in output equals only s_H times the percentage increase in the labor input.

of equilibrium employment is the effect of household income on *consumption demand*, rather than the effect of income on *labor supply*.

Another implication of Figure 1 worthy of note concerns the effect of government purchases on measured labor productivity. Since output increases more and hours increase less, in the case of μ greater than one, it is obvious that measured productivity rises more (falls less) in the case of imperfect competition. The behavior of the Solow residual, which is used as a measure of exogenous productivity growth in the real business cycle literature, is especially noteworthy. Recall that the Solow residual is given by

$$\Delta Y_t - s_H \Delta H_t - s_K \Delta K_t \tag{4.2}$$

where ΔY_t represents the change in the logarithm of Y_t and analogously for ΔH_t and ΔK_t. By contrast, differentiation of (1.13) using (1.14) and keeping I constant establishes that

$$\gamma_t^z = \frac{\Delta Y_t - \mu s_H \Delta H_t - \mu s_K \Delta K_t}{\mu s_H} \tag{4.3}$$

The numerator of (4.3) thus must be invariant (to first order) to changes in G_t. But this implies that if μ is greater than one and $\dot{H}_t$ rises for any reason other than a technological shift (so that γ_t^z is unchanged) the expression in (4.2) will rise. In other words, an increase in hours induced by an increase in government purchases or any other non-technology shock raises the standard Solow residual. The reason is that, with imperfect competition, an increase in the labor input must necessarily raise the value of output by more than it raises labor costs (since firms make profits on the marginal units). This results in an increase in measured productivity.

This shows that ignoring imperfect competition when it is actually present is dangerous. It leads to incorrect measures of total factor productivity. If one wants to measure true changes in z, one must use the formula (4.3) which depends on μ in addition to depending on observable magnitudes.

The fact that the Solow residual (4.2) is not a correct measure of true technical progress may explain some observed anomalies concerning Solow residuals. For example, a number of authors have observed that, in postwar U.S. data, Solow residuals are correlated with various measures of government purchases (Hall (1988), Baxter and King (1991), Burnside, Eichenbaum and Rebelo (1993)). One might argue that this simply indicates that government purchases are not exogenous with respect to technology shocks. This might be true of some components of government purchases, especially spending by local governments that

25

are forced to have balanced budgets. But it is hard to defend the "reverse causation" thesis in the case of national defense-related goods, that have moved mainly in response to changed perception of the threat posed by communist regimes. Yet this component is positively correlated with Solow residuals as Hall (1988) and Baxter and King (1991) show. Our model can explain this observation insofar it predicts that hours should increase in response to an exogenous increase in government purchases and, as a result, the Solow residual should increase as well if $\mu > 1$. [16]

Imperfect competition might similarly explain the observations of Hall (1988) that changes in world oil prices (that, again, should be exogenous with respect to the state of U.S. productivity) are correlated with Solow residuals, [17] and the findings of Evans (1990) that various measures of monetary policy shocks forecast future Solow residuals. [18]

The ability of imperfect competition to explain these anomalies provides not only an argument that imperfect competition is important but may also be the basis for a quantitative estimate of its importance. Indeed, this is the basis for Hall's (1988) estimates of μ. The parameter μ can be estimated using (4.3) if one observes a variable v_t, such as military purchases or changes in the world oil price, that is both correlated with output and hours changes and is known to be orthogonal to the change in technology γ_t^z. Then v_t should be orthogonal to the right hand side of (4.3). In particular

$$\text{Cov}(v_t, \Delta Y_t - \mu s_H \Delta H_t - \mu s_K \Delta K_t) = 0 \tag{4.4}$$

Hall's proposal is to estimate μ so as to minimize (under an appropriate metric) the extent to which the moment condition (4.4) fails to hold. In the case where v_t is a single variable, the expression in (4.4) is actually zero if μ is replaced by the instrumental variable estimate

$$\mu^s = \frac{\sum_t v_t \Delta Y_t}{\sum_t v_t (s_H \Delta H_t + s_K \Delta K_t)}$$

Hall's estimates indicate estimates of the wedge μ of over 1.8 for all seven 1-digit industries he considers. Subsequent work by Domowitz, Hubbard and Petersen (1988), uses gross industry output so that it estimates

---

[16] Alternative possible explanations of the anomaly that do not depend on imperfect competition are proposed by Baxter and King (1991) and Burnside, Eichenbaum and Rebelo (1993).

[17] For a complete dynamic equilibrium model of the effects of oil prices which is constructed along the lines of this paper, see Rotemberg and Woodford (1993).

[18] Explicit development of this last idea would require, of course, a model where monetary policy shocks affect economic activity, a topic we do not take up here. Imperfect competition in product markets does not in itself imply any real effect of monetary policy. On the other hand, as we mentioned in footnote 2, imperfect competition is often a crucial element of models in which monetary non-neutrality results from price rigidity.

the markup γ. Their estimates of γ range between 1.4 and 1.7 for 17 out of their 19 industries. These two sets of estimates do not contradict each other because, as long as the materials share s_M is positive, μ is smaller than γ as can be seen from (1.66). For example, μ equal to 2.3 (a typical value for Hall's industries) and a materials share of .5 (which is also typical for those industries) imply $\gamma = 1.4$ which is in the range of those estimated by both Domowitz, Hubbard and Petersen (1988) and Morrison (1990).

5 Responses to Technology Shocks

In this section, we show that the level of the average markup matters when it comes to the response of the economy to changes in technology. Before we can compute the economy's response, we must postulate a stochastic process for the level of technology itself and for the number of firms. It is not enough to simply compute the size of the technology shocks using (4.3). The reason is that, once again, household's decisions of how much labor to supply and output to consume depend upon their expectations of the future state of productivity. We are also unable, in this case, to ignore the issue of variation in the rate of entry of new firms, as we do not wish to assume that technology shocks have a purely transient effect on productivity. In fact, in this section, we follow King, Plosser and Rebelo (1988b) and Plosser (1989) in assuming that productivity is a random walk, so that γ_t^i is an independently distributed random variable. [19]

As we showed in (1.12), the number of firms must grow with z for profits to remain equal to zero in the steady state. On the other hand, we do not believe the number of firms adjusts very rapidly to a technology shock. We thus wish to concentrate on short run effects of technology shocks by assuming that entry plays only a small role in these short term dynamics. To do this, we let I_t follow an error-correction process of the form

$$\log I_t = \kappa \log(I z_t N_t) + (1 - \kappa) \log I_{t-1} \tag{5.1}$$

where I and κ are positive constants with $\kappa < 1$. This process implies that with a stationary $\{\gamma_t^i\}$, $\{I_t\}$ is itself stationary as we assumed in Section 2. By letting κ be small we ensure that there is little immediate entry in response to a technology shock, even when it is permanent. We conjecture that our results hinge

[19] We made this assumption because the results we obtained for the case where z_t is stationary were unsatisfactory. We show below that there are exist two different methods of computing the variance of output implied by the model as a response to technology shocks. While the methods are different, they ought to give the same answer for the volatilities of output and hours. The two methods do so when we assume that z is a random walk but we lead to very different theoretical variances for output and hours when we assumed that z is stationary. For more on the sensitivity of theoretical variances of output as one varies the stochastic process for the technology shocks, see Eichenbaum (1991).

27

mainly on the fact that the response of entry with a small κ is slow and not on the precise specification of (5.1). In particular, a small κ preserves comparability between our results and those obtained by Hornstein (1993) for a model without permanent technology shocks in which the effects of technology shocks on entry are ignored.

Differencing (5.1), we obtain

$$\Delta I_t = \kappa(\gamma_t^z + \gamma^N) + (1 - \kappa)\Delta I_{t-1}$$

which implies that

$$\Delta \tilde{I}_t = \sum_{j=0}^{\infty} \kappa(1 - \kappa)^j \tilde{\gamma}_{t-j}^z \qquad (5.2)$$

where $\Delta \tilde{I}_t$ denotes ΔI_t minus the unconditional mean of that stationary variable while, as before, $\tilde{\gamma}_t^z$ denotes γ_t^z minus its mean. [20]

Differentiation of (1.13) now yields

$$\gamma_t^z = \frac{\Delta Y_t - \mu s_H \Delta H_t - \mu s_K \Delta K_t + (\mu - 1)\Delta I_t}{\mu s_H} \qquad (5.3)$$

This implies that the change in the number of firms enters the corrected Solow residual (analogous to (4.3)), in the case of imperfect competition. This is because entry of firms increases fixed costs and thus requires an increase in productivity if a given quantity of output is to be produced with the same inputs. Thus an assumption such as (5.1) is necessary in order to measure technology shocks. Removing means from (5.3) and substituting (5.2), we obtain

$$\tilde{\gamma}_t^z = \Omega(L) \left[\Delta \tilde{Y}_t - \mu s_H \Delta \tilde{H}_t - \mu s_K \Delta \tilde{K}_t \right] \qquad (5.4)$$

where $\Omega(L)$ is a polynomial in the lag operator whose coefficients depend on μ and κ. This polynomial has no roots inside the unit circle; in particular, it equals 1 when μ is one. Using (5.4), we construct a series for γ_t^z for the U.S. economy using quarterly data from 1947:1 until 1989:4 on private output and private sector hours while assuming that the change in capital is constant. [21] The construction of our output and hours series is discussed in Rotemberg and Woodford (1992).

The first part of Table 2 presents the measured variance of γ_t^z for various values of μ. It shows that the measured variance of technology falls as we increase the markup. The reason is that typical U.S. business

---

[20] Note that with this notation, if a variable X is stationary, then $\Delta \tilde{X}$ is the first difference of $\tilde{X}$.

[21] This is not strictly correct but, because investment is such a small fraction of capital, it does not induce a large bias into our calculation.

cycles involve procyclical movements in output, hours and the standard Solow residual. If μ is greater than one, one expects the Solow residual to move procyclically even in the absence of technology shocks as explained in the previous section. Thus a model with μ greater than one can explain some fraction of the variation in the Solow residual leaving smaller unexplained variations in total factor productivity.

We computed the model's responses to shocks in γ^z using the King, Plosser and Rebelo (1988a) parameters listed in Table 1, a share of materials of one half and, again, values μ equal to 1 and 1.4, We assumed that κ is equal to 0.02 so as to make sure that immediate entry had a relatively trivial effect on the results. The model's prediction for the effect on output and hours of a 1% shock to γ_t^z is presented in Figure 3. We see that, for a given increase in z, the response of output is higher under imperfect competition. The reason is that, an increase in z represents, in effect, an increase in the effective units of labor that firms hire. Because firms with market power set the marginal product of labor higher than the wage, an increase in the effective labor input raises output more under imperfect competition. This can be seen directly from equation (2.15) which shows that the change in output for given hours and capital input is equal to μs_H times γ^z. This is why a higher μ raises the response of output.

By contrast, the response of hours worked is smaller when the markup is higher. Indeed, for our chosen value of μ, hours are nearly insensitive to changes in technological possibilities. That the response of hours to a change in technological opportunities is ambiguous is well known. Because capital is fixed in the short run, an increase in z raises temporarily the marginal product of labor and the wage and this leads workers to substitute current work for future leisure. On the other hand, an increase in z also makes people wealthier and this reduces labor supply. The net effect depends on whether the intertemporal substitution or the wealth effect is larger. Imperfect competition increases the size of the wealth effect, as in the previous section, and so reduces the extent to which labor supply increases. If μ is made slightly larger than 1.4, a positive technology shock actually reduces equilibrium hours, though output still increases.

This result, that employment fluctuations are smaller with imperfect competition, is also obtained by Hornstein (1993) who discusses it in terms of the variances of output and hours that can be explained by technology shocks. We next turn to the consequences of imperfect competition for this type of exercise.

There are two different ways of computing the variance of output and hours that can be attributable to changes in z. They both rely on the impulse response functions which are plotted in Figure 3. Let the

impulse response functions be written as

$$\Delta \hat{Y}_t = \sum_{i=0}^{\infty} \xi_i^Y \hat{\gamma}_{t-i}^z \tag{5.5}$$

$$\Delta \hat{H}_t = \sum_{i=0}^{\infty} \xi_i^H \hat{\gamma}_{t-i}^z \tag{5.6}$$

Then one estimate of the implied variance of the change in output is

$$\sum_{i=0}^{\infty} (\xi_i^Y)^2 \text{Var}(\gamma_t^z) \tag{5.7}$$

where $\text{Var}(\gamma_t^z)$ is the variance of the technology shock reported in Table 2. The variance of the change in hours can be computed analogously.

An alternative computation of the variances of output and hours relies on the historical time series for the technology shock that can be computed using (5.4). As in Plosser (1989), we can compute the values of output and hours that these time series for technology shocks predict. If $\bar{\gamma}_t^z$ gives the (de-meaned) historical series for technology shocks, the predicted series for (de-meaned) changes in output and hours are, respectively

$$\Delta \bar{Y}_t = \sum_{i=0}^{\infty} \xi_i^Y \bar{\gamma}_{t-i}^z \qquad \Delta \bar{H}_t = \sum_{i=0}^{\infty} \xi_i^H \bar{\gamma}_{t-i}^z \tag{5.8}$$

We can then compute the sample variance of the series $\{\Delta \bar{Y}_t\}$ and $\{\Delta \bar{H}_t\}$.

Column 2 of the second and third part Table 2 give empirical variances of output and hours growth while column 3 gives their theoretical variances computed using the sample variances of (5.8). We do not report the variances computed using (5.7) because they are nearly identical. [22] For the standard case where μ is equal to one, the predicted variance of output is somewhat smaller than the actual variance while the predicted variance of hours is quite a bit smaller than the actual variance. Increases in μ lower both predicted variances though they have a more substantial effect on the variance of hours for the reasons that we gave above.

We also present in Table 2 a statistic that provides a further test of the empirical plausibility of the model's predictions regarding the effects of technology shocks. This statistic relates to the orthogonality

---

[22] As we said in footnote 19 above, this is not true when technology shocks are assumed to induce only transitory changes in technology. When we assumed that z_t followed an autoregressive process we obtained much smaller theoretical variances for output changes when using the method of Kydland and Prescott (1982) than when we asked about the variability of output generated by the model in response to the entire time series of Solow residuals. This was true even when we made the autoregressive coefficient equal to .99.

of the predicted movements in output and hours on the one hand, and the prediction errors on the other. We have mentioned above that a correct measure of technical progress should be independent of the other shocks that affect the economy. In the previous section we showed that if some of the other shocks can be directly measured, this gives rise to an orthogonality condition that can be used to test the validity of the method of measuring technology shocks. But even when the other shocks cannot be directly measured, the independence principle can be used as the basis for a specification test of a model of the effects of technology shocks. For if the technology shocks are correctly measured and the theoretical impulse response coefficients ξ^Y and ξ^H are correct, then the prediction errors $\Delta Y - \Delta \bar{Y}$ and $\Delta H - \Delta \bar{H}$ should be expressible as distributed lags of the other exogenous shocks. [23] It follows that

$$\text{Cov}\{\Delta \tilde{Y}_t, (\Delta Y_t - \Delta \tilde{Y}_t)\} = 0 \tag{5.9}$$

$$\text{Cov}\{\Delta \tilde{H}_t, (\Delta H_t - \Delta \tilde{H}_t)\} = 0 \tag{5.10}$$

Testing the validity of these moment conditions provides a specification test which does not require knowledge of the other types of shocks. To provide a convenient measure of the extent to which the moment conditions (5.9) and (5.10) are violated we decompose the variance of actual changes in output as follows

$$\text{Var}(\Delta Y_t) = \text{Var}(\Delta \tilde{Y}_t) + \text{Var}(\Delta Y_t - \Delta \tilde{Y}_t) + 2\text{Cov}\{\Delta \tilde{Y}_t, (\Delta Y_t - \Delta \tilde{Y}_t)\}$$

so that

$$1 = \frac{\text{Var}(\Delta \tilde{Y}_t)}{\text{Var}(\Delta Y_t)} + \frac{\text{Var}(\Delta Y_t - \Delta \tilde{Y}_t)}{\text{Var}(\Delta Y_t)} + \omega^Y$$

where

$$\omega^Y = \frac{2\text{Cov}\{\Delta \tilde{Y}_t, (\Delta Y_t - \Delta \tilde{Y}_t)\}}{\text{Var}(\Delta Y_t)}$$

Thus ω^Y measures the extent to which the sum of the variance of output that the model attributes to technology shocks and the variance it attributes to other sources fails to equal the total variance of output. The size of the departure of ω^Y from zero thus gives one some idea how seriously one can take a statistic such as $\text{Var}(\Delta \tilde{Y}_t)/\text{Var}(\Delta Y_t)$ as a measure of the degree to which observed variations in output growth can be explained by technology shocks.

---

[23] Here we rely on the validity of the log-linear approximation introduced in section 2.

The fifth column of Table 2 gives values of ω^Y and of the analogously defined ω^H for different values of μ. For the model to be correct, both ω's ought to be zero. Hence, these provide additional moment restrictions that can be used to estimate μ. Such an estimate is then based entirely upon the way in which one parameter value as opposed to another improves the model's ability to produce empirically plausible predictions regarding the part of aggregate fluctuations that are due to aggregate technology shocks.

As we can see, raising μ from 1 to 1.4 lowers both ω's which reach a minimum for μ between 1.4 and 1.6. Thus, the results of this estimation procedure also yield relatively important departures from perfect competition. What is perhaps even more interesting is that the resulting estimate of μ implies that technology shocks lead to practically no fluctuations in hours worked. Thus, essentially, the entire movement in hours worked must be due to some other type of shock.

6 Fluctuations Due to Self-Fulfilling Expectations

The introduction of imperfect competition and increasing returns also makes possible an entirely new source of equilibrium fluctuations in economic activity. In particular, equilibria may exist in which economic activity fluctuates in response to random events that do not involve any change in underlying fundamentals ("sunspots"). These fluctuations are caused simply by changes in people's expectations of the future path of the economy. This need not be because individual agents' expectations are incorrect – instead, in such a "sunspot equilibrium" it is correct to expect a different future path for the economy, if everyone else's expectations and actions change in response to the event in question. Equilibria of this kind are now known to be possible in many kinds of intertemporal equilibrium models (see, e.g., Guesnerie and Woodford (1992)). They are not, however, possible in the case of the standard neoclassical growth model. For there, the equilibrium allocation of resources must maximize the expected utility of the representative household, and there is a unique allocation with this property.

Once we introduce imperfect competition, as above, the first welfare theorem no longer holds, and as a result one cannot show in such a simple way that equilibrium must be unique. Indeed, it need not be, as Benhabib and Farmer (1992) show in the case of a model with monopolistic competition like that described here in sections 1-2. [24] Recall that in section 2 we observe that, for our calibrated parameter values, the

---

[24] The indeterminacy of rational expectations equilibrium, and the possibility of endogenous equilibrium fluctuations, were first discussed in the context of a model of this kind by Hammour (1988). For other examples of stationary sunspot equilibria

matrix $A^{-1}B$ (where A and B are the matrices in (2.20)) has one real eigenvalue with absolute value less than one, and another with absolute value greater than one; this allows us to compute a locally unique stationary solution to (2.20). This need not be true, however, for all parameter values, and Benhabib and Farmer show that if μ and η are sufficiently large, $A^{-1}B$ instead has two eigenvalues with modulus less than one (which for some parameter values are a complex pair). In this case, there exists a large multiplicity of stationary rational expectations equilibria, including equilibria in which output fluctuates in response to "sunspot" events.

Consider, for simplicity, the case in which there are no stochastic variations in any of the exogenous "fundamentals" $\{\gamma_t^z, \hat{G}_t, \hat{I}_t\}$. Then (2.20) becomes simply

$$A \begin{pmatrix} E_t \hat{\lambda}_{t+1} \\ \hat{K}_{t+1} \end{pmatrix} = B \begin{pmatrix} \hat{\lambda}_t \\ \hat{K}_t \end{pmatrix}$$

A stationary solution is given by the bivariate stochastic process

$$\begin{pmatrix} \hat{\lambda}_{t+1} \\ \hat{K}_{t+1} \end{pmatrix} = A^{-1}B \begin{pmatrix} \hat{\lambda}_t \\ \hat{K}_t \end{pmatrix} + \begin{pmatrix} u_{t+1} \\ 0 \end{pmatrix}$$

where $\{u_t\}$ is a mean-zero white noise "sunspot" variable, and the realization of u_{t+1} becomes known only at date $t+1$. [25] In fact, all stationary solutions must be of this form (for some sunspot variable $\{u_t\}$). Thus the multiplicity of equilibria does not mean that theory lacks testable implications about the character of aggregate fluctuations. In the absence of exogenous shocks and to the extent that the log-linear approximation is accurate, all of the stationary equilibria are simply scalar multiples of a single equilibrium. Thus the model does not predict the amplitude of the fluctuations, but one is able to obtain definite numerical predictions about the *relative* variability of output, hours, investment, real wages and so on, as well as definite predictions about the serial correlation and cross correlation of all these series. [26] Even when other shocks are added, the set of stationary equilibria is a set of finite dimension (linear combinations of a small number of possible types of fluctuations), so that relatively strong restrictions are placed on the data.

Farmer and Guo (1993) show that in the case of a model of this kind that is "calibrated" in a relatively standard fashion, except for the large values assumed for μ and η, the predicted fluctuations in aggregate

---

in models with imperfectly competitive product markets and increasing returns, see Woodford (1991), Hammour (1991), and Gali (1991).

[25] This solution to the log-linearized equilibrium conditions (2.20) approximates a solution to the exact equilibrium conditions, in the case that the amplitude of "sunspot" fluctuations are sufficiently small. See Woodford (1986) for details.

[26] Early illustrations of this were given in Woodford (1988, 1991).

quantities, in the absence of any exogenous shocks, exhibit relative variabilities and co-movements similar to those observed in de-trended U.S. data. They argue that in this respect the model's predictions are not clearly less consistent with the facts than are those of a standard RBC model, with perfect competition, constant returns, and exogenous technology shocks. The standard model, of course, also seeks to explain the amplitude of aggregate fluctuations, given that measured Solow residuals can be taken to indicate the size of the exogenous technology shocks, while the amplitude of fluctuations remains unexplained in the Farmer-Guo model. The Farmer-Guo model, however, can in principle explain the co-movement of measured Solow residuals with other aggregate variables; for the model predicts variation in measured Solow residuals in response to "sunspot" events that cause variations in output, for the reasons discussed in section 4. [27]

The degree of market power and increasing returns assumed by Farmer and Guo ($\mu = 1.72, \eta = 1.61$) is not completely outside the range of values suggested by the empirical evidence discussed in sections 3 and 4. However, it should be noted that in their model, assuming somewhat lower values does not simply reduce the *magnitude* of the equilibrium response to the sunspot events; it eliminates it altogether. [28] For the stationary sunspot equilibria exist only if both eigenvalues are inside the unit circle. Since one of them must be outside the unit circle in the case of perfect competition (because of the first welfare theorem), and since the eigenvalues vary continuously as one varies the parameters μ and η toward one, there must be a point at which μ and η still exceed one but one eigenvalue has a modulus greater than one. The quantitative reasonableness of their assumptions about μ and η is therefore a critical issue, more so than in the case of other consequences of imperfect competition and increasing returns, such as Hall's interpretation of productivity variations. The reason is that, as shown by Hall, any departures from perfect competition will generate *some* procyclical productivity even if the departures are not big enough to explain all the variations in productivity. On the other hand, Farmer and Guo require that the departures be significant. Indeed, in the case of Farmer and Guo's calibration, μ and η cannot be made much smaller than the values that they assume. Further research on the magnitude of these parameters thus seems crucial to establish

---

[27] The possibility that the observed co-movement of measured Solow residuals with output could be consistent with a model in which all fluctuations are due to self-fulfilling expectations was first illustrated in Woodford (1991).

[28] The notion that the possible amplitude of the sunspot equilibria is unaffected by the parameter values is to some extent an artifact of the log-linear approximation used here. In the case of the log-linear equilibrium conditions (2.20), if any sunspot solution exists, solutions exist with fluctuations of arbitrary amplitude. In the case of the exact equilibrium conditions instead, it is possible that the set of possible stationary sunspot equilibria includes only fluctuations over a certain range of amplitudes, the bounds on which collapse to zero as the critical parameter values are approached at which local sunspot equilibria cease to be possible. However, the general point derived from analysis of the log-linear system, that stationary sunspot equilibria cease to be possible while some amount of market power and increasing returns still exist, remains valid.

34

the empirical validity of their model.

7 Models with Exogenously Varying Markups

Up to this point we have considered a model in which, as in the perfectly competitive model, the markup is constant. One benefit of considering models with imperfect competition is that this class of models also contains models where markups vary over time. This leads to specifications that are considerably richer than those with constant markups. In particular, markup variations may play an important role in generating fluctuations in equilibrium employment. To show this, we first consider the effect of exogenous changes in markups. In the next section, we take up models of endogenous markup variation.

Suppose that there are exogenous variations in each firm's elasticity of demand $D'(1)$, resulting from variations in the degree of substitutability of the various differentiated goods. Giving this elasticity a subscript t, equation (1.7) becomes

$$\gamma_t^i = [1 + 1/D_t'(1)]^{-1} \tag{7.1}$$

Equation (1.8) then implies that the value added markup varies as well. This variation in the markup implies a variation in the ratio of the marginal product of labor to the wage. In particular, (1.14b) becomes

$$z_t F_2(K_t, z_t H_t) = \mu_t w_t \tag{7.2}$$

Thus varying markups imply labor demand shifts, i.e.,changes in the amount of labor that firms will hire at a given real wage. In this respect changes in markups are similar to changes in the productivity parameter z_t. In terms of the detrended variables this becomes

$$F_2\left(\frac{\tilde{K}_t}{\gamma_t^i}, \tilde{H}_t\right) = \mu_t \tilde{w}_t \tag{2.10'}$$

The variability of markups also changes equation (2.13) which becomes

$$1 = \beta \left(\gamma_t^z\right)^{-\sigma} E_t \left\{ \left(\frac{\tilde{\lambda}_{t+1}}{\tilde{\lambda}_t}\right) \left[\frac{F_K(\tilde{K}_{t+1}/\gamma_{t+1}^z, \tilde{H}_{t+1})}{\mu_t} + (1-\delta) \right] \right\} \tag{2.13'}$$

These are the only equilibrium conditions affected by the variability of the markup. Their linearization yields

$$\frac{s_K}{\epsilon_{KH}}(\hat{K}_t - \gamma_t^z - \hat{H}_t) = \hat{w}_t + \bar{\mu}_t \tag{2.16'}$$

$$-\sigma\hat{\gamma}_t^z + E_t \left\{ \hat{\lambda}_{t+1} - \hat{\lambda}_t + \left(\frac{r+\delta}{1+r}\right)\frac{s_H}{\epsilon_{KH}}(\hat{H}_{t+1} - \hat{K}_{t+1} + \hat{\gamma}_{t+1}^z) - \left(\frac{r+\delta}{1+r}\right)\hat{\mu}_t \right\} = 0 \tag{2.19'}$$

35

where $\tilde{\mu}_t$ is the logarithmic deviation of the markup from its steady state value (which we shall denote by μ). To solve the current model, we substitute (2.15), (2.16') and (2.17) into (2.18) and (2.19') to obtain two difference equations which now also involve the stationary random variable $\tilde{\mu}_t$.

To compute how the system responds to markup shocks, we postulate a stochastic process for $\tilde{\mu}_t$. In particular, we assume that

$$\tilde{\mu}_t = \rho^\mu \tilde{\mu}_{t-1} + \nu_t^\mu \tag{7.3}$$

We use the same parameter values as in previous simulations. [29] For the markup, we assume as before that the mean of μ is 1.4. We experiment with a variety of values of ρ^μ. Figure 4 reports the response of output, hours, consumption and investment to a unit shock to ν^μ when ρ^μ equals 0 while Figure 5 reports the same responses when ρ^μ equals 0.9. Figure 6 reports the response of real wages for both values of ρ^μ. The instantaneous responses of all three variables do not depend to any significant extent on ρ^μ. Not surprisingly, the higher value of ρ^μ makes the responses of output, hours and wages more persistent.

Figure 6 shows that real wages decline when markups rise. This is to be expected since increases in markups lower labor demand. What is important about this is that it shows that, in response to markup shocks, real wages move procyclically.

One notable feature of figures 4 and 5 is that, in the immediate aftermath of the shock, consumption moves less than output while investment moves more. Thus, our model with markup shocks makes consumption less variable than output whereas investment is more variable. This higher relative variability of investment is a robust feature of business cycles which the model reproduces. The success of the model along this dimension is due to the relative unwillingness of consumers with concave utility functions to substitute their consumption intertemporally coupled with the small sensitivity of the marginal product of capital to short run changes in investment. These also account for the relative variability of consumption and investment in standard real business cycle models.

We also observe that the short run impact on hours of a markup shock exceeds that on output. When ρ^μ equals 0, a unit increase in ν^μ lowers hours by 1.38 while output falls only by 1.13. This rather large change in hours must be contrasted with the negligible effect of a unit technology shock that we displayed above.

---

[29] Hence, as earlier, the matrix $A^{-1}B$ has exactly one eigenvalue with a modulus less than one, and so the equilibrium response to the markup variations is determinate, just as in the case of the responses to variations in government purchases and to technology shocks.

The reason that hours fall so dramatically when μ rises, even though they hardly change when z rises is that, in the latter case, wealth and substitution effects work in opposite directions. In particular, an increase in z raises wealth which discourages work. By contrast, increases in μ have only very small wealth effects. The reason is that, to first order, the increase in firm's profits is exactly offset by the losses to consumers. Thus, the main effect of increases in μ is to lead households to substitute leisure for consumption. The small importance of wealth effects probably accounts also for the similarity of the responses to temporary and more persistent changes in μ.

The difference in the response of output and hours is due to the value of μs_H. For our choice of parameters $\mu s_H < 1$. As a result, (2.15) implies that, holding z constant, the change in output is smaller than the change in hours. This conclusion would be reversed if μs_H were bigger than one.

We have also computed the variance of the log changes in hours and in output for a variety of values of μ and ρ^μ. In Table 3 we display these variances as ratios of the variance of the log changes in μ. We normalize by the variance of the changes in μ so that the variances do not grow spuriously as we raise ρ^μ. We see that for large values of μ, which imply that μs_H is greater than one, the variance of output actually exceeds the variance of hours. However, for smaller values of μ the variance of hours changes exceeds the variance of output changes. If parameter values of this type are entertained, markup shocks cannot possibly be the only shocks impinging on the economy. The reason is that, as we saw earlier, the variance of hours is in fact smaller than the variance of output. However, a larger predicted variance of H can be reconciled with the model as long as markup shocks coexist with technology shocks. Since the latter affect mainly the variance of output, a combination of the two types of shocks can potentially explain both why hours fluctuations are significant and why output variability is larger than the variability of hours.

This discussion raises the question of whether average markups actually vary, and thus of how much of the observed variation in hours can be attributed to them. Important markup variations, with markups being much lower in booms than in recessions were found by Bils (1987) and Rotemberg and Woodford (1991). Using a somewhat different specification of technology, Rotemberg and Woodford (1991) find that the assumption of an average markup of 1.6 implies that the variance of quarterly changes in markups equals 1.33%. This number is not directly comparable to the variance in markups in our theoretical model because of differences in specification and because the actual stochastic process for markups is not (7.3). Nonetheless,

37

it is worth noting that the numbers of Table 3 suggest that such a large variability in markups implies that output and hours should vary even more than they actually do.

8 Models of Endogenous Markups

In the last section we showed that markup variations of a plausible magnitude can explain the fluctuation in hours worked. The problem with the analysis of that section is that exogenous changes in markups do not appear to be particularly plausible. Models of imperfect competition would be more attractive if markup changes could, themselves, be explained by changes in other variables. Of particular interest in this regard are changes in aggregate demand, *i.e.*,changes in desired purchases away from the future and towards the present. [30] Changes in aggregate demand include many of the variables traditionally held responsible for business fluctuations, including changes in current government purchases, changes in the expected future profitability of current investment (as opposed to changes in the productivity of the existing capital stock). and changes in "consumer sentiment".

In this section we briefly describe three models of endogenous markup determination and their ability to fit business cycle facts. These models can be separated in two types. The first has markups depend only on the size of aggregate demand at different points in time but makes the markup independent of the composition of demand. In this category are the models of markup determination discussed by Rotemberg and Woodford (1991). The second type of model makes markups depend on the composition of demand but not on the level of aggregate demand itself. In the latter category fall the models of Bils (1989) and Gali (1991).

We first consider two models of the first type. The first is a customer market model based on Phelps and Winter (1970) which is similar to Phelps (1992), the second is an implicit collusion model based on Rotemberg and Woodford (1992).

---

[30] From a macroeconomic perspective another potentially important determinant of average markups, and hence of aggregate activity, is the level of inflation. Benabou (1992) shows both that many search models imply a connection between inflation and markups, and that markups in the retail sector constructed along the lines of Rotemberg and Woodford (1991) are in fact negatively correlated with the rate of inflation.

8.1 The Customer Market Model

The customer market model we consider is based on Phelps and Winter (1970). As before, firms maximize profits with respect to their own markup taking the markup charged by all other firms as given. It differs from the earlier model in that demand has a dynamic pattern. A firm that lowers its current price not only sells more to its existing customers, but also expands its customer base. Having a larger customer base raises future sales for any given future price. It would be attractive to obtain such a specification of demand from underlying aggregator functions for consumers such as (1.1) which would depend on previous purchases. Unfortunately, we are unable to do so and capture the basic idea by simply writing the quantity demanded from firm i at time t, q_t^i as

$$q_t^i = \frac{Q_t}{I_t} \psi \left(\frac{\mu_t^i}{\mu_t} \right) m_t^i, \qquad \psi' < 0, \quad \psi(1) = 1. \tag{8.1}$$

The variable m_t^i is the fraction of average demand Q_t/I_t that goes to firm i if it charges the same price as all other firms. The ratio of markups in (8.1) represents the relative price of firm i's good, since marginal cost is independent of the scale of operation and the same for all firms (as in section 1). Thus (8.1) is a straightforward generalization of (1.2). The market share m^i depends on past pricing behavior according to the rule

$$m_{t+1}^i = g \left(\frac{\mu_t^i}{\mu_t} \right) m_t^i \qquad g' < 0, \quad g(1) = 1 \tag{8.2}$$

so that a temporary reduction in relative price raises firm i's market share permanently. Equations (8.1) and (8.2) are intended to capture the idea that customers have switching costs, in a manner analogous to the models of Gottfries (1986), Klemperer (1987), Farrell and Shapiro (1988) and Beggs and Klemperer (1992). A reduction in price attracts new customers who are then reluctant to change firms for fear of having to pay these switching costs. One obvious implication of (8.1) and (8.2) is that the long run elasticity of demand, i.e.,the response of eventual demand to a permanent increase in price, is larger than the short run elasticity of demand. In our case, a firm that charges a higher price than its competitors eventually loses all its customers, though this is not essential for our analysis.

Ignoring fixed costs, firm i's profits at t are given by

$$\frac{\gamma_t^i - 1}{\gamma_t} \frac{Q_t}{I_t} \psi \left(\frac{\mu_t^i}{\mu_t} \right) m_t^i$$

39

Using (1.8) and an analogous condition for individual markups as well as the fact that Y equals $(1 - s_M)Q$, these profits also equal

$$\frac{\mu_t^i - 1}{\mu_t} \frac{Y_t}{I_t} \psi\left(\frac{\mu_t^i}{\mu_t}\right) m_t^i \tag{8.3}$$

Thus, the firm's expected present discounted value of profits from period t onward is

$$E_t \sum_{j=0}^{\infty} \alpha^j \beta^j \frac{\lambda_{t+j}}{\lambda_t} \left(\frac{\mu_{t+j}^i - 1}{\mu_{t+j}}\right) \frac{Y_{t+j}}{I_{t+j}} \psi\left(\frac{\mu_{t+j}^i}{\mu_{t+j}}\right) m_t^i \prod_{s=0}^{j-1} g\left(\frac{\mu_{t+s}^i}{\mu_{t+s}}\right). \tag{8.4}$$

where our earlier analysis implies that $\beta^j \frac{\lambda_{t+j}}{\lambda_t}$ is the pricing kernel for valuing contingent securities that pay off in period $t + j$. The quantity $1 - \alpha$ represents the probability that a firm will, for random reasons, be assigned a market share in the next period that is independent of its past pricing behavior. For example the firm might cease to exist with this probability. Firm i chooses μ_t^i to maximize (8.4), taking as given the stochastic processes $\{\mu_t\}$, $\{\lambda_t\}$ and $\{Y_t/I_t\}$. Therefore

$$\psi\left(\frac{\mu_t^i}{\mu_t}\right)\frac{Y_t}{I_t} + \qquad\qquad \psi'\left(\frac{\mu_t^i}{\mu_t}\right)\left[\frac{\mu_t^i - 1}{\mu_t}\right]\frac{Y_t}{I_t} + $$
$$g'\left(\frac{\mu_t^i}{\mu_t}\right)E_t\sum_{j=1}^{\infty}\alpha^j\frac{q_{t+j}}{q_t}\left[\frac{\mu_{t+j}^i - 1}{\mu_{t+j}}\right]\psi\left(\frac{\mu_{t+j}^i}{\mu_{t+j}}\right)\frac{Y_{t+j}}{I_{t+j}}\prod_{s=1}^{j-1}g\left(\frac{\mu_{t+s}^i}{\mu_{t+s}}\right) = 0 \quad (8.5)$$

At a symmetric equilibrium where all firms charge the same price, each has a an equal share m^i equal to one, and g equals one in all periods. So the expectation term in (8.5) is equal to X_t/I_t where

$$X_t = I_t E_t \sum_{j=1}^{\infty} \alpha^j \frac{q_{t+j}}{q_t} \left(\frac{\mu_{t+j} - 1}{\mu_{t+j}}\right) \frac{Y_{t+j}}{I_{t+j}} \tag{8.6}$$

Note that X_t can be interpreted as the aggregate profits expected in the future by all the existing firms. Using (8.6), equation (8.5) can be transformed so that μ_t is

$$\mu_t = \mu(X_t/Y_t) \equiv \frac{\psi'}{1 + \psi' + g'(1)X_t/Y_t}. \tag{8.7}$$

Because ψ' and $g'(1)$ are both negative, the derivative of μ with respect to X/Y is negative. An increase in X_t means that profits from future customers are high so that each firm lowers its price in order to increase its market share. An increase in Y_t means that profits stem mostly from current sales so that increasing market share is relatively unimportant. The result is that firms raise their price.

Equation (8.7) replaces the exogenous stochastic process (7.3) that we employed in section 7. The other equations of section 7, however, continue to hold.

8.2 The Implicit Collusion Model

The model in this section is a simplified presentation of Rotemberg and Woodford (1992) which is itself based on Rotemberg and Saloner (1986). In this model, there are two levels of aggregation needed to go from individual products to aggregate output. First, there is an aggregator function like (1.1) that gives total output as a function of the output of a measure of industries. The output of each industry is itself given by a homogeneous of degree one aggregator function which depends on the output of n constituent firms. The goods produced by each of the n firms that constitute a particular industry are very good substitutes and, to prevent what would seem the inevitable fall in price until price is close to marginal cost, the firms in each industry collude implicitly. Collusion is implicit in the sense that there is no enforceable cartel contract, there exists only an implicit agreement that firms that deviate from the collusive understanding will be punished.

The firms in each industry, even when acting in concert, take other industries' prices, the level of aggregate demand, and the level of marginal cost as given. We will consider symmetric equilibria and the profitability of deviating from this equilibrium by either a single firm or by an industry as a whole. We thus consider the demand for firm i in industry j at t when its price corresponds to an inefficiency wedge of μ_t^{ij}, all other firms in its industry charge a price which corresponds to μ_t^j and all firms in other industries charge a price which corresponds to μ_t. Given the homogeneity of demand, we can write the demand faced by firm i in industry j as

$$q_t^{ij} = D^i\left(\frac{\mu_t^{ij}}{\mu_t}, \frac{\mu_t^j}{\mu_t}\right)\frac{Q_t}{I_t} \qquad D^i(1,1) = 1/n. \tag{8.8}$$

Using the same substitution that led to (8.3), profits for this firm equal

$$\pi_t^{ij} = \frac{\mu_t^{ij} - 1}{\mu_t} D^i\left(\frac{\mu_t^{ij}}{\mu_t}, \frac{\mu_t^j}{\mu_t}\right)\frac{Y_t}{I_t}. \tag{8.9}$$

If each firm existed for only one period, it would maximize (8.9) with respect to its own markup treating the markups of all other firms as given. The resulting Bertrand equilibrium would have relatively low prices and low profits. If the firms in an industry charge more than the Bertrand price, individual firms would benefit from undercutting the industry's price. Higher prices, with their attendant higher profits, can only be sustained as a subgame perfect equilibrium if deviators are punished after a deviation. If firms interact repeatedly and have an infinite horizon, there are many equilibria of this type and these differ in the price

41

that is charged in equilibrium.

We assume that firms succeed in implementing that symmetric equilibrium that is jointly best for them. That is, their implicit agreement maximizes the present discounted value of expected equilibrium profits for each firm in industry j, taking as given the stochastic processes for $\{\mu_t\}$, $\{\lambda_t\}$ and $\{Y_t/I_t\}$. As shown by Abreu (1986), this requires that the punishment for any deviation be as severe as possible. Because of the possibility of exit, the voluntary participation of the firm that is being punished precludes it earning an expected present value lower than zero after a deviation. This leads us to assume that a deviator earns a present discounted value of zero after his deviation. Sufficient conditions for this punishment to be feasible and subgame perfect are given in Rotemberg and Woodford (1992).

Because the punishment is independent of the size of the deviation, a deviating firm sets its price at t to maximize (8.9). Let X_t^j denote, by analogy to (8.5), the expected present discounted value of the profits that each firm in industry j can expect to earn in subsequent periods if there are no deviations. Then, if the expected present value of profits after a deviation equals zero, firms in industry j will not deviate as long as

$$\max_{\mu_t^{ij}} \pi_t^{ij} \leq \pi_t^j + X_t^j \tag{8.10}$$

where π_t^j is the value of π_t^{ij} when firm i charges the same price as the other firms in its industry. We consider the case where the incentive compatibility constraint (8.10) is always binding. Thus, firms are indifferent between the additional profits from deviating in the present and the future loss of X. Since X is what firms who deviate give up, α can be given a different interpretation. The quantity $(1-\alpha)$ remains the probability that sales will be independent of the history of prices. This can now mean that $(1-\alpha)$ is the probability that the collusive arrangement is renegotiated and a firm that has deviated in the past and has exited to avoid punishment can reenter. [31]

At a symmetric equilibrium, all industries have the same markup, so that each firm sells Q_t/nI_t and X_t^j equals X_t/nI_t. Assuming (8.10) holds with equality, (8.9) implies that

$$\max_{\rho} \left[\rho - \frac{1}{\mu_t}\right] D(\rho, 1) \frac{Y_t}{I_t} = \left[1 - \frac{1}{\mu_t}\right] \frac{Y_t}{nI_t} + \frac{X_t}{nI_t} \tag{8.11}$$

---

[31] In Rotemberg and Woodford (1992) we give conditions under which a deterministic steady state exists in which (8.10) is always binding. We also show that, for small enough stochastic shocks, there continues to exist a perturbed equilibrium in which (8.10) always binds. This case is clearly most plausible if X_t^j is not too large a multiple of a single period's profits, which is to say if α is considerably less than one.

42

where ρ represents the relative price chosen by the deviating firm. Equation (8.11) can be solved for μ_t, yielding once again

$$\mu_t = \mu(X_t/Y_t) \tag{8.12}$$

In this case, however, the derivative of μ with respect to X/Y is positive. The reason is that an increase in X/Y raises the size of the punishment (the foregone profits represented by X_t) relative to the size of current sales (as represented by Y_t). It thus allows the firms in each oligopolistic industry to charge higher markups without fearing deviations. The theoretical model also implies an upper bound on the elasticity of the markup with respect to X/Y. We show in Rotemberg and Woodford (1992) that this upper bound is equal to $\mu - 1$.

The two models we have presented both make the markup depend on X_t/Y_t. Because the sign of the effect of X/Y differs, they are empirically distinguishable and this is pursued in Rotemberg and Woodford (1991). Here we are interested in seeing what this dependence of the markup on X/Y implies about the effect of aggregate demand.

8.3 The Response of the Model to Exogenous Changes in Government Purchases

To compute the responses of the model to exogenous changes in aggregate demand we have to linearize (8.6) and (8.12) around their steady states. For this purpose we assume that $\bar{X}_t$ is given by $\frac{X_t}{z_t N_t}$ while $\bar{X}$ and $\bar{\mu}$ give the steady state level of expected profitability and of the markup respectively. We then let $\hat{X}_t$ and $\hat{\mu}_t$ represent the logarithmic deviations of $\bar{X}_t$ and μ_t from their respective steady state values. In terms of these values, and assuming a constant number of firms I_t, equations (8.6) and (8.12) become approximately

$$\hat{X}_t = E_t\left\{\hat{\lambda}_{t+1} - \hat{\lambda}_t + \left(\frac{r-g}{1+r}\right)\left(\frac{1}{\mu-1}\hat{\mu}_{t+1} + \hat{Y}_{t+1}\right) + \left(\frac{1+g}{1+r}\right)\hat{X}_{t+1}\right\} \tag{8.13}$$

$$\hat{\mu}_t = \epsilon_\mu(\hat{X}_t - \hat{Y}_t) \tag{8.14}$$

The linearized model now consists of (2.15), (2.16'), (2.17), (2.18), (2.19'), (8.13) and (8.14). Since we are interested in the effect of temporary changes in government spending, we shall assume that the number of firms stays constant so that $\hat{I}_t$ is zero in these equations. We can solve (2.16') and (2.17) for $\hat{H}_t$ and $\hat{w}_t$ as functions of $\hat{\lambda}_t, \hat{K}_t$, and $\hat{\mu}_t$. Substitution into (2.15) gives $\hat{Y}_t$, and substitution into (8.14) then allows us to solve for $\hat{X}_t$, both again as functions of the same three state variables. We can then eliminate these four

state variables from the remaining three equilibrium conditions, obtaining a system of difference equations of the form (2.20) except that it has three endogenous variables $\{\lambda, K, \mu\}$ instead of only $\{\lambda, K\}$.

Assuming once again that the stochastic process for government purchases takes the form given in (4.1) with ρ^G equal to 0.9, we can compute the economy's response to the shock ν_t^G. These responses are computed using the parameters of Table 1 with the average inefficiency wedge μ set equal to 1.4. We have set α equal to 0.9 because this is consistent with (8.10) holding as an equality with about ten firms. Finally, we have looked at two values of ϵ_μ. In the first case, it has the sign suggested by the customer market model and equals -1. In the second, it has the sign implied by the implicit collusion model and equals 0.39. We chose 0.39 because it is just below the upper bound of 0.4 which applies for our average markup.

The response of output and hours is displayed in Figure 7 while Figure 8 displays the response of real wages. We see in figure 7 that the response of output and hours is most pronounced in the case of ϵ_μ equal .39 and least pronounced in the case where it equals -1. The constant markup case of Figure 1 yields an intermediate answer. The reason is easy to understand. An increase in military purchases requires that individuals postpone their consumption so that interest rates rise. This lowers X. Since the increase in military purchases also raises Y, X/Y unambiguously falls. Thus in the customer market model markups rise, which by (7.2) lowers the demand for labor at any given real wage. The result is that output and the labor input do not rise as much as in the case where the markup is constant. By contrast, in the implicit collusion model the fall in X/Y lowers markups which raises the demand for labor. This accentuates the increase in output.

The difference between the models' implications for the demand for labor is even more apparent in Figure 8 where we plot the responses of the real wage w_t to changes in ν_t^G. We see that the customer market model accentuates the reduction in real wages that we displayed in Figure 2 for the constant markup case. The reason is that, as we saw, the reduction in X/Y lowers the demand for labor at any given real wage. By contrast, in the implicit collusion model with ϵ_μ equal to .39 the real wage actually rises, albeit only slightly. This occurs because the fall in X/Y is so pronounced that the increased demand for labor from the fall in the markup actually exceeds the increase in labor supply. In Rotemberg and Woodford (1992) we obtained much more pronounced rises in real wages by assuming a smaller elasticity of labor supply ϵ_{Hw}.

The conclusion from this exercise is that the implicit collusion model can generate procyclical movements

44

in real wages in response to changes in government purchases of goods. Because the key ingredient in generating these responses is the increase in interest rates that leads to a fall in X/Y, similar responses would be observed to other shocks that increase aggregate demand such as increases in firms desire to invest because of changing perceptions of future profitability.

In Rotemberg and Woodford (1992) we considered the actual response of real wages to changes in military purchases. We showed that, indeed, real wages have tended to rise in the United States following increases in military purchases. That paper also considers alternative explanations for the finding such as the fact that the government raises the size of its military personnel at the same time as it increases its national defense purchases of produced goods. It shows that, even after this is taken into account, reductions in markups of the sort implied by the implicit collusion model are needed to explain the reaction of real wages. In Rotemberg and Woodford (1993) we demonstrated that increases in oil prices tend to raise markups in this type of model. We showed that this is consistent both with the large size of output reductions that follow actual oil price increases and with the failure of the value added deflated real wage to rise on these occasions.

8.4   A Model with Composition Effects

In this subsection we describe an example of a model (that of Gali (1991)) in which markups are affected by the composition of aggregate demand. In Gali's model both households and firms purchase the entire range of differentiated goods, but (unlike our assumption in section 1) their aggregator functions f_t are different. In particular, the elasticity of substitution between different goods, again evaluated at the case of uniform prices for all goods, is different for the two types of purchasers. In all other respects, the model is one of static monopolistic competition like that described in sections 1-2.

Profit maximization by price-setting firms now implies a markup that depends on the share θ_t of aggregate demand that consists of demand by firms so that

$$\gamma_t = \gamma(\theta_t) \equiv \left[1 + \left(\theta_t D_f'(1) + (1 - \theta_t) D_h'(1)\right)^{-1}\right]^{-1}$$

where D_f' and D_h' indicate the elasticities of demand of firms and households respectively. This reduces to (1.7) in the case that $D_f = D_h = D$. In the case argued by Gali to be of greatest interest, that in which $D_f'(1) < D'h(1) < -1$, one finds that γ is a monotonically decreasing function of θ_t. Because the inefficiency

45

wedge μ_t remains the same function of γ_t as before (given by (1.8)), μ_t is also a monotonically decreasing function of θ. Finally, identifying

$$\theta_t = \frac{I_t + M_t}{Q_t} = s_M + (1 - s_M)s_{It}$$

where s_{It} is the share of investment spending in value added, we obtain

$$\mu_t = \mu(s_{It}) \tag{8.15}$$

where $\mu(s_I)$ is a monotonically decreasing function. Gali's model is then essentially the model of section 7 with (8.15) added in place of the exogenous markup process.

In this model, shocks that affect the composition of aggregate demand affect equilibrium markups. This introduces a channel through which some kinds of increases in aggregate demand may have additional expansionary effects (for example an increase in investment due to a change in tax incentives). But, it is not a model where increases in aggregate demand as such have an expansionary effect through a change in desired markups. It depends entirely on the category of demand increase. For example, if the government uses the same aggregator as households (an issue not discussed in Gali) an increase in government purchases will increase markups. This would imply that government purchases are even less expansionary than in the constant markup model.

The predictions of such a model depend critically upon the elasticities of demand of different purchasers. Gali provides no direct evidence on this although he does show that the average markup (measured using a method similar to that of Rotemberg and Woodford (1991)) shows a strong negative association with the investment share, even when one controls for the levels of aggregate output and hours. (Thus the association is not simply a reflection of the fact that the average markup is countercyclical).

Another consequence of Gali's model is that, for some choices of the parameters, aggregate fluctuations can occur in equilibrium in the absence of exogenous shocks. He shows that low expected future markups increase current investment demand. The reason is that the low expected markups raise the expected demand for labor and thereby increases the expected marginal product of capital for any given capital stock. In addition, the reduction in expected markups lowers the wedge between the marginal product of capital and its user cost. This means that low expected future markups raise current investment which, in turn, reduces current markups. For some parameter values this effect is so strong as to make the expectation of

46

low markups self-fulfilling. Gali demonstrates this possibility, and analyzes the character of the fluctuations that can exist due to self-fulfilling expectations, using techniques like those discussed in section 6.

9   Conclusions

In this paper we have shown that imperfect competition matters; it affects the way in which the economy responds to a great variety of shocks. It is thus not possible to ignore departures from perfect competition if one wants quantitatively accurate assessments of the importance of various disturbances. The reason imperfect competition matters so much is that it affects the relationship between the marginal product of labor and the real wage. It thus affects the relationship between output, the labor input and the wage. Because so many of the puzzles in macroeconomics, including Okun's law and the Dunlop-Tarshis observation relate to these three variables, imperfect competition is central to the concerns of business cycle analysis.

We have also shown that incorporation of imperfectly competitive product markets into standard dynamic models of aggregate fluctuations is relatively simple. The models that we have presented can still be analyzed numerically using standard techniques. At most, a small increase in the state space is required. We have also shown that imperfect competition need introduce only a small number of additional parameters to be calibrated and that both empirical evidence and theoretical considerations can be used to bound the plausible range of variation in the new parameters. The degree to which this direction of generalization of standard models can improve the ability of such models to explain observed aggregate fluctuations is the subject of continuing research.

47

10   References

Abreu, Dilip:"Extremal Equilibria of Oligopolistic Supergames," *Journal of Economic Theory*, 39, 1986, 191-225.

Barro, Robert J.:"Output Effects of Government Purchases," *Journal of Political Economy*, 89, 1981.

Basu, Susanto:"Intermediate Goods, Menu Costs, and Business Cycles," mimeo, 1992.

Baxter, Marianne and Robert King :"Productive Externalities and Cyclical Volatility," mimeo, 1990.

Beggs, Alan and Paul Klemperer: "Multi-Period Competition with Switching Costs," *Econometrica*, 60. May 1992, 651-66.

Benabou, Roland:"Inflation and Markups: Theories and Evidence from the Retail Trade Sector", *European Economic Review*, 1992.

Benassy, J. P.:"Monopolistic Competition," in W. Hildenbrand and H. Sonnenschein eds. *Handbook of Mathematical Economics vol. 4*, Amsterdam: North Holland, 1991.

Benhabib, J. and R. E. A. Farmer:"Indeterminacy and Increasing Returns," mimeo, 1992.

Bils, Mark: "Pricing in a Customer Market," *Quarterly Journal of Economics*, November 1989.

Blanchard, Olivier Jean and Charles M. Kahn: "The Solution of Linear Difference Models under Rational Expectations," *Econometrica*, 48, 1305-13, July 1980.

Bils, Mark:"The Cyclical Behavior of Marginal Cost and Price," *American Economic Review*, 77, December 1987, 838-57.

Burnside, Craig, Martin Eichenbaum and Sergio Rebelo:"Labor Hoarding and the Business Cycle," *Journal of Political Economy*, 101, 1993, 245-273. 1990.

Chamberlin, E.:*The Theory of Monopolistic Competition*, Cambridge, Mass., Harvard University Press. 1933.

Chappell, Jr. H.W. and R.P. Wilder: "Multiproduct Monopoly, Regulation and Firm Costs: Comment," *Southern Economic Journal*, 52, 1168-74, 1986.

Christensen, L. and W. Greene: "Economies of Scale in U.S. Power Generation," *Journal of Political Economy*, 655-76, 1976.

Dixit, Avinash and Joseph Stiglitz: "Monopolistic Competition and Optimum Product Diversity," *American Economic Review*, 67, 297-308, 1977.

Domowitz, Ian, R. Glenn Hubbard, and Bruce C. Petersen, "Market Structure and Cyclical Fluctuations in U.S. Manufacturing," *Review of Economic Statistics*, February 1988.

Eichenbaum, Martin: "Real Business Cycle Theory: Wisdom or Whimsy," *Journal of Economic Dynamics and Control*, 15, 1991, 607-26.

Evans, Charles: "Productivity Shocks and Real Business Cycles," mimeo, 1990

Farmer, R. E. A. and J.T. Guo: "Real Business Cycles and the Animal Spirits Hypothesis," mimeo, 1993.

Farrell, Joseph and Carl Shapiro: "Dynamic Competition with Switching Costs," *Rand Journal of Economics*, 19, Spring 1988, 123-37.

Gali, Jordi: "Monopolistic Competition, Business Cycles, and the Composition of Aggregate Demand," mimeo, 1992.

Gottfries, Nils: "Price Dynamics of Exporting and Import-Competing Firms," *Scandinavian Journal of Economics*, 88, 417-436, 1986.

Guesnerie, Roger and Michael Woodford: "Endogenous Fluctuations," in Jean Jacques Laffont, ed. *Advances in Economic Theory: Proceedings of the 6th World Congress of the Econometric Society Vol 2*, Cambridge University Press, Cambridge, 1992.

Hairault, Jean-Olivier and Franck Portier: "Money, New-Keynesian Macroeconomics and the Business Cycle," mimeo, 1992, *European Economic Review*, forthcoming

Hall, Robert E.: "The Relation Between Price and Marginal Cost in U.S. Industry," *Journal of Political Economy*, 96, October 1988, 921-48.

——: "Invariance Properties of Solow's Productivity Residual," in Peter A. Diamond ed.Growth, Productivity and Unemployment, Essays to Celebrate Bob Solow's Birthday, Cambridge: MIT Press, 1990.

Hammour, Mohamad. L. : "Increasing Returns and Endogenous Business Cycles," ch 1 of Ph. D. Dissertation at Massachusetts Institute of Technology, 1988

—— : "Overhead Costs and Economic Fluctuations," Discussion Paper 547, Columbia University Working paper, 1991.

Hansen, Lars Peter and Thomas Sargent: "Formulating and Estimating Dynamic Linear Rational Expectations Models," *Journal of Economic Dynamics and Control*, 2, 7-46, February 1980.

Hornstein, Andreas:"Monopolistic Competition, Increasing Returns to Scale and the Importance of Productivity Changes," *Journal of Monetary Economics*, 31, 1993, 299-316.

King, Robert G., Charles I.Plosser and Sergio Rebelo: "Production, Growth and Business Cycles: I. The Basic Neoclassical Model," *Journal of Monetary Economics*, 21, March 1988, 195-232.

———"Production, Growth and Business Cycles: II. New Directions," *Journal of Monetary Economics*. 21, March 1988, 309-342.

Klemperer, Paul D.:"Markets with Consumer Switching Costs,"*Quarterly Journal of Economics, 102*. May 1987, 375-94.

Kydland, Finn E. and Edward C. Prescott:"Time to Build and Aggregate Fluctuations," *Econometrica*. 50, (November 1982): 1345-70.

Mankiw, N. Gregory:"Imperfect Competition and the Keynesian Cross," *Economic Letters*. 26, 1988, 7-13.

Morrison, Catherine J.:"Market Power, Economic Profitability and Productivity Growth Measurement: An Integrated Structural Approach," *NBER Working Paper 3355*, May 1990.

Panzar, John C.:"Technological Determinants of Firm and Industry Structure," in R. Schmalensee and R.D. Willig, eds., *Handbook of Industrial Organization*, North-Holland, Amsterdam, 1989.

Phelps, Edmund S.:"Consumer Demand and Equilibrium Unemployment in a Working Model of the Customer-Market Incentive-Wage Economy," *Quarterly Journal of Economics*, 107, August 1992. 1003-32

Phelps, Edmund S. and Sidney G. Winter:"Optimal Price Policy under Atomistic Competition," in E. Phelps ed., *Microeconomic Foundations of Employment and Inflation Theory*, W. W. Norton and Co., New York, 1970.

Plosser, Charles I.:"Understanding Real Business Cycles," *Journal of Economic Perspectives*, 3, Summer 1989, 51-78.

Rotemberg, Julio J. and Garth Saloner:"A Supergame-Theoretic Model of Price Wars during Booms," *American Economic Review*, 76, June 1986, 390-407.

Rotemberg, Julio J. and Michael Woodford:"Markups and the Business Cycle," *Macroeconomics Annual*. 1991, 63-128.

—— and ——:"Oligopolistic Pricing and the Effects of Aggregate Demand on Economic Activity," *Journal of Political Economy*, 100, December 1992, 1153-1207.

—— and ——:"Imperfect Competition and the Effects of Energy Price Increases on Economic Activity," mimeo, 1993

Silvestre, Joaquim: "The Market-Power Foundations of Macroeconomic Policy," *Journal of Economic Literature*, 31, March 1993, 105-141.

Startz, Richard: "Monopolistic Competition as a Foundation for Keynesian Macroeconomic Models," *Quarterly Journal of Economics*, 104, November 1989, 737-52.

Summers, Lawrence: "Taxation and Corporate Investment: A q Theory Approach," *Brookings Papers on Economic Activity*, January 1981, 67-127.

Tellis, Gerald J.: "The Price Elasticity of Selective Demand: A Meta-Analysis of Econometric Models of Sales," *Journal of Marketing Research*, 25, November 1988, 331-41.

Woodford, Michael: "Stationary Sunspot Equilibria: The Case of Small Fluctuations around a Deterministic Steady State," unpublished, University of Chicago, 1986.

—— :"Expectations, Finance and Aggregate Instability," in M. Kohn and S.-C. Tsiang eds., *Finance Constraints, Expectations and Macroeconomics*, Oxford: Oxford University Press, 1988

—— (1991),"Self-Fulfilling Expectations and Fluctuations in Aggregate Demand," in N. G. Mankiw and D. Romer eds.,*New Keynesian Economics*, Cambridge: MIT Press.

Yun, Tack: "Nominal Price Rigidity, Endogenous Money and Business Cycles," University of Chicago, mimeo, 1993.

Table 1
The Calibrated Parameters

Parameter	Defined by	Values	Description
γ_z		1.004	steady state quarterly growth rate of technology
γ_N		1.004	steady state quarterly growth rate of population
s_C	$\frac{\tilde{C}}{Y}$	0.587	share of private consumption expenditure in value added
s_G	$\frac{\tilde{G}}{Y}$	0.117	share of government purchases of goods in value added
s_I	$(g+\delta)\frac{\tilde{K}}{Y}$	0.296	share of private investment expenditure in value added
δ		0.025	rate of depreciation of capital stock (per quarter)
s_H	$\frac{F_H \tilde{H}}{F}$	0.58	share of labor costs in total costs
r	$\frac{F_K Q_V}{\mu}-\delta$ or $\gamma_z^\sigma \beta^{-1}$-1	0.016	steady state real rate of return (per quarter)
$1/\sigma$		1	Intertemporal elasticity of substitution of consumption holding hours worked constant
ϵ_{Hw}		4	Intertemporal elasticity of labor supply
ϵ_{KH}	$\frac{F_K F_H}{F_{KH} F}$	1	elasticity of substitution between capital and hours
μ		1,1.4	steady state value-added markup (efficiency wedge)
κ		0.02	rate at which entry adjusts to changes in technology
ϵ_μ	$\frac{\tilde{X}\mu'}{Y\mu}$	-1,.39	elasticity of the markup (endogenous markup model)
α		0.9	probability that future sales depend on price history
ρ^G		0.9	serial correlation of government purchases
ρ^μ		0,0.9	serial correlation of μ (exogenous markup model)

Except for population growth, parameters displayed above μ take the values in King, Plosser and Rebelo (1988a).

TABLE 2

THE EFFECT OF TECHNOLOGY SHOCKS
PREDICTED VERSUS ACTUAL SECOND MOMENTS

μ	variance of measured total factor productivity
1.0000	2.1653
1.2000	1.5132
1.4000	1.1510
1.6000	0.9365
1.8000	0.8038
2.0000	0.7192

OUTPUT STATISTICS

μ	(1)	(2)	(3)	(4)
1.0000	1.3216	1.2263	0.5286	-0.3279
1.2000	1.3216	0.9448	0.6332	-0.1940
1.4000	1.3216	0.7745	0.7632	-0.1635
1.6000	1.3216	0.6685	0.8980	-0.1853
1.8000	1.3216	0.6016	1.0288	-0.2336
2.0000	1.3216	0.5596	1.1522	-0.2952

HOURS STATISTICS

μ	(1")	(2")	(3")	(4")
1.0000	0.8282	0.1833	0.9365	-0.3520
1.2000	0.8282	0.0271	0.8658	-0.0780
1.4000	0.8282	0.0010	0.8383	-0.0134
1.6000	0.8282	0.0076	0.8055	0.0182
1.8000	0.8282	0.0218	0.7644	0.0507
2.0000	0.8282	0.0373	0.7187	0.0872

----------------------- --

(1) var(dY(t))
(2) var(dYth(t))
(3) var(dY(t)-dYth(t))
(4) rho=2*cov(dYth,dY- dYth)/(var(dYth)+var(dY-dYth))
(1") var(dH(t))
(2") var(dHth(t))
(3") var(dH(t)-dHth(t))
(4") rho=2*cov(dHth,dH-dHth)/(var(dHth)+var(dH-dHth))

TABLE 3

VARIANCE OF THE LOG DIFFERENCE OF OUTPUT, HOURS AND WAGE

VARIANCE OF OUTPUT CHANGES DIVIDED BY VARIANCE OF MARKUP CHANGES

μ	$rho^{\mu}=0$	$rho^{\mu}=.3$	$rho^{\mu}=.6$	$rho^{\mu}=.9$
1.2000	0.9269	0.9363	0.9580	1.0577
1.4000	1.2285	1.2295	1.2317	1.2430
1.6000	1.5626	1.5494	1.5210	1.4145
1.8000	1.9262	1.8926	1.8216	1.5726
2.0000	2.3166	2.2559	2.1303	1.7180

VARIANCE OF HOURS CHANGES DIVIDED BY VARIANCE OF MARKUP CHANGES

μ	$rho^{\mu}=0$	$rho^{\mu}=.3$	$rho^{\mu}=.6$	$rho^{\mu}=.9$
1.2000	2.0055	2.0244	2.0672	2.2589
1.4000	1.9681	1.9676	1.9664	1.9597
1.6000	1.9316	1.9128	1.8720	1.7159
1.8000	1.8960	1.8601	1.7840	1.5151
2.0000	1.8613	1.8094	1.7020	1.3480

VARIANCE OF WAGE NORMALIZED BY VARIANCE OF MARKUP

μ	$rho^{\mu}=0$	$rho^{\mu}=.3$	$rho^{\mu}=.6$	$rho^{\mu}=.9$
1.2000	0.1498	0.1484	0.1455	0.1367
1.4000	0.1521	0.1533	0.1562	0.1704
1.6000	0.1546	0.1584	0.1668	0.2030
1.8000	0.1571	0.1634	0.1775	0.2340
2.0000	0.1596	0.1685	0.1881	0.2634

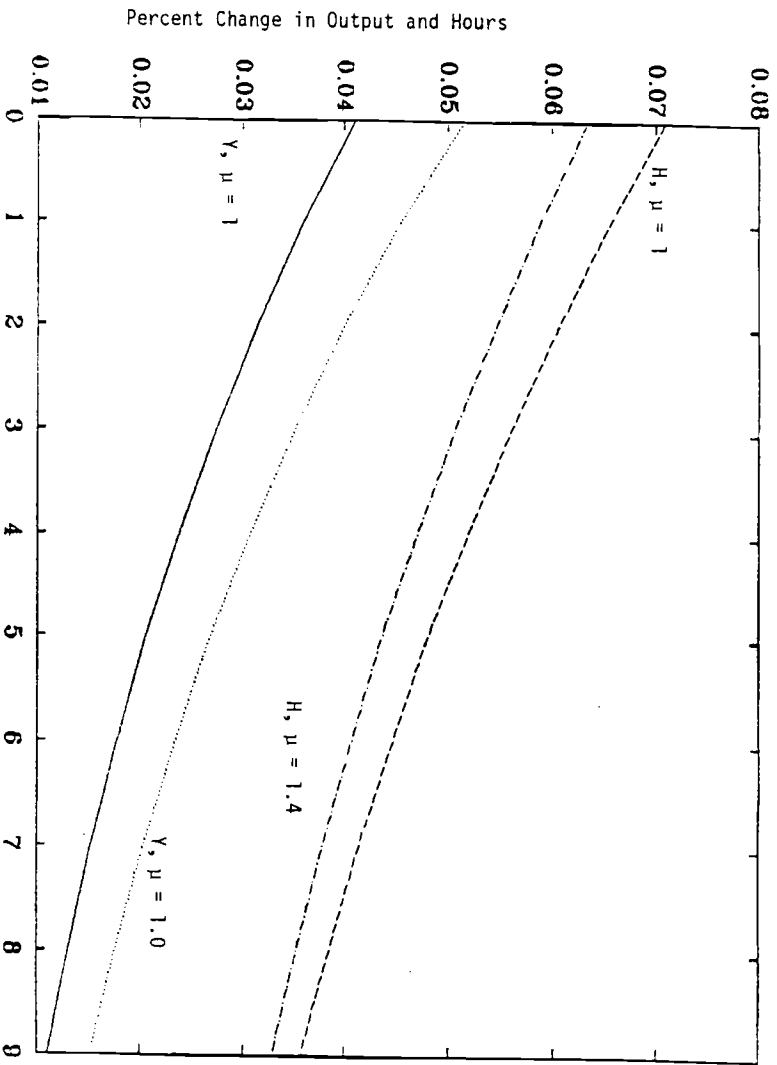---
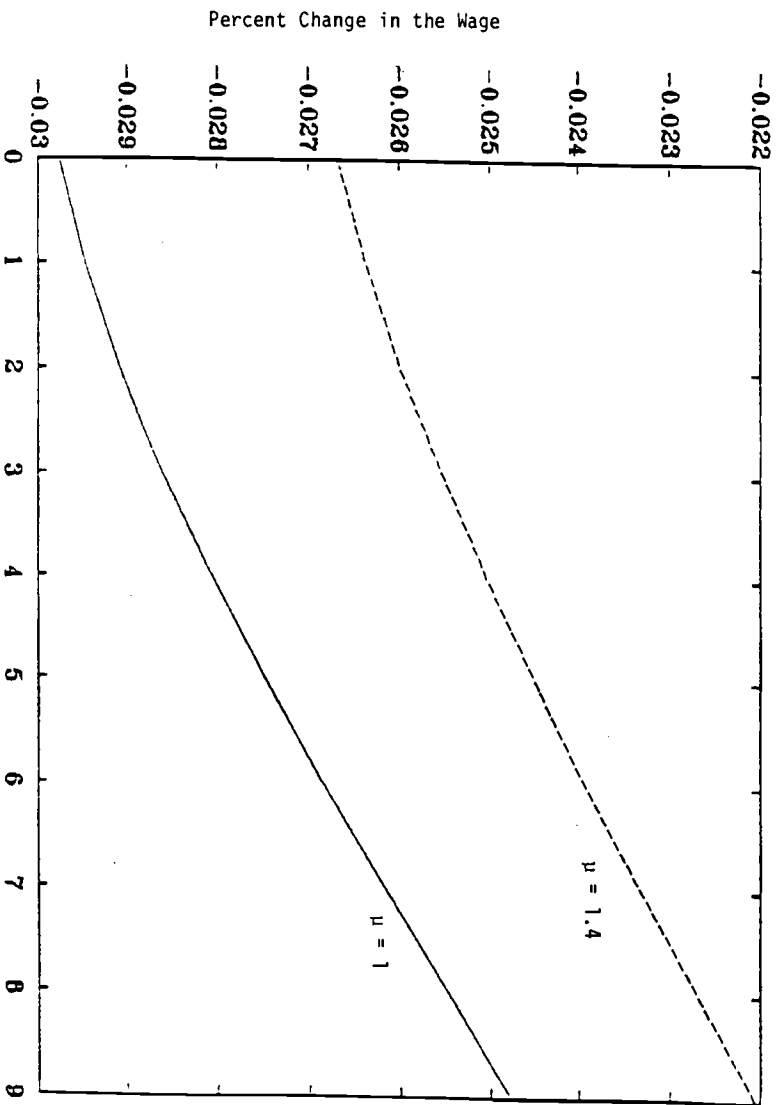
FIGURE 1. The Response of Output and Hours to a One Percent Shock to Government Purchases

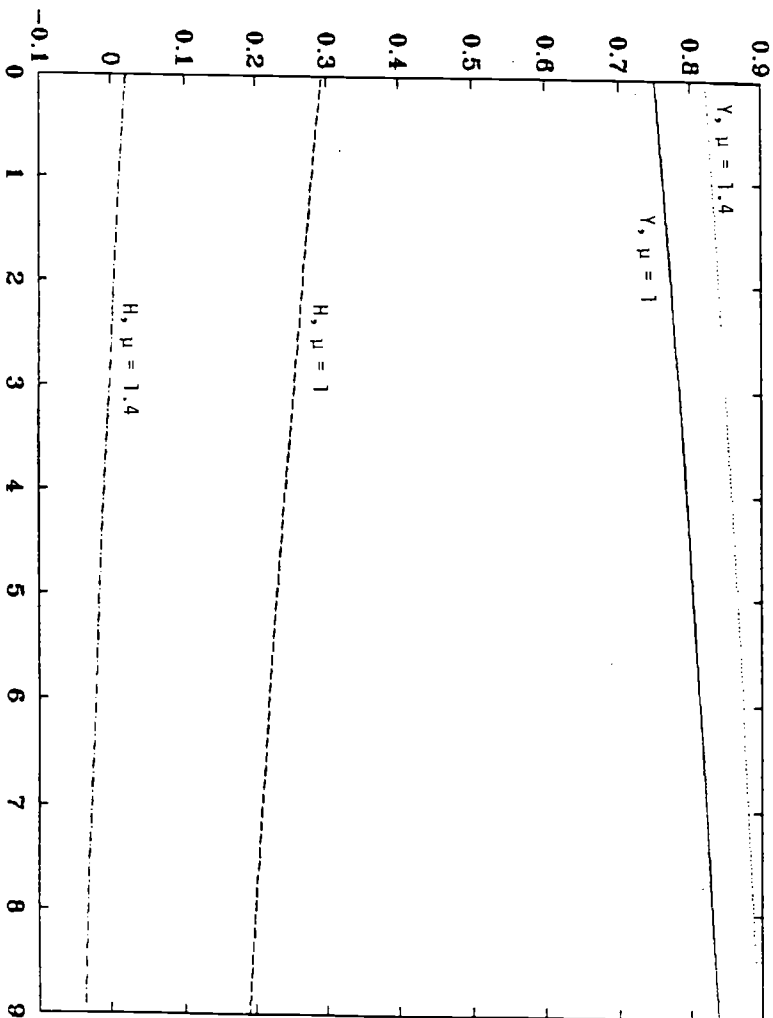FIGURE 2. The Response of the Real Wage to a One Percent Shock to Government Purchases

FIGURE 3. The Response of Output and Hours to a One Percent Technology Shock
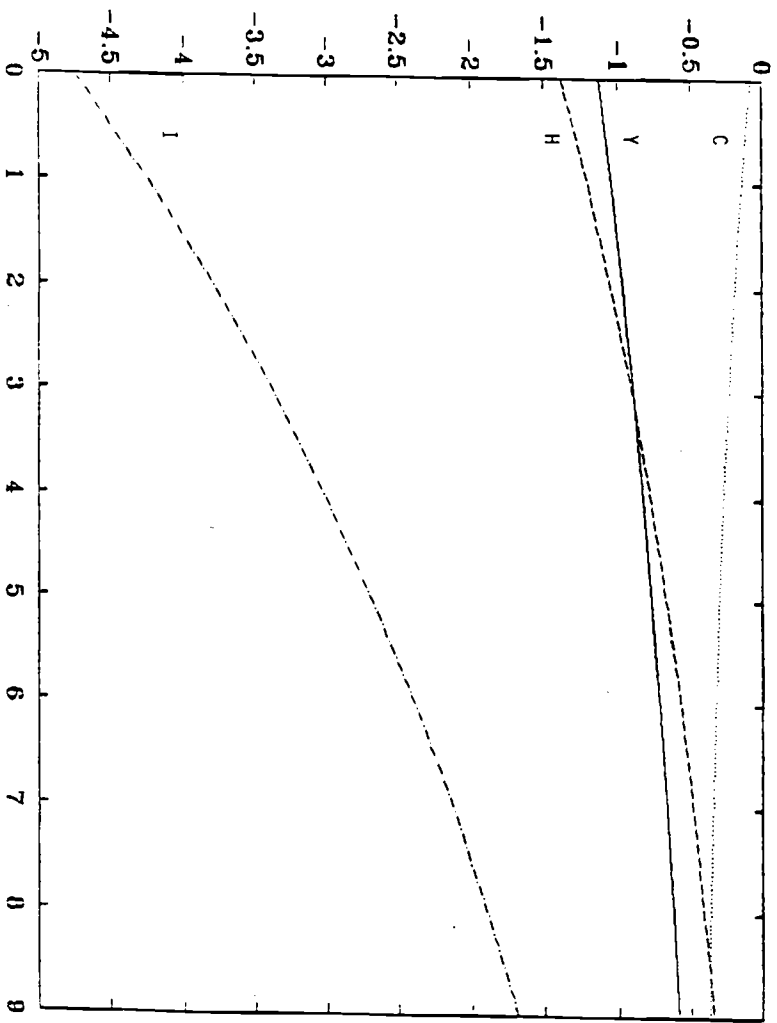
Percent Change in Output and Hours

FIGURE 4. The Response of Output, Hours, Consumption and Investment to a One Percent Markup Shock, $\rho^\mu = 0.9$

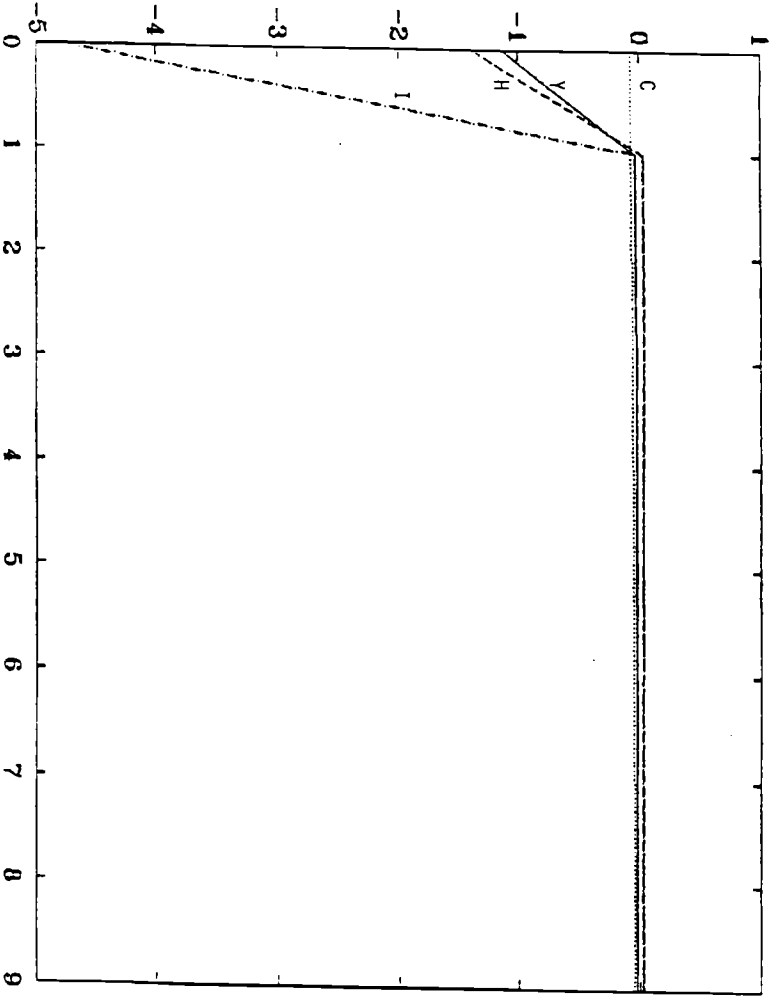Percent Change in Output, Hours, Consumption and Investment

FIGURE 5. The Response of Output, Hours, Consumption, and Investment to a Markup Shock, $\rho^* = 0$
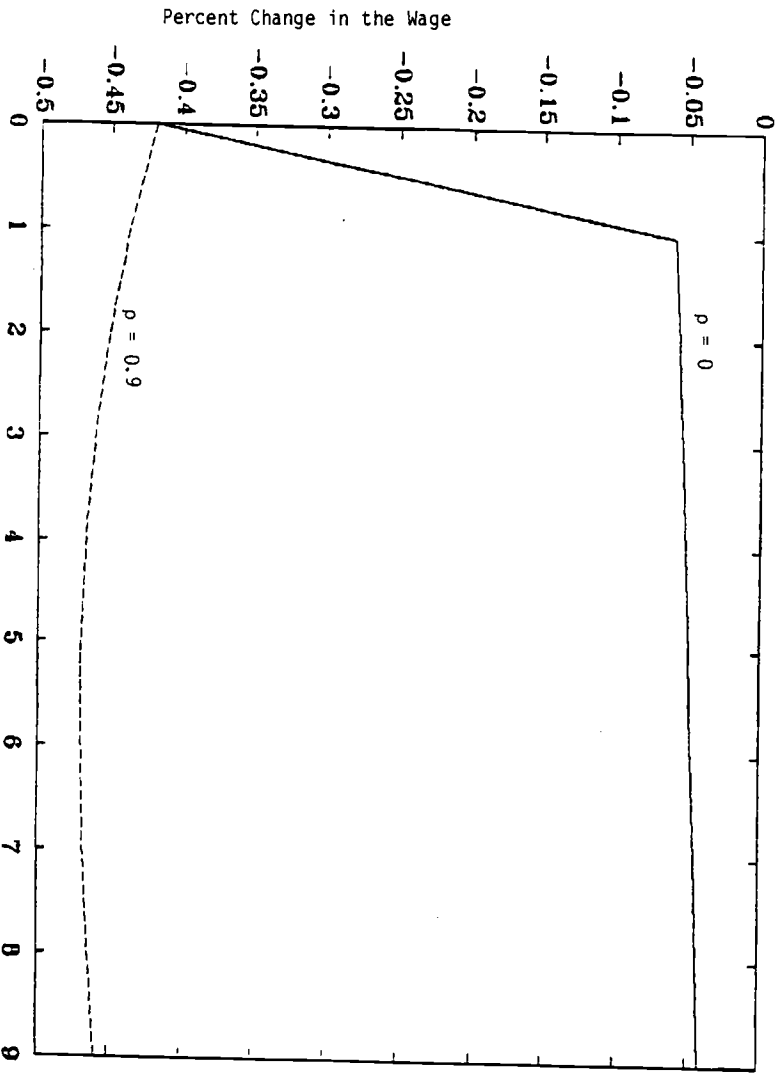
FIGURE 6. The Response of the Real Wage to Markup Shocks.
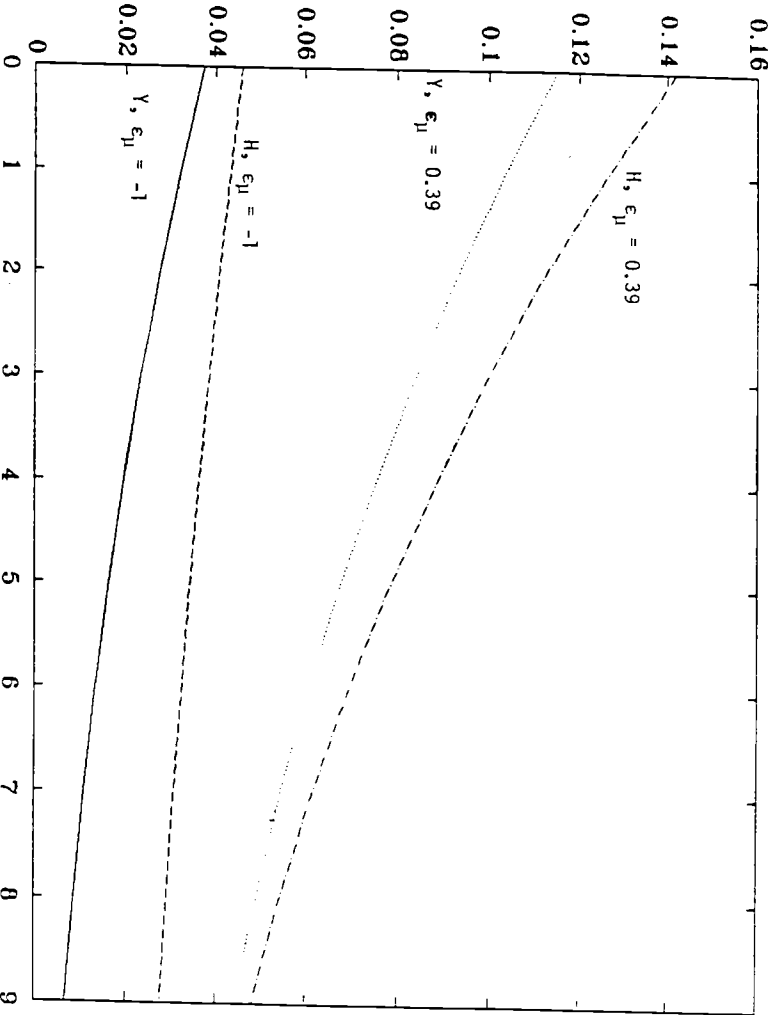
Percent Change in Output and Hours

FIGURE 7. The Response of Output and Hours to Changes in Government Purchases: The Case of Endogenously Varying Markups.

Y, ϵ_μ = 0.39

H, ϵ_μ = 0.39

Y, ϵ_μ = -1

H, ϵ_μ = -1

FIGURE 8. The Response of the Real Wage to Changes in Government Purchases: The Case of Endogenously Varying Markups.