Volume Title: NBER Macroeconomics Annual 1991, Volume 6

Volume Author/Editor: Olivier Jean Blanchard and Stanley Fischer, editors

Volume Publisher: MIT Press

Volume ISBN: 0-262-02335-0

Volume URL: http://www.nber.org/books/blan91-1

Conference Date: March 8-9, 1991

Publication Date: January 1991

Chapter Title: Markups and the Business Cycle

Chapter Author: Julio J. Rotemberg, Michael Woodford

Chapter URL: http://www.nber.org/chapters/c10982

Chapter pages in book: (p. 63 - 140)

Julio J. Rotemberg and Michael Woodford
SLOAN SCHOOL OF MANAGEMENT, MIT AND NBER/UNIVERSITY OF CHICAGO AND NBER

*Markups and the Business Cycle**

1. *Cyclical Markups and the Transmission of Aggregate Demand Variations to the Labor Market*

Perfectly competitive models of the effects of aggregate demand variations on output and employment have great difficulty generating patterns of comovement among aggregate variables that resemble typical "business cycle" patterns (Barro and King, 1984). We wish to emphasize two such difficulties here, and to argue that an imperfectly competitive model of product markets can avoid these difficulties.

First, price-taking firms and a standard neoclassical production technology imply that, in the absence of shifts in the production function (which are not what we would mean by "aggregate demand" shocks), output and employment fluctuations should be associated with *countercyclical movements in the real wage.* For firms' profit-maximizing labor demand will be determined by the condition

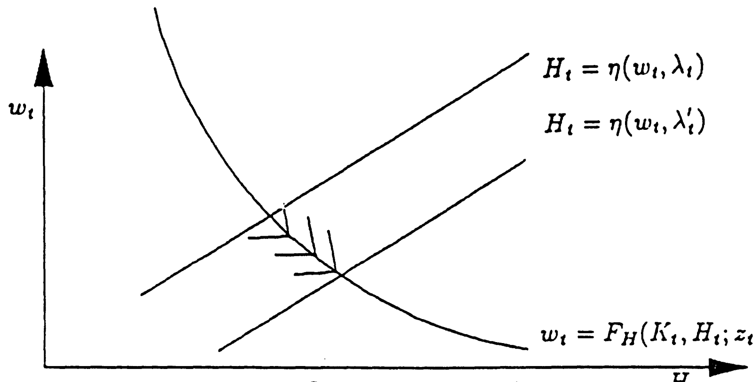
$$F_H(K_t, H_t; z_t) = w_t \tag{1.1}$$

where K_t represents the capital stock, H_t hours of labor demanded, w_t the real wage, and $F(K_t, H_t; z_t)$ represents output given the state of technology z_t. For a given capital stock K_t and state of technology z_t, this implies a downward-sloping demand curve for labor H_t as a function of w_t. Furthermore, an aggregate demand shock at time t (e.g., an increase in government purchases) cannot shift this labor demand curve, since the capital stock is predetermined and the state of technology is exogenous

*We wish to thank Joseph Beaulieu, Roland Benabou, Mark Bils, Alan Blinder, John Cochrane, Stanley Fischer, Robert King, N. Gregory Mankiw, Rob Gertner, and Hal Varian, as well as participants at the NBER Summer Institute for suggestions. We also thank Changyong Rhee for providing us with his data on q, Matthew Shapiro for his two-digit data, and the NSF for research support.

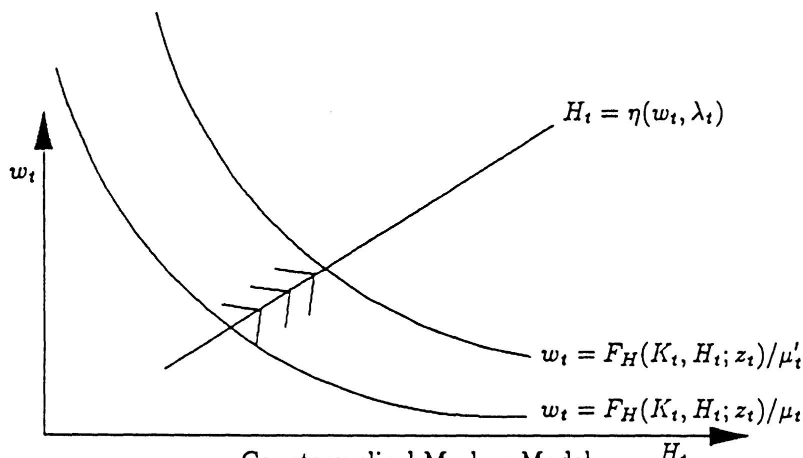
Figure 1

(a)



Competitive Model

(b)



Countercyclical Markup Model

with respect to such shocks. Hence if hours and output are to increase, it must be through a shift *along* this curve, as shown in Figure 1a. But this implies a reduction in the real wage.[1]

Under standard assumptions about the aggregate production technology (an elasticity of substitution between capital and labor near one), the

1. More complicated competitive models are considered in Appendix 1, where we argue that the problems discussed are not easily avoided through the assumption of a more complex technology.

countercyclical movement of the real wage should be substantial: one-third to one-half a percent change for each 1% variation in output. But such countercyclical real wages are not observed; indeed, this embarrassment to the neoclassical theory of labor demand has been noted at least since Dunlop (1938) and Tarshis (1939). Recent studies have generally upheld this finding. In fact, when correction is made for bias resulting from cyclical variation in the composition of the work force (there is a greater proportion of lower-wage and presumably lower-productivity labor hired during booms), many authors find significantly *procyclical* real wage movements (e.g., Stockman, 1986; Kydland and Prescott, 1988; Barsky and Solon, 1989).

A number of other objections may be raised to the use of data on average hourly earnings as a measure of the cost to firms of marginal hours. Most of the corrections implied by these considerations strengthen our argument as well, i.e., they provide further reason to conclude that the typical cyclical behavior of wages is inconsistent with the joint hypothesis of competitive firms and variations in output due to aggregate demand variations. For example, in the presence of convex adjustment costs for changes in the labor input, (1.1) becomes instead

$$F_H(K_t, H_t; z_t) = w_t + \phi_t$$

where ϕ_t represents the adjustment cost of adding an hour of work. But ϕ_t should be positive when hours are increasing (due to current adjustment costs) or higher than they are expected to be in the future (due to expected future adjustment costs), and similarly negative when hours are decreasing or lower than expected to be in the future. Hence $w_t + \phi_t$ should be even more procyclical than the real wage alone, creating an even greater problem. Alternatively, if firms insure their workers by smoothing their wage payments, we would expect that payments to labor in recessions should exceed the social cost of that labor, while labor payments in booms should fall short of the social cost. This too would mean that the true shadow cost of an additional unit of labor is even more procyclical than the measured real wage.

Finally, there is the distinction between straight time and overtime hours. These differ in two respects. First, as stressed by Hansen and Sargent (1988), these may not be perfect substitutes because the use of overtime hours lengthens the period over which capital is utilized. Second, as stressed by Bils (1987), overtime hours command a higher wage. As we discuss more fully in Appendix 2 and summarize in Section 4.3 both of these matter largely because overtime hours rise disproportionately in booms. Insofar as the 2 hours are not perfect substitutes this implies that the marginal product of overtime hours declines sharply in booms, thereby deepening the puzzle of why real wages do not decline. Even if they are perfect substitutes, the disproportionate increase in overtime hours means that marginal hours of labor have a higher overtime component in booms. Insofar as the legally mandated overtime premium is allocative, this implies that the cost of marginal hours is more procyclical than the average wage. Since it is the cost of marginal hours that should be equated to the marginal product of hours in (1.1), this correction also makes it harder to assign demand shocks an important role in a competitive model of fluctuations.

A second (though related) difficulty can be expressed without reference to data on real wages. In a competitive model, an aggregate demand shock can produce an increase in output and hours only through an outward and downward shift of the short-run labor supply curve, as in Figure 1a. If we assume a representative household (or more properly, conditions under which aggregation is valid) and time-separable preferences, labor supply in a given period can be written in the Frisch form

$$H_t = \eta(w_t, \lambda_t) \tag{1.2}$$

where λ_t represents the marginal utility of wealth in period t. Here η is necessarily increasing in w_t, and assuming normal goods, it is increasing in λ_t as well. Hence the labor supply shift shown in Figure 1a must result from an increase in the marginal utility of wealth for the representative household.

This is certainly a theoretical possibility; for example, an increase in government purchases at time t could increase λ_t, either through an increase in expected real rates of return (which would increase λ_t for any given expectations about the future marginal utility of wealth) or through an increase in the expected future marginal utility of wealth (due to an expectation of eventual tax increases). But it would imply *countercyclical movements of aggregate consumption.*[2] Again, under the assumption that both leisure and consumption are normal goods, desired leisure can fall (in the face of a real wage decline) only because total expenditure (on consumption and leisure together) has fallen. But that should imply a decline in consumption demand a fortiori (given the real wage decline as well). Yet consumption is clearly procyclical in typical aggregate fluctuations.

These problems relate to the competitive theory of labor demand, and are in fact not much dependent on assuming an instantaneously clearing

2. See Barro and King (1984) for an early discussion of this point.

labor market or neoclassical labor supply. Suppose that one has, instead, an efficiency wage model of the labor market. Equation (1.1) still applies, and hence one still obtains the prediction of a countercyclical real wage. Furthermore, a specification similar to (1.2) still applies as well, although it must now be interpreted as a "quasi-supply curve" for labor, indicating the efficiency wage as a function of the hours demanded by firms (see, e.g., Shapiro and Stiglitz, 1984; Jullien and Picard, 1989).[3] If the efficiency wage depends only on the current level of employment, then there is no way that aggregate demand shocks can shift the efficiency wage locus, and so no variations in equilibrium employment and output in response to such shocks are possible. If, on the other hand, as is plausible in many efficiency wage models, the efficiency wage is lower when households have a higher marginal utility of wealth, then a specification of the form (1.2) is obtained. Aggregate demand shocks can increase employment and output only insofar as they are associated with increases in λ_t, and, as before, this should imply countercyclical consumption demand.

Alternatively, if one assumes imperfectly indexed wage contracts, with firms free to choose their desired quantity of hours ex post, given the contractually specified schedule of compensation as a function of the hours demanded, condition (1.1) again applies. Consequently this type of theory also implies a countercyclical real wage (hence Dunlop and Tarshis' criticism of the *General Theory*).

Other considerations as well lead us to be unhappy with the view that aggregate demand affects equilibrium output and employment primarily through shifts in the labor supply (or quasi-supply) curve. For one, the ability of demand shocks to affect the marginal utility of wealth (and so to shift the labor supply curve in Figure 1a) often depends on an assumption that the suppliers of labor participate in economywide financial markets. If, instead, workers are liquidity constrained, neither an increase in the expected future marginal utility of wealth nor an increase in real rates of return need imply an increase in λ_t. Furthermore, other aspects of the effects of business cycles on the labor market also suggest that the demand for labor at any given real wage moves procyclically. For instance, vacancies are procyclical, suggesting that, as in Blanchard and Diamond (1989), firms are willing to hire more workers at the going wage in booms. Similarly, quits are higher in booms, suggesting that the increased employment at such times is not due to workers' having relaxed their demands as to the acceptable terms of employment.

3. If, as in the papers cited, effort is a zero-one decision variable, then (1.1) still applies, where F represents output assuming that no workers shirk. Condition (1.2) represents the lowest wage consistent with workers in fact choosing not to shirk.

One obvious response to these problems with a competitive theory of business cycles due to aggregate demand variations is to suppose that aggregate fluctuations are instead due mainly to technology shocks. This has recently become a popular view. But this solution is subject to important objections. On the one hand, it is not obvious that one sees evidence of the kind of large variations in aggregate production possibilities at business cycle frequencies that are assumed in such an explanation (Barro and King, 1984; Summers, 1986). The technology shocks may be inferred from the discrepancy between the predictions of a competitive model with smooth technological progress and the facts (i.e., the puzzles just cited, and the related problem of the failure of average labor productivity to move countercyclically), but the absence of more direct evidence has led to continued skepticism about this hypothesis.[4]

Furthermore, if this explanation were correct, the neoclassical predictions should be observed to be correct on those occasions when fluctuations in output and employment are largely due to demand shocks. For example, if increases in military purchases result in increased output and employment, then one should see reduced real wages and reduced consumption spending on those occasions,[5] even if real wages and consumption are procyclical most of the time. But, as is discussed further in Section 3, increased military purchases appear to stimulate higher output and employment without any associated reduction of real wages or consumption.[6] They also seem to be accompanied by increases in vacancies and quits, which is further evidence for the view that increases in military purchases do not affect the labor market only through an effect on labor supply.

An alternative explanation of the failure of real wages to be coun-

4. The mere observation of a procyclical Solow productivity residual need not indicate the existence of technology shocks at business cycle frequencies, as a number of authors have noted (e.g., Baxter and King, 1990; Burnside, Eichenbaum, and Rebelo, 1990; Gordon, 1990; Hall, 1987, 1988a; Lucas, 1989; Rotemberg and Summers, 1990; Summers, 1986). Indeed, Solow (1964) rejected this interpretation. For evidence that measured Solow residuals do not behave in a way that appears to be consistent with their interpretation as exogenous shocks to technology, see in particular Hall, Baxter and King, Burnside et al., and Evans (1990).
5. With regard to the latter prediction, it should be noted that even spending on *nondurable* consumer goods and services should be reduced, at least if one assumes that the utility from services from consumer durables is additively separable from the utility from nondurable consumption and leisure, in which case the above arguments extend directly to a model with durable consumer goods.
6. Our argument here is parallel to that of Hall (1987, 1988a), who rejects the technology shock explanation of procyclical Solow residuals on the ground that Solow residuals also exhibit positive covariance with variables such as growth in military purchases. Like Hall, we interpret our findings as evidence of imperfectly competitive product markets.

tercyclical has been discussed since the 1930s (Kalecki, 1938; Keynes, 1939), and recently revived as part of many modern accounts of the effects of fluctuations in aggregate demand.[7] In this view, the flaw in the above arguments is the assumption of perfectly competitive product markets. Instead, we will argue, not only are prices frequently above marginal cost, but the extent to which this is true varies over the business cycle.

If product markets are imperfectly competitive, (1.1) becomes instead

$$F_H(K_t, H_t; z_t) = \mu_t w_t \tag{1.3}$$

where μ_t denotes the desired markup (ratio of price to marginal cost) in period t. If the markup is variable, then, like the state of technology z_t, it becomes a shift variable for the labor demand curve. In particular, if for some reason an increase in aggregate demand were to result in a reduced markup, the labor demand curve would shift up and to the right, as shown in Figure 1b.[8] This would make possible an increase in output and hours that coincides with an increase in real wages. Furthermore, because of the increase in the real wage, a reduction in leisure could coincide with an increase in consumption. Hence both of the puzzles cited above about the effects of military purchases could be explained (and the other evidence suggesting an effect on labor demand as well). In addition, such a theory would reduce the need to rely on technology shocks as the driving force behind typical fluctuations in aggregate output and employment.

Such a theory also provides an attractive explanation of certain relative price movements over the business cycle. Raw materials prices are most procyclical, intermediate good prices less so, and finished goods prices least of all. Furthermore, Murphy, Shleifer, and Vishny (1989) show that, for many industries, output prices move countercyclically relative to input prices. For technology shocks to explain these facts, they would have to be highly correlated across sectors, which seems implausible. On

7. See, e.g., Bils (1987, 1989), Lindbeck and Snower (1987), Phelps (1989), Rotemberg and Woodford (1989), Stiglitz (1984), Woodford (1990), and Zink (1989). The shift in emphasis in current theories of nominal rigidity, from an emphasis on wage rigidity to an emphasis on price rigidity (see, e.g., Rotemberg, 1987) may also be seen as part of the same general tendency, insofar as it directs attention to product market imperfections rather than labor market imperfections alone, and insofar as theories of nominal price rigidity imply countercyclical markups (even if the desired markup is not the crucial choice variable in such models).

8. In the figure, the labor supply curve is not shown to move. This is not because a demand shock should not in general have some effect on it (again, through an effect on λ_t). We simply wish to indicate that increases in output, hours, and the real wage are possible, regardless of the sign of the effect on λ_t.

the other hand, these facts are consistent with simultaneous reductions in all markups in response to increases in aggregate demand (or, as below, in response to a higher real rate of interest). This would result in least procyclical prices for goods that are latest in the production chain; these firms not only reduce their markups but also purchase inputs from firms with reduced markups, and so on.

It remains, of course, to be explained how an increase in aggregate demand could in fact result in the reduction in markups needed for this explanation. We review three models of endogenous markup determination in the next section. Each of them has been discussed elsewhere; our point here is to show how they all imply a common specification, expressing the markup as a function of two aggregate state variables. The functional relationship is, however, different in the three cases. Given this simple, common specification, we can estimate its coefficients and determine which, if any, of these models is consistent with U.S. data.

The three models we review are the following. In the first, firms are monopolistic competitors whose elasticity of demand depends on the level of sales. According to this model, the markup is a function of current aggregate output (or perhaps output relative to trend). In the second, the "customer market" model of Phelps and Winter (1970), firms are again monopolistic competitors, but current prices affect demand both immediately and in the future. Pricing then involves a tradeoff between increasing market share in the future (by lowering price now) and exploiting existing customers (by raising price now). As a result, the markup now depends on the present discounted value of profits from future sales as well as on current sales. If the present value of future profits is high, the firm gains by reducing its markup to build its customer base. By contrast, high current demand relative to the present value of future profits raises the incentive to exploit current customers by raising the markup.

In the third model, firms belong to oligopolies that collude implicitly as in Rotemberg and Saloner (1986). This collusion is maintained by the threat that reductions in price, that would raise the current profits of a deviating firm, lead to a price war, which reduces future profits. An increase in expected future profits thus reduces the incentive to deviate and allows the oligopoly to maintain markups at a higher level. By contrast, an increase in current demand, relative to this present value, raises the incentive to deviate, so the oligopoly must lower its markup in order to maintain discipline. Hence in this model the same two state variables determine pricing incentives as in the customer market model. The difference is that the implicit collusion model asserts that competition is most fierce when current demand is strong relative to the present discounted value of future profits, while the customer market model asserts the reverse.

After introducing these models, we test the empirical adequacy of the markup equations that they imply. In Section 3, we construct a time series for markup variations in the United States over the postwar period. As in Bils (1987), this requires us to make assumptions about the form of the production function, and to use data on both output and factors of production to control for technology shocks. Like Bils, we find that markups are quite strongly countercyclical.

Section 4 is then devoted to estimating the relationship between markups, current output, and the expected present value of future profits using aggregate U.S. data. Because this present value is hard to measure, we provide a number of different methods for estimating the relationship. Some of our estimates rely on Tobin's q, while others rely only on measures of expected rates of return. Section 5 instead analyzes markup behavior at a more disaggregated level. We look both at time series variation of markups at the two-digit level, and at two case studies. The advantage of the case studies is that both the industry structure and the shocks that affect markups are clearer. Section 6 concludes.

2. *Models of Endogenous Markup Determination*

Many models of possible effects of demand variations on the relationship between price and marginal cost have been proposed.[9] We narrow the scope of the present inquiry by considering only models where desired markups depend on the timing and level of total demand but do not depend on changes in the composition of demand. This is not the only possible type of theory of variable markups. For example, as shown by Bils (1989) and Lindbeck and Snower (1987), changes in the composition of demand can affect the price elasticity of demand perceived by the typical firm, thus changing its desired markup.

We do not pursue these ideas here for two reasons. First, we wish to preserve the traditional view that all increases in aggregate demand, whatever their origin, have the same expansionary effects. Second, models where the markup depends only on the level of total demand are simpler. This simplicity is particularly important when one wishes to close the models in a complete general equilibrium framework (as is necessary for policy simulations like those in Rotemberg and Woodford, 1989). The incorporation of compositional shifts would seem to require additional state variables. For instance, if the poor and the rich have

9. Stiglitz (1984) surveys a number of these.

different demand functions and their relative importance in sales varies over the cycle, the income distribution must be represented by state variables. Similarly, if durables replacement purchases have different characteristics than "upgrade" purchases, the evolution of the stock of durables would have to be modeled explicitly.

2.1 THE BASIC SETUP

We consider economies with many symmetric firms. We focus on symmetric equilibria, so that in equilibrium all firms charge the same price at time t, P_t. For simplicity we treat the output of these symmetric firms as the numeraire so that, in units of the numeraire, P_t is one.

These symmetric firms have access to a technology of the form

$$y_t^i = F[K_t^i, z_t(H_t^i - \bar{H}_t)] \tag{2.1}$$

where y_t^i, H_t^i, and K_t^i represent, respectively, firm i's output, labor input, and capital input at time t. The variable z_t represents the state of technology at time t, so that a higher z corresponds to a more productive period, while $\bar{H}_t$ is the amount of labor devoted to fixed costs. The allowance for an overhead labor requirement is a way of introducing decreasing average costs, of the kind needed to reconcile an assumed markup of price over marginal cost with the apparent absence of significant pure profits in U.S. industry.[10]

Each firm has access to competitive markets for labor and capital services. At time t, firm i must pay a wage w_t for each unit of labor and it must pay r_t for each unit of capital that it rents. Assuming F is homogeneous of degree one and competitive factor markets, marginal cost at t is independent of the number of units that the firm produces and is equal to

$$\min_{h,k} w_t h + r_t k \qquad \text{s.t.} \qquad F(k, z_t h) = 1. \tag{2.2}$$

The assumption that F is homogeneous of degree one so that marginal cost is constant is not essential for the models to be presented below. However, it simplifies our analysis by allowing us to write the ratio of two firms' prices as the ratio of their respective markups. We denote the equilibrium markup by μ_t; this is the equilibrium ratio of the price charged by all firms to marginal cost. Since both w_t and r_t are denominated in the units of the typical firm's output, marginal cost in (2.2) is simply equal to $1/\mu_t$. Letting firm i's ratio of price to marginal cost be

10. For evidence on the existence of increasing returns, in the sense of average costs in excess of marginal cost on average, in U.S. industry, see Hall (1987).

denoted by μ_t^i, firm i's profits gross of fixed costs in units of the numeraire are equal to

$$\Pi_t^i = \left(\frac{\mu_t^i - 1}{\mu_t} \right) y_t^i. \tag{2.3}$$

In a symmetric equilibrium all firms charge the same price and sell the same quantity y_t. This is related to the aggregate level of sales Y_t through the relation

$$Y_t = nI_t y_t$$

where I_t denotes the number of industries in period t and n the number of firms per industry. (In the case of a monopolistic competition model, n equals one and I_t is the number of differentiated goods produced in period t.) It is assumed that I_t grows deterministically at a constant rate, $I_{t+1}/I_t = \gamma$, where $\gamma \geq 1$.[11] This growth in the variety of goods produced can be one source of growth in the aggregate overhead labor requirement. We furthermore assume that goods may disappear from production; each industry in existence in period t is assumed to have a probability α of existence in period $t + 1$, where $0 < \alpha \leq 1$, with the probabilities of disappearance being independent across industries and over time.

Within symmetric equilibria, we denote by x_t each firm's expected present discounted value at t of the stream of individual profits from period $t + 1$ onward

$$x_t = E_t \sum_{j=1}^{\infty} \alpha^j \frac{q_{t+j}}{q_t} \left(\frac{\mu_{t+j} - 1}{\mu_{t+j}} \right) y_{t+j}. \tag{2.4}$$

Here E_t takes expectations conditional on information available at t, and q_{t+j}/q_t is the stochastic variable such that any random yield z_{t+j} (in units of period $t + j$ goods) has a present discounted value in period t of $E_t(q_{t+j}z_{t+j}/q_t)$. The expectational variable x_t is of critical importance in both of the "dynamic" models of markup determination below.

We now distinguish among three models that differ in both the specification of demand and of market structure.

11. The variable I_t takes continuous rather than integral values. In fact, we assume a continuum of industries, so that each has a negligible effect on factor markets and on the average price of output.

2.2 THE STATIC MONOPOLISTIC COMPETITION MODEL

In this model each firm behaves like a monopolistic competitor in that it takes as given the prices of all other firms, the level of marginal cost, and the level of aggregate demand. As in the "symmetric" monopolistic competition model of Dixit and Stiglitz (1977), we assume that the demand for firm i depends on the ratio of its price to the average price charged by all other firms. Equivalently, firm i's demand at t depends on the ratio of its own markup μ_t^i to the markup charged by all other firms in the symmetric equilibrium we will consider, μ_t. Thus we write firm i's demand as

$$y_t^i = D\left(\frac{\mu_t^i}{\mu_t}, y_t\right) \tag{2.5}$$

where the firm's demand depends on aggregate demand through the average level of sales y_t. To preserve symmetry we require that the demand for each firm be equal to y if they all charge the same price. Thus we require that $D(1,y) = y$. A special case to which we will return has homothetic preferences so that demand is the product of a function of relative prices and average demand y_t. In this special case both D and the partial derivative of D with respect to relative prices, D_1, are proportional to y.

Since the firm's problem is static we can obtain its decision rule by substituting (2.5) into (2.3) and maximizing with respect to μ_t^i. This yields the familiar formula

$$D + \frac{\mu_t^i - 1}{\mu_t} D_1 = 0. \tag{2.6}$$

In a symmetric equilibrium all firms charge the same markup, so that the markup can rise if and only if $-D_1(1,y)/D(1,y) = -D_1/y$, the elasticity of demand evaluated at the point where all prices are the same, falls. Thus the markup can rise with a change in y_t if and only if preferences are not homothetic (as in Robinson, 1932). There is little a priori reason to expect either direction of deviation from homotheticity, so that markups seem as likely to rise with increased sales as to fall.

The nonhomothetic case has two disadvantages relative to the homothetic case. First, it leads the markup to be nonstationary if output is itself nonstationary. The existence of nonstationary markups would seem to demand more explicit modeling of the dynamic evolution of fixed costs and of entry more generally. Moreover, as we shall see below, the existence of nonstationary markups would considerably complicate our computation of markup variations. That computation is based on approximating the behavioral equations around the constant value of the markup in the economy's deterministic steady-state growth path. Computations of this type are much more complicated if the markup is nonstationary.

A possible alternative view (and the main reason for our considering the nonhomothetic case in our estimates below) is that the elasticity of demand really depends not on y_t but on the deviation of y_t from its trend path. This would allow stationary fluctuations in the markup, and would justify the log-linear specifications used in our empirical work. The theoretical interpretation of such a specification, however, would probably have to rely on cyclical changes in the composition of demand (for which the deviation of output from trend would be a proxy), in which case a more adequate analysis should specify those changes in composition explicitly.

The second disadvantage of the nonhomothetic case is that aggregation of demand across different types of purchasers (consumers, firms, and the government) becomes more difficult; similarly, the use of a representative consumer to model private consumption demand becomes problematic. In the nonhomothetic case, the composition of demand must itself matter since the elasticity of demand depends on the level of each type of spending instead of depending on the overall level of spending. Yet, as we explained earlier, models where the composition of demand matters are inherently more complicated and possibly unsatisfactory in their implications.

2.3 THE CUSTOMER MARKET MODEL

The customer market model is based on Phelps and Winter (1970). It continues to have each firm maximizing profits with respect to its markup taking the markup in all other firms as given. It differs 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 leads future sales to be higher at any given price. One simple formulation that captures this idea involves writing the demand for firm i at time t as

$$y_t^i = \eta\left(\frac{\mu_t^i}{\mu_t}, y_t\right) m_t^i, \qquad \eta_1 < 0, \qquad \eta(1,y) = y. \tag{2.7}$$

In the homothetic case, once again, η and η_1 are proportional to y. The variable m_t^i is the fraction of average demand y_t that goes to firm i if it charges the same price as all other firms. The market share m^i depends on past pricing behavior according to the rule

$$m^i_{t+1} = g\left(\frac{\mu^i_t}{\mu_t}\right) m^i_t, \qquad g' < 0, \qquad g(1) = 1 \tag{2.8}$$

so that a temporary reduction in price raises firm i's market share permanently. Equations (2.6) and (2.7) are intended to capture the idea that customers have switching costs, in a manner analogous to the models of Gottfries (1986), Klemperer (1987), and Farrell and Shapiro (1988).[12] 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 (2.6) and (2.7) 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.

The firm's expected present discounted value of profits from period t onward is thus

$$E_t \sum_{j=0}^{\infty} \alpha^j \frac{q_{t+j}}{q_t} \left(\frac{\mu^i_{t+j} - 1}{\mu_{t+j}}\right) \eta\left(\frac{\mu^i_{t+j}}{\mu_{t+j}}, y_{t+j}\right) m^i_t \prod_{z=0}^{j-1} g\left(\frac{\mu^i_{t+z}}{\mu_{t+z}}\right). \tag{2.9}$$

Firm i chooses $\{\mu^i_t\}$ to maximize (2.9), taking as given the stochastic processes $\{\mu_t\}$ and $\{y_t\}$. Therefore

$$\eta\left(\frac{\mu^i_t}{\mu_t}, y_t\right) + \eta_1\left(\frac{\mu^i_t}{\mu_t}, y_t\right)\left[\frac{\mu^i_t - 1}{\mu_t}\right] + g'\left(\frac{\mu^i_t}{\mu_t}\right) E_t \sum_{j=1}^{\infty} \alpha^j \frac{q_{t+j}}{q_t}$$

$$\left[\frac{\mu^i_{t+j} - 1}{\mu_{t+j}}\right] \eta\left(\frac{\mu^i_{t+j}}{\mu_{t+j}}, y_{t+j}\right) \prod_{z=1}^{j-1} g\left(\frac{\mu^i_{t+z}}{\mu_{t+z}}\right) = 0 \tag{2.10}$$

where subscripts denote partial derivatives. At a symmetric equilibrium where all firms charge the same price, each has a share m^i equal to one, and g equals one in all periods. So the expectation term in (2.10) is equal to the common present discounted value of future profits given by (2.4). Therefore, (2.10) gives the markup μ_t as

12. This idea has been used in general equilibrium macroeconomic models by Greenwald and Stiglitz (1988), Phelps (1989), and Gottfries (1990). It has been applied to the analysis of international pricing issues by Gottfries (1988) and Froot and Klemperer (1989).

$$\mu_t = \mu(x_t, y_t) \equiv \frac{\eta_1(1, y_t)}{y_t + \eta_1(1, y_t) + g'(1)x_t}. \tag{2.11}$$

Because η_1 and $g'(1)$ are both negative, the derivative of μ with respect to x is negative. An increase in x means that profits from future customers are high so that each firm lowers its price in order to increase its market share. The effect of current sales y_t on the markup is more ambiguous. In the homothetic case where η_1 is proportional to y, (2.11) implies that the markup depends only on the ratio x_t/y_t; the elasticity of the markup with respect to y is equal to the negative of the elasticity with respect to x. A high value of y means that current customers are relatively profitable so that, in the homothetic case, raising prices and exploiting existing customers are relatively attractive. This intuition must be modified when the elasticity of demand facing an individual firm depends on the level of sales. Differentiating (2.11) and ignoring time subscripts, the derivative of μ with respect to y is

$$\frac{-\mu + (1 - \mu)\eta_{12}}{y + \eta_1(1, y) + g'(1)x}$$

which is positive in the homothetic case where η_{12}, the second partial of η with respect to relative prices and y, equals η_1/y. This derivative can be negative if η_{12} is smaller so that demand becomes more elastic as output rises.

Put broadly, Equation (2.11) says that lower prices are a form of investment, an investment in market share. Such an investment is attractive when the present discounted value of the future returns from investment (x) are high relative to its cost, which depends on the level of current sales (y). Hence, a new variable (x) affects the equilibrium markup. This can be thought of in terms somewhat similar to those used in the case of the static model. Because the long-run elasticity is higher than the short-run elasticity, conditions that lead firms to be more concerned about future sales (high x for a given y) mean that they effectively face a more elastic demand curve. They thus lower their markups.

2.4 THE IMPLICIT COLLUSION MODEL

The model in this section is a simplified presentation of Rotemberg and Woodford (1989). We consider an economy with many industries, each of which consists of n firms. The n firms in each industry collude implicitly in the sense that there is no enforceable cartel contract, but only an implicit agreement that firms that deviate from the collusive understanding will be punished. On the other hand, 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. Abusing the language somewhat, we can view industries as monopolistic competitors in the usual sense, while the firms within each industry collude implicitly.

Keeping this distinction in mind, we write the demand for firm i in industry j as

$$y_t^{ij} = D^i\left(\frac{\mu_t^{1j}}{\mu_t}, \ldots, \frac{\mu_t^{nj}}{\mu_t}, y_t\right), \qquad D^i(1, \ldots, 1, y) = y. \tag{2.12}$$

The function D^i is symmetric in its first n arguments except the ith, and the functions D^i (for $i = 1, \ldots, n$) are all the same after appropriate permutation of the arguments. Using (2.3), profits for firm i in industry j when all other firms in industry j charge the markup μ_t^j, while firms in other industries all charge μ_t, equal

$$\pi_t^{ij} = \frac{\mu_t^{ij} - 1}{\mu_t} D^i\left(\frac{\mu_t^j}{\mu_t}, \ldots, \frac{\mu_t^{ij}}{\mu_t}, \ldots, \frac{\mu_t^j}{\mu_t}, y_t\right). \tag{2.13}$$

If each firm existed for only one period, it would maximize (2.13) with respect to its own markup treating the markups of all other firms as given. The resulting Bertrand equilibrium in the industry would have a markup equal to $\mu^B(\mu_t, y_t)$. If the firms in an industry charged more than $\mu^B(\mu_t, y_t)$, individual firms would benefit from undercutting the industry's price. Higher prices, with their attendant higher profits, can be sustained as a subgame perfect equilibrium only 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 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 $\{\mu_t\}$ and $\{y_t\}$. As shown by Abreu (1986), the punishment for any deviation is as severe as possible in the optimal symmetric equilibrium. Therefore, a deviating firm sets price to maximize current period profits π_t^{ij}. The result is that the single period profits of a deviating firm equal

$$\pi_{dt}^j = \max_{\mu_t^{ij}/\mu_t} \frac{\mu_t^{ij} - 1}{\mu_t} D^i\left(\frac{\mu_t^j}{\mu_t}, \ldots, \frac{\mu_t^{ij}}{\mu_t}, \ldots \frac{\mu_t^j}{\mu_t}, y_t\right). \tag{2.14}$$

After any deviation, the firms in the industry punish the deviator to the maximum possible extent. 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. We give conditions that ensure that a deviator indeed earns a present discounted value of zero in Rotemberg and Woodford (1989).[13]

Let x_t^j denote, by analogy to (2.3), 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

$$\pi_{dt}^j \leq \pi_t^j + x_t^j \tag{2.15}$$

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 (2.15) is always binding.[14]

At a symmetric equilibrium, all industries have the same markup, so that each firm sells y_t and x_t^j equals x_t. Using $D(\rho, y)$ to denote $D^i(1, \ldots, \rho, \ldots, 1, y)$, we then have from (2.13)–(2.15)

$$\max_{\rho} \left[\rho - \frac{1}{\mu_t} \right] D(\rho, y_t) = \left[1 - \frac{1}{\mu_t} \right] y_t + x_t \tag{2.16}$$

where ρ represents the relative price chosen by the deviating firm. Equation (2.16) can be solved for μ_t, yielding once again $\mu_t = \mu(x_t, y_t)$. The

13. The main condition requires that there exist a $\hat{\mu}$ smaller than one such that when all firms in industry j charge a markup of $\hat{\mu}$ while the firms in other industries charge a markup greater than or equal to one, a deviating firm cannot sell positive quantities by charging a price in excess of marginal cost. This assumption requires that the goods produced by firms in the industry be relatively good substitutes. It ensures that the deviating firm cannot make positive profits in the periods following a deviation by deviating from the behavior it is expected to follow after the deviation.

14. In Rotemberg and Woodford (1989) we give conditions under which a deterministic steady state exists in which (2.15) is always binding. We also show that, for small enough stochastic shocks, there continues to exist a perturbed equilibrium in which (2.15) 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. In the present case, we need not interpret a low value of α as referring to rapid disappearance of goods from the market; instead, it might be taken to indicate a limit on the ability of firms to punish their competitors for past undercutting. For example, we may suppose that in each period there is a probability α that the previous collusive agreement will be played, including punishment if the previous agreement calls for it, but also a probability $1 - \alpha$ that a new collusive agreement will be negotiated, in which case the prior history of play becomes irrelevant.

relevant solution of (2.16) is the one where μ_t exceeds the Bertrand level, so that deviators undercut the equilibrium price and ρ is less than one.

Denoting by μ_X the derivative of μ with respect to X, (2.16) yields

$$\mu_X = \frac{\mu^2}{D(\rho, y) - y}. \tag{2.17}$$

Since ρ is less than one, $D(\rho, y) > D(1, y) = y$ and μ_X is positive. An increase in x, which raises the cost of deviating, raises the equilibrium markup. Such an increase in the markup is necessary to maintain the equality between the costs and the benefits of deviating.

We can also bound the response of the markup to changes in x from above. In particular

$$x = (\rho - 1/\mu)D(\rho,y) - (1 - 1/\mu)y < (1 - 1/\mu)[D(\rho, y) - y] = \frac{\mu(\mu - 1)}{\mu_X} \tag{2.18}$$

where the first equality follows from (2.16), the inequality from $\rho < 1$, and the last equality from (2.17). Therefore, the elasticity of μ with respect to x, while positive, is smaller than $\mu - 1$.

The effects of changes in y are more ambiguous. In the homothetic case, where $D_Y = D/y$ for all prices, (2.16) implies that μ depends only on the ratio x/y. Thus an increase in y raises the benefits to deviating now and the markup falls. More generally, μ_Y is negative as long as increases in y raise the left-hand side of (2.16) more than they raise the right-hand side. This occurs as long as

$$\frac{\pi_d(\mu, y)D_2(\rho, y)}{D(\rho, y)} > \frac{\pi(\mu, y)}{y}.$$

While this must hold in the homothetic case where D_2/D equals $1/y$, it could fail more generally if yD_2/D is sufficiently less than one for $\rho < 1$. This quantity is increasing in ρ only if the elasticity of demand faced by a deviating firm, $-\rho D_1(\rho, y)/D(\rho, y)$, is a decreasing function of y. For goods that are close substitutes, the optimal deviating ρ is only slightly less than one, even though π_d is much larger than π. Since $yD_2(1, y)/D(1, y) = 1$, it seems likely that yD_2/D is not much smaller than one, so that $\mu_Y > 0$ is implausible in this model.

We consider small deviations of the markup, output and x around their trend values. Variables that are hatted, for example $\hat{\mu}_t$, will thus denote the logarithmic deviation at t of the markup around its trend. The three models we have considered then imply that

$$\hat{\mu}_t = \epsilon_X \hat{x}_t - \epsilon_Y \hat{y}_t. \tag{2.19}$$

Where the theories differ is in their implications for the elasticities ϵ_X and ϵ_Y. These implications can be summarized as follows:

	General case	*Homothetic case*
Static	$\epsilon_X = 0$	$\epsilon_X = \epsilon_Y = 0$
Customer market	$\epsilon_X < 0$	$\epsilon_X = \epsilon_Y < 0$
Implicit collusion	$0 < \epsilon_X < \mu - 1$	$0 < \epsilon_X = \epsilon_Y < \mu - 1$

Note that the predictions of the three models about the possible parameter values are all mutually inconsistent (especially in the homothetic case). Hence estimation of these elasticities allows us to discriminate among the three models.

Finally, note that in (2.19) we can interpret $\hat{y}_t$ as the logarithmic deviation of aggregate output Y_t, and similarly $\hat{x}_t$ as the logarithmic deviation of aggregate profit expectations X_t, where $X_t = nI_t x_t$, or

$$X_t = E_t \sum_{j=1}^{\infty} \left(\frac{\alpha}{\gamma}\right)^j \frac{q_{t+j}}{q_t} \left(\frac{\mu_{t+j} - 1}{\mu_{t+j}}\right) Y_{t+j}. \tag{2.20}$$

These are the variables in terms of which we work in our analysis of aggregate U.S. data below.

An alternative to the methods pursued there, where we try to ascertain how markups vary with $\hat{x}$ and $\hat{y}$, is to analyze the response of the economy to a change in aggregate demand. These changes are akin to exogenous changes in demand because, for the United States, changes in military purchases are arguably due either to break-outs of hostilities in foreign countries (World War I, World War II, the Korean War, the Vietnam War) or to exogenous changes in attitudes toward defense (the Reagan buildup). They also ought to be independent of changes in the private sector's ability to convert inputs into final output. Therefore, any shift in labor demand that they induce ought to be due to markup variation.

Such an analysis is contained in Rotemberg and Woodford (1989) where we study the economy's response to changes in military expenditures. We find that an increase in national defense purchases raises output produced by the private sector (as in Garcia-Milà, 1987), raises that sector's productivity (as in Hall, 1988a), and, most important for our purposes, raises real product wages paid in the private sector. We find that this is true using both quarterly post-War data and annual data starting in 1890. We now inquire which, if any, of the models we consider are consistent with these increased real wages.

Consider first the static model, which makes the markup a function of the level of output. Since the markup μ is only a function of Y, (1.3) can be replaced by

$$F_H(K_t, H_t, z_t) = \mu(K_t, H_t, z_t)w_t.$$

This describes a relationship between H_t and w_t that depends only on K_t and z_t, so that it cannot be affected by aggregate demand. Aggregate demand can affect employment only by shifting labor supply. Increases in real wages following increases in aggregate demand could still be consistent with this story if the derivative of μ with respect to Y were so large that the labor demand curve sloped upward. As is discussed above, this is possible only by having large, and problematic, departures from homothetic demand. Moreover, such a story seems difficult to reconcile with the increases in vacancies and quits that are shown by Rotemberg and Woodford (1989) to accompany increases in military purchases. These too suggest increases in labor demand.

Consider next the customer market model. In this model, increases in military purchases affect the markup insofar as they affect expected rates of return or the relationship between current and expected future output. From a theoretical viewpoint, we would expect increases in military purchases to raise the rate of return. This is also consistent with the evidence in Rotemberg and Woodford (1989). An increase in rates of return should raise markups in the customer market model, thus leading to a fall in labor demand. This is precisely the sort of paradoxical (and unappealing) result presented by Phelps (1989).

Finally, consider the implicit collusion model. In this model, the increases in rates of return lower equilibrium markups and raise the demand for labor. That model is thus consistent with the qualitative features of the empirical responses. Its quantitative fit is discussed at more length in Rotemberg and Woodford (1989).

A different form of evidence on these models can be obtained if one is willing to make more precise assumptions about production functions. In this case, one can construct markup series that one can confront with the models without having to identify demand shocks explicitly.

3. Construction of a Time Series for Markup Variations

3.1 METHOD

We assume (as in the theoretical models discussed above) an aggregate production function of the form (2.1).[15] As in (1.3), the markup of price over marginal cost is then

$$\mu_t = \frac{F_H[K_t, z_t(H_t - \bar{H}_t)]}{w_t}. \tag{3.1}$$

We can thus construct a markup series from aggregate time series for output, factor inputs, and real wages, given a quantitative specification of the production function F (including a value for $\bar{H}_t$), and given a time series for the productivity shocks $\{z_t\}$. The productivity shocks present an obvious difficulty, since they are not directly observed. In our previous paper (Rotemberg and Woodford, 1989), we measured the effects of a particular type of aggregate demand shock on the markup by choosing a shock (innovations in real military purchases) that could be argued to be uncorrelated with variations in $\{z_t\}$. This will not, however, suffice if we wish to construct a time series for cyclical variations in the markup over the entire postwar period. Here we propose instead to construct a series for $\{z_t\}$ from (2.1), using what is essentially the familiar Solow (1957) method, corrected for the presence of imperfect competition and increasing returns to scale.[16]

We consider a log-linear approximation to (2.1) around a steady-state growth path along which H_t grows at the same rate as $\bar{H}_t$, while K_t and Y_t grow at the same rate as $z_t\bar{H}_t$.[17] This approximation yields

15. Our results are little affected by the choice of the functional form (2.1) over the form (5.1) used in the analysis of sectoral data below. By contrast, the assumed size of the fixed costs in relation to total costs (or more generally, of average cost in relation to marginal cost), represented here by the average size of $\bar{H}_t/H_t$, is important to our conclusions.
16. Bils (1987) avoids the need to construct a series for $\{z_t\}$ by assuming a Cobb–Douglas production function with no overhead requirement (at least for production hours) so that F_H in (3.1) can be replaced by $\alpha Y_t/H_t$. We show that this restrictive functional form is not necessary, and are able to consider the consequences of alternative assumptions regarding factor substitutability and the size of fixed costs.
17. The assumption that the overhead labor requirement grows at a constant rate allows us to obtain a stationary equilibrium with growth (in which, among other things, the ratio of fixed costs to total costs fluctuates around a constant value). This could be due to growth in the variety of goods produced as the economy grows, although we do not impose such an interpretation. We could have assumed instead that the overhead labor requirement is constant in per capita terms. Because per capita hours appear stationary, this too would have allowed us to apply our techniques.

$$\hat{y}_t = \left(\frac{F_1K}{Y}\right)\hat{k}_t + \frac{zF_2(H-\bar{H})}{Y}\left[\hat{z}_t + \left(\frac{H}{H-\bar{H}}\right)\hat{h}_t\right] \tag{3.2}$$

where hatted lower case variables refer to log deviations from trend values, and where the other expressions represent constant coefficients evaluated at the steady-state growth path.

We assume that, for both factors, the marginal product equals μ^* times the factor price in the steady-state growth path, where μ^* is the steady-state markup. Therefore, F_1K/Y and zF_2H/Y are, respectively, equal to $\mu^* s_K$ and $\mu^* s_H$, where s_K and s_H are payments to capital and labor as a share of output's value. Because F is homogeneous of degree one, Euler's equation implies that

$$\mu^* s_K + \mu^* s_H \frac{H-\bar{H}}{H} = 1. \tag{3.3}$$

Using (3.3), (3.2) can be written as

$$\hat{z}_t = \frac{\hat{y}_t - \mu^* s_K \hat{k}_t - \mu^* s_H \hat{h}_t}{1 - \mu^* s_K}. \tag{3.4}$$

This allows us to construct a time series for $\hat{z}_t$ from the variations in detrended output and factor inputs, given average factor shares, and given a value for the single free parameter μ^*. This parameter is set to one in Solow's original method.[18]

Assuming that w_t and z_t have the same trend growth rates, the analogous log-linear approximation of (3.1) yields

$$\hat{\mu}_t = \hat{z}_t - \hat{w}_t + \frac{\mu^* s_K}{e}\left(\hat{k}_t - \hat{z}_t - \frac{\mu^* s_H}{1 - \mu^* s_K}\hat{h}_t\right) \tag{3.5}$$

where e represents the elasticity of substitution between the two factors in F, evaluated at the factor ratio associated with the steady-state growth path. Substituting (3.4) for $\hat{z}_t$ this becomes

$$\hat{\mu}_t = \frac{e - \mu^* s_K}{e - e\mu^* s_K}\hat{y}_t + \frac{(1-e)\mu^* s_K}{e - e\mu^* s_K}\hat{k}_t - \frac{\mu^* s_H}{1 - \mu^* s_K}\hat{h}_t - \hat{w}_t. \tag{3.6}$$

18. Technically, Solow's calculation also differs from (3.4) in allowing the factor shares to be time-varying. This amounts to preserving some higher-order terms in the Taylor series expansion of (2.1), but there is then little reason to drop other second-order terms. We thus stick here to a simple log-linear approximation.

Hence we need to specify only the parameters e and μ^* in addition to the observable factor shares to construct our markup series. Assigning numerical values to e and μ^* is admittedly somewhat problematic. Our basic strategy is to determine ranges of plausible values, and then to check the degree to which our results are sensitive to the exact values chosen for e and μ^* within those ranges. The parameter e is often "calibrated" in real business cycle studies on the basis of observed long-run trends. The absence of a significant trend in factor shares, in the face of a significant trend in relative factor prices over the last century, is sometimes taken to indicate an elasticity of substitution near one. But this is not a particularly persuasive justification. First, this fact might simply indicate that most technical progress is labor augmenting, as assumed in (2.1), rather than a long-run elasticity of one.

Second, there need not be much relationship between the long-run elasticity and the short-run elasticity (relevant for our purposes). On the one hand, if one assumes a "putty-clay" technology, the short-run elasticity of substitution might be much less than that indicated by long-run trends. But, on the other hand, cyclical variations in capital utilization might make the relevant short-run elasticity even greater than the long-run elasticity.

As is discussed in Appendix 1, when utilization varies, the relevant production function for short-term analysis is the reduced form (A.4). Thus, in the above calculations, e is the elasticity associated with F. But, in the long run, utilization may well be constant. In this case, the elasticity one would infer from growth observations would be that associated with the production function in (A.1), $\tilde{F}$, evaluated at constant u. Then the measured long-run elasticity of substitution would be smaller than the relevant short-run elasticity. We must thus admit that the relevant elasticity is not easily measured. We take as our baseline case the value $e = 1$ (Cobb–Douglas), the value most often used in real business cycle studies, but we also consider the possibilities $e = 0.5$ and $e = 2$.

We are similarly unable to directly observe μ^*. Hall (1988a) proposes to measure it on the basis that the $\hat{z}_t$ series given by (3.4) should be orthogonal to changes in variables such as real military purchases or the party of the President. Hall uses value added as his measure of output and finds values above 1.8 for all seven of his one-digit industries. Domowitz, Hubbard, and Petersen (1988) use gross output instead and obtain smaller estimates of μ^* for most industries; a value of around 1.6 is typical of their findings. These smaller estimates do not contradict Hall's findings. In an industry that uses materials inputs, the markup calculated using the value added data, μ^{VA}, exceeds the markup calculated using gross output data, μ^{GO}. If materials inputs vary proportionally with gross output, the theoretical relationship is

$$\mu^{VA} = \frac{1 - s_M}{1/\mu^{GO} - s_M} \tag{3.7}$$

where s_M represents the share of materials in the value of gross output.

When we study aggregate data, we find it more convenient to use value-added data so that the estimates of Domowitz, Hubbard, and Petersen (1988) would have to be adjusted upward to be appropriate for our analysis.[19] Nonetheless, we take 1.6 as our baseline case for the aggregate data, but also consider the value 2. As some readers may be skeptical about the existence of markups even as high as 60%, we present some results for a markup variation series constructed under the assumption $\mu^* = 1.2$, although we regard this as an extremely conservative choice.

3.2 AGGREGATE DATA

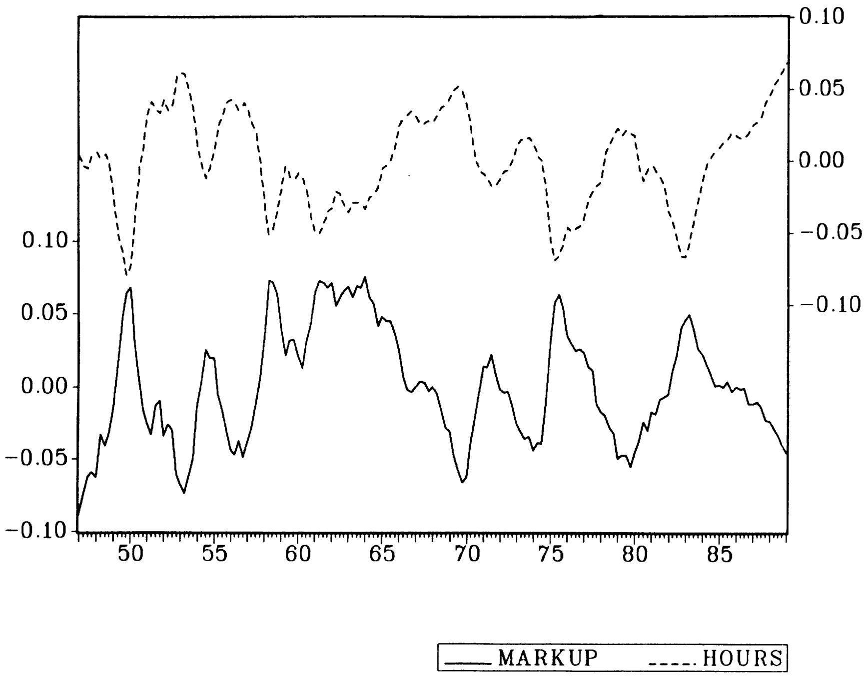
Our time series for Tobin's q comes from Blanchard, Rhee, and Summers (1990). Our measure of the output (value added) of the private sector is obtained from the NIPA as the difference between GNP and the value added by the Federal, State, and local governments. Our index of the prices of goods is the ratio of nominal to real private value added. Our measure of private hours is obtained from the establishment survey as the difference between total hours in nonagricultural payrolls and hours employed by the government. These hours do not have exactly the same coverage as our output series. Thus, for our measures to be strictly accurate, the percentage changes in agricultural hours must equal the percentage changes in the hours of private nonagricultural establishments.

We employ two measures of wages. The principal one is a measure of hourly compensation. This measure equals private employee compensation from the NIPA (i.e., total compensation minus government compensation) over our measure of private hours. The second measure is average hourly earnings in manufacturing. One advantage of the compensation series is that it has a larger coverage both in terms of the sectors whose payments are recorded and in terms of the forms of compensation that are included.[20]

19. Other industry studies using gross output data, such as Morrison (1990), find somewhat lower values for μ^{GO}, ranging between 1.2 and 1.4. Assuming a typical materials share of 0.5 these correspond to μ^{VA} ranging between 1.5 and 2.3.

20. A second advantage is that there is reason to believe the compensation series has smaller measurement error, at least in the way we use it. We use the real wage only to

Figure 2 DEVIATIONS FROM TREND OF HOURS AND THE MARKUP ($e = 1$, $\mu^* = 1.6$)



3.3 BASIC PATTERNS IN THE AGGREGATE DATA

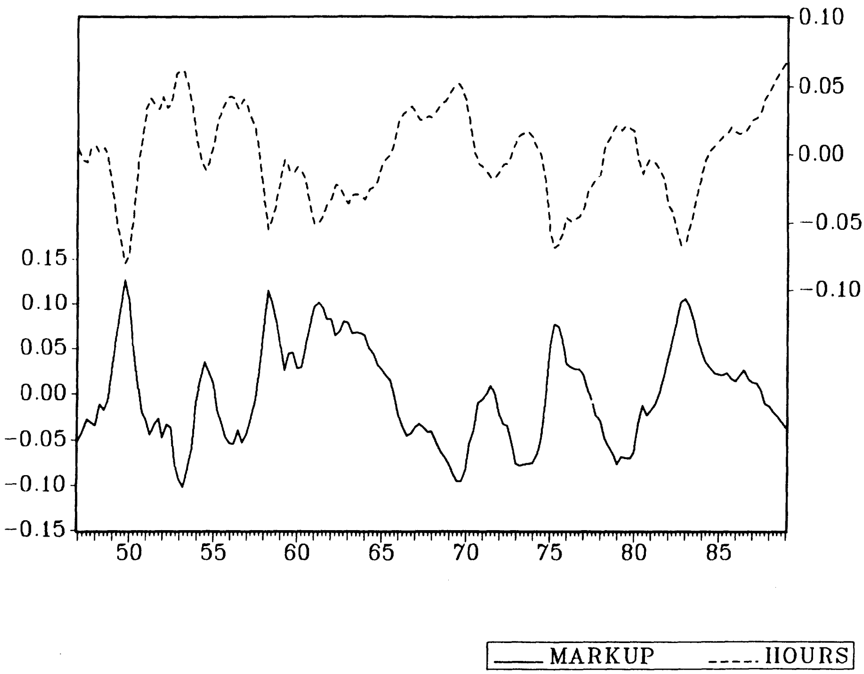
Figures 2, 3, and 4 illustrate the constructed series for the logarithmic deviation of the markup from trend over the postwar period, under different assumptions regarding μ^* and e. These are constructed by ignoring the departures of capital from trend, $\hat{k}$. Because we make an

construct our series on markups. Ignoring fluctuations in capital, Equation (3.6) gives the detrended markups as a function of the detrended levels of output, $\hat{y}_t$, hours, $\hat{h}_t$, and the real wage, $\hat{w}_t$. A simple transformation allows one to write the detrended markups as a function of the detrended labor share ($\hat{s}_{Ht} = \hat{w}_t + \hat{h}_t - \hat{y}_t$), detrended output and detrended hours. The use of the two different wage series is thus equivalent to the use of the corresponding two series for fluctuations in the labor share. To see which series has more classical measurement error we use U.S. data from 1947.III to 1989.I to run regressions of the logarithm of one share on the other including a trend and a correction for first-order serial correlation. When the share using hourly earnings is on the right-hand side its coefficient equals 0.73 and is statistically different from one. When that using compensation is on the right-hand side, its coefficient is 0.93 and is not statistically different from one. We thus cannot reject the hypothesis that the earnings share equals the compensation share plus noise.

assumption about the average level of the markup in order to construct the series, we present here only our constructed series for the deviation from trend, to make it clear that we do not pretend to have directly measured the level. Figure 2 represents our baseline cases, $\mu^* = 1.6$, $e = 1$. Figure 3 shows the consequences of assuming instead $e = 0.5$, while Figure 4 presents the case $\mu^* = 2$, $e = 1$. In each case, the deviation of the logarithm of hours from trend is shown as well; it is clear that for each of these sets of parameters the constructed series displays strongly countercyclical markup variations.

The effects of parameter variation are easily understood. Assuming a lower elasticity e implies a sharper decline in the marginal product of hours in booms, and so increases the amplitude of the countercyclical variation in the series constructed for $\hat{\mu}_t$. Assuming a higher μ^* implies a higher steady state $\bar{H}/H$ because of (3.3), and hence a larger estimate of the percentage increase in $H_t - \bar{H}_t$ for any given observed increase in H_t. For any given e, this then implies a sharper decline in the marginal

Figure 3 DEVIATIONS FROM TREND OF HOURS AND THE MARKUP ($e = 0.5$, $\mu^* = 1.6$)



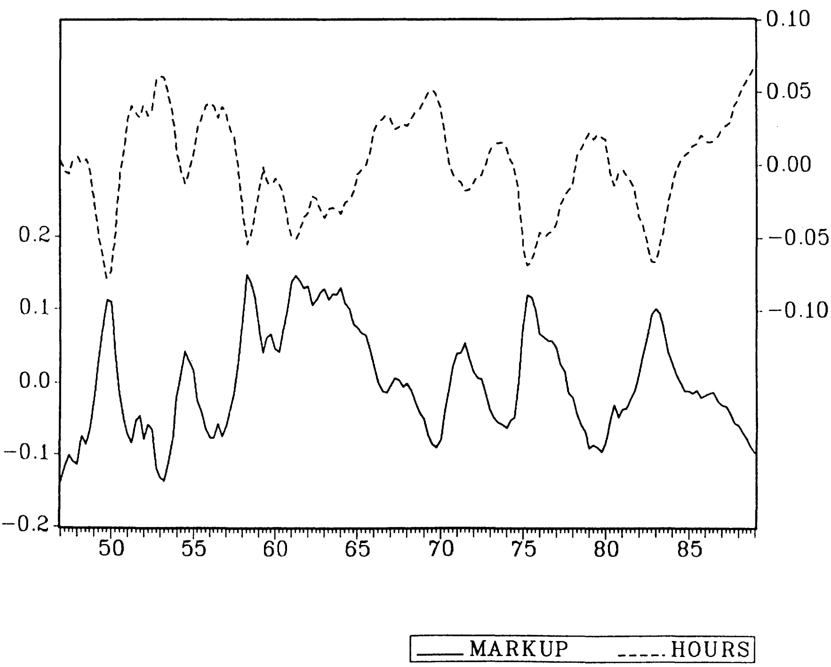
product of hours in booms, so that a higher μ^* results in a greater amplitude of countercyclical variation in $\hat{\mu}_t$. (Note the different scales for the markup series in Figs. 2–4.)

Our result that markups are countercyclical confirm the conclusion of Bils (1987), although we obtain this result for a different reason. Focusing on the baseline case of $e = 1$, (3.6) becomes

$$\hat{\mu}_t = \hat{y}_t - \frac{\mu^* s_H}{1 - \mu^* s_K}\hat{h}_t - \hat{w}_t = -\hat{s}_{Ht} - \left(\frac{\mu^* s_H}{1 - \mu^* s_K} - 1\right)\hat{h}_t \tag{3.8}$$

where $\hat{s}_{Ht}$ denotes log deviations of the share of hours. If μ^* equals one, and given that $s_H + s_K = 1$ (which then implies the absence of fixed costs), $\hat{\mu}_t$ is simply the negative of $\hat{s}_{Ht}$, which is not very strongly cyclical. But if we assume $\mu^* > 1$ (and hence increasing returns), then a countercyclical term is added to $\hat{\mu}_t$. Bils assumes instead a production function with the implication that the marginal product and the average product of production workers' hours decrease in proportion to one another [which

Figure 4 DEVIATIONS FROM TREND OF HOURS AND THE MARKUP ($e = 1$, $\mu^* = 2$)



amounts, in our notation, to deletion of the final term in (3.8)], but he points out that the relevant wage $\hat{w}_t$ is the marginal wage (the wage paid for marginal hours) rather than the average wage. These two quantities can differ if the utilization of overtime labor is cyclical and if overtime hours must be paid more than straight-time hours. With this correction, he obtains

$$\hat{\mu}_t = -\hat{s}_{Ht} - \hat{u}_t$$

where $\hat{u}_t$ represents the log deviation of the ratio of the marginal wage to the average wage. In Appendix 2 we show how to compute this correction with our data. Bils' method for estimating $\hat{u}_t$ depends crucially on regarding the overtime premium as allocative. For a criticism, see Hall (1988b). Because we are uncertain of the extent to which Bils' treatment of the overtime premium is justified, we present most of our results without this correction.

Even in the absence of any premium, the variation in the use of overtime would affect our calculations if straight-time and overtime hours are not perfect substitutes. This may well be the case, as an increase in the number of hours per worker may increase the number of hours that capital is in use, while an increase in the number of employees who work a standard shift does not. This is assumed in Hansen and Sargent (1988), and indeed helps explain the systematic cyclical variation in the use of overtime hours.

In Appendix 2, we show that in the baseline case of $e = 1$, with no fixed costs ($\mu^* = 1$), and assuming no premium for overtime hours, equation (3.8) takes the form

$$\hat{\mu}_t = \hat{s}_{Ht} - \frac{1}{\epsilon_{12}}(\hat{h}_{2t} - \hat{h}_t) \tag{3.9}$$

where $\hat{h}_{2t}$ and $\hat{h}_t$ represent the percentage deviation of overtime and total (straight-time plus overtime) hours, respectively, while ϵ_{12} equals the elasticity of substitution between the two kinds of hours. Hence, if $\epsilon_{12} < \infty$, markups become more countercyclical the more procyclical is the movement of overtime hours relative to total hours. In Appendix 2, we show that overtime hours increase by 7% for each 1% increase in total hours. Hence, if $\epsilon_{12} = 6$, (3.9) implies

$$\hat{\mu}_t = -\hat{s}_{Ht} - \hat{h}_t.$$

This is exactly our baseline markup series [the one implied by (3.8) in the case of $e = 1$ and $\mu^* = 1.6$]. Hence the degree of countercyclical markup variation indicated by Figure 2 could easily result even in the complete absence of increasing returns. It should be noted that the elasticity $\epsilon_{12} = 6$ is more than twice the value assumed by Hansen and Sargent ($\epsilon_{12} = 1/0.36$) who assume that adding overtime hours has no effect on the marginal product of straight-time hours. It is thus hardly outside the range of plausibility. But because the connection between overtime and the work week of capital is hard to measure directly, we implicitly assume $\epsilon_{12} = \infty$. It should be clear, however, that assuming a lower value for ϵ_{12}, together with a lower value for μ^*, would result in constructed series for markup variations very similar to those we use.

Our specification of production possibilities is obviously overly simple in many respects, and many of its shortcomings deserve more careful attention in the future. As we noted in the introduction, the cost of an additional hour of work probably differs from the wage. However, the most obvious corrections make this cost more procyclical so that markups are even more countercyclical than is implied by our method.[21]

4. *The Evidence from the Aggregate Data*

4.1 THREE METHODS FOR EVALUATING THE COMPETING THEORIES

In the next two subsections we estimate the coefficients of (2.19). The problem with estimating (2.19) is that we lack direct observations on $\hat{x}_t$. We have three methods for dealing with this issue. The first uses measurements of Tobin's q, the ratio of firms' market value to the value of their capital in place. The total market value of all firms V is equal to

21. One defect of average wages is that they abstract from the heterogeneity of different workers' hours. As many studies have shown (e.g., Kydland and Prescott, 1988; Barsky and Solon, 1989), the most important such bias has to do with the greater cyclical variability of low-wage (and presumably low-productivity) hours. Suppose that low-wage and high-wage hours are two distinct factors of production, and assume a Cobb–Douglas production function. We can measure the markup as the ratio of the marginal product of low-wage hours to the low wage. Then, corresponding to (3.8), one obtains

$$\hat{\mu}_t = -\hat{s}_{HLt} - \left(\frac{\mu^* s_{HL}}{1 - \mu^* s_K} - 1 \right) \hat{h}_{Lt}$$

where $\hat{h}_{Lt}$ represents the log deviation of low-wage hours from trend, s_{HL} represents the trend value of the share of payments to low-wage hours in output, and so on. Both $\hat{s}_{HLt}$ and $\hat{h}_{Lt}$ should be more procyclical than the corresponding $\hat{s}_{Ht}$ and $\hat{h}_t$ in (3.8). These considerations tend to make $\hat{\mu}_t$ more countercyclical. On the other hand, s_{HL} is smaller than s_H, so the direction of the overall bias is not certain.

$$V_t = (1 + \lambda_t)K_t + X_t - \Phi_t \tag{4.1}$$

where K_t equals the replacement cost of capital, λ_t equals the ratio of the shadow price of adjusting capital to capital's replacement cost, and Φ_t is the present value of fixed costs. The term Φ_t includes the present discounted value of taxes levied from firms as well as random misvaluations of the stock market. Then the logarithmic deviation of Tobin's q should equal

$$\hat{q}_t = \hat{v}_t - \hat{k}_t = \frac{(1 + \lambda)K - V}{V}\hat{k}_t + \frac{(1 + \lambda)K}{V}\hat{\lambda}_t + \frac{X}{V}\hat{x}_t - \frac{\Phi}{V}\hat{\phi}_t \tag{4.2}$$

where the ratios with (V) in the denominator represent steady state values, and where $\hat{\lambda}_t$ represents the logarithmic deviation of $(1 + \lambda)$ from its steady-state value.

Assuming that on average, equilibrium pure profits are zero ($X = \Phi$) (4.2) becomes

$$\hat{q}_t = \hat{\lambda}_t + \frac{X}{(1 + \lambda)K}(\hat{x}_t - \hat{\phi}_t). \tag{4.3}$$

Thus, the variations in $\hat{\lambda}_t$ and in $\hat{\phi}_t$ prevent $\hat{q}_t$ from being a perfect proxy for $\hat{x}_t$. Absent these variations, one could substitute (4.3) into (2.19) and obtain

$$\hat{\mu}_t = \frac{(1 + \lambda)K}{X}\epsilon_x\hat{q}_t - \epsilon_Y\hat{y}_t. \tag{4.4}$$

Equation (4.4) can be estimated by ordinary least-squares with $\hat{\mu}_t$ as the dependent variable if classical measurement error in $\hat{\mu}_t$ is the main source of error. This is not likely given that our procedure for constructing $\hat{\mu}_t$ uses variables that are correlated with q and y. Any specification error is likely to be correlated with these variables.

An alternative is to run a regression of $\hat{q}_t$ on the other variables. This will recover the coefficients in (4.4) if the main error term in (4.4) comes from shocks to $\hat{\phi}_t$ that are uncorrelated with $\hat{x}_t$. Examples of such shocks might include regulatory changes and random misvaluations of the stock market. However, even these shocks may have a direct effect on demand so that they affect all the other variables. Running the regression might also be justified if there are important fluctuations in $\hat{\lambda}_t$ as long as these have one important feature. Investment (and so $\hat{\lambda}_t$) would have to respond to *short-term* sales expectations, which are largely orthogonal to the variations in *long-run* sales expectations that affect $\hat{x}_t$. The obvious problem with this reverse regression is that changes in rates that affect $\hat{x}_t$ and $\hat{\phi}_t$ by similar amounts have little effect on $\hat{q}_t$. Thus, the coefficient on $\hat{\mu}_t$ (which is affected by these shocks) will be biased downward.

All attempts to use data on q as a proxy for X are clearly problematic, given our inability to observe either $\hat{\lambda}_t$ or $\hat{\phi}_t$ directly. Furthermore, the expected profits variable X_t occurring in (4.1) may not be the same as the one that affects markup determination in the theories described in Section 2. Suppose for example that, as discussed earlier, the parameter $(1 - \alpha)$ is taken to indicate not the probability of disappearance of an industry but rather the probability of renegotiation of the collusive agreement among oligopolists. Then the discounted profits that determine the size of the maximum feasible penalty for deviation involve discounting of future profits by the factor α as in (2.20), but the discounted profits that determine the value of the stock market should not involve discounting by this factor. This provides another possible source of misspecification in (4.4). Hence it is desirable to find another way of making inferences about variations in X.

Our second procedure starts from the observation that (2.20) implies

$$X_t = E_t\left\{\frac{\alpha}{\gamma}\frac{q_{t+1}}{q_t}[\Pi_{t+1} + X_{t+1}]\right\} \tag{4.5}$$

where Π_t denotes aggregate profits in period t. In the steady state where capital, output, and profits grow at the rate g, the trend value of X_t equals the trend value Π_t times $\delta/(1 - \delta)$, where

$$\delta = \frac{\alpha(1 + g)}{\gamma(1 + r^*)}$$

and r^* is the trend value of the real rate at which profits are discounted. Therefore, the log-linearization of (4.5) gives

$$\hat{x}_t = E_t\{(1 - \delta)\hat{\pi}_{t+1} + \delta\hat{x}_{t+1} - \hat{r}_{t+1}\} \tag{4.6}$$

where $\hat{r}_t$ is the log deviation from trend of the gross real rate of return between $t - 1$ and t. Moreover, linearizing (2.3) gives

$$\hat{\pi}_t = \hat{y}_t + \hat{\mu}_t/(\mu^* - 1). \tag{4.7}$$

We can use these two linearizations to estimate the coefficients of (2.19) using two alternative procedures. The first involves substituting for both $\hat{x}_t$ and $\hat{x}_{t+1}$ in (4.6) using (2.19) and (4.7). This gives

$$\hat{\mu}_t + \epsilon_Y \hat{y}_t = E_t\left\{\left[\delta + (1-\delta)\frac{\epsilon_X}{\mu^* - 1}\right]\hat{\mu}_{t+1} + \left[\delta\epsilon_Y + (1-\delta)\epsilon_X\right]\hat{y}_{t+1} - \epsilon_X \hat{r}_{t+1}\right\}. \quad (4.8)$$

If one eliminates the expected value operator from (4.8) one obtains an equation whose residual is supposed to be uncorrelated with information available at t. Following the suggestions of Hansen (1982) we estimate this equation by instrumental variables. The great advantage of this method over the one based on observations of q is that changes in $\hat{\phi}_t$ and in $\hat{\lambda}_t$ do not affect the estimates. We can also avoid the problems associated with the possible difference between the rate at which the stock market discounts future profits and the relevant rate for markup determination. To implement this procedure, we need a value for δ. In our baseline case we will let δ take on a value equal to 0.9. However, we also consider letting α/γ equal to one so that δ equals just $(1 + g)/(1 + r^*)$, which, in the case of stock returns, is 0.987.

Our third procedure involves substituting (4.7) in (4.6) and solving forward so that

$$\hat{x}_t = E_t \sum_{j=0}^{\infty} \delta^j \left[(1-\delta)\left(\hat{y}_{t+j+1} + \frac{\hat{\mu}_{t+j+1}}{\mu^* - 1}\right) - \hat{r}_{t+j+1}\right]. \quad (4.9)$$

To obtain estimates of this level of $\hat{x}_t$, we use techniques analogous to those in Hansen and Sargent (1980) and Campbell and Shiller (1988). In other words, we estimate a vector autoregression including at least the variables $\hat{\mu}$, $\hat{y}$, and $\hat{r}$. We can write this vector autoregression in compact notation as $z_t = Az_{t-1} + \epsilon_t$, where the vector z_t includes both current and lagged realizations of the included variables. We let the first three elements of z_t be $\hat{y}_t$, $\hat{\mu}_t$, and $\hat{r}_t$. The resulting estimate of $\hat{x}_t$ is then $v' A[I - \delta A]^{-1} z_t$, where v is a vector whose first three elements are given by $(1 - \delta)$, $(1-\delta)/(\mu^*-1)$, and (-1), respectively, while its other elements all equal zero.

We use these estimates of $\hat{x}_t$ to run regressions of the form of (2.19) and thereby obtain estimates of ϵ_X and ϵ_Y. We also use these estimates of $\hat{x}$ to discover whether $\hat{x} - \hat{y}$ is pro- or countercyclical. Here the customer market and the implicit collusion model make opposite predictions, at least in the homothetic case.

An issue that our methods do not solve is that, in practice, there probably exist changes in markups that are not due to changes in either $\hat{x}_t$ or $\hat{y}_t$. These specification errors could well affect our estimates since exogenous changes in markups might, in turn, affect output and X. Even here, the specification in (4.8) might be more robust to the presence of such errors than that in (4.4). The reason is that the markup and the level of output enter in (4.8) essentially as first differences and, in addition, its parameters are estimated via instrumental variables. So, as long as whatever predictable exogenous changes in markups exist do not also affect expected rates of return and expected growth in output, the estimates remain valid.

The linearizations that lead to our estimating equations involve the logarithmic deviations from trend values. Instead of prior detrending, we include instead the logarithm of the variables and add a constant and a deterministic trend. In particular, we compute markup variations using the logarithm of output, hours and real wages in (3.6).

We present results for our three estimation methods in three subsections. The first covers the estimates from (4.4) by ordinary least-squares. The second discusses the estimates from estimating (4.8) by instrumental variables. Finally, the third presents the results when we use (4.9) to obtain a proxy for the level of $\hat{x}$.

4.2 ESTIMATES BASED ON TOBIN'S q

Our baseline markup variation series is constructed assuming an average markup μ^* equal to 1.6 and an elasticity of substitution of capital for labor e equal to 1.0, and ignoring overtime. We estimate this equation in two ways. First, we estimate it in levels. The residuals from this estimation are highly serially correlated, so that we report standard errors constructed using the procedure suggested by Newey and West (1987), which is also robust to the presence of heteroscedasticity. Second, we estimate it assuming the residuals have first order serial correlation. In this later case, ρ is the autocorrelation of the residual. Using data for the period 1952.II to 1988.IV, these two estimation procedures yield

$$\begin{aligned}\mu_t = {} & \underset{(0.5)}{0.77} + \underset{(0.0007)}{1.4 \times 10^{-5}t} - \underset{(0.08)}{0.63y_t} + \underset{(0.015)}{0.058q_t} \\ & R^2 = 0.983 \qquad DW = 0.16\end{aligned}$$

$$\begin{aligned}\mu_t = {} & \underset{(0.6)}{-0.72} \;\; \underset{(0.0007)}{0.002t} \;\; - \;\; \underset{(0.09)}{0.42y_t} + \underset{(0.014)}{0.035q_t} \\ & \rho\ 3\ 0.934 \qquad R^2 = 0.997 \qquad DW = 1.54.\end{aligned}$$

The coefficients and standard errors of the levels and quasi-differenced regressions are similar. In both cases, the coefficient on output is negative while that on q is positive as required by the implicit collusion model, and thus of the opposite sign than the coefficients predicted by the customer market model. Moreover, since both coefficients are significantly different from zero at conventional significance levels, the customer market model is statistically rejected. The fact that ϵ_X is statistically different from zero also leads us to reject static models of the markup where the only determinant of the markup is the current level of output.

According to (4.4), the coefficient on y_t is $-\epsilon_Y$ while that on q_t is $[(1+\lambda)K\epsilon_X]/X$. Ignoring the average value of λ, which is presumably small, we need to multiply the latter by X/K to obtain an estimate of ϵ_X. According to our model, this expression equals $\delta(1-1/\mu^*)Y/(1-\delta)K$, which equals $3.75Y/K$ for our baseline case. Since Y/K is roughly 0.1,[22] the implied values for ϵ_X are just over 0.01 and just over 0.02 for the two cases. Both are certainly smaller than $\mu^* - 1$ as the implicit collusion model requires.

We show in Table 1 how the coefficients from the quasi-differenced form vary as we vary μ^* and e. Increases in μ^* raise the variability of the markup. In particular, they amplify the reduction in $\hat{\mu}_t$ for a given increase in $\hat{h}_t$. As a result, a given increase in $\hat{y}_t$ reduces the markup by more. This explains why the coefficient on y_t falls as μ^* rises. What is somewhat more unexpected is that increases in μ^* also raise the coefficient on q so that the implied value of ϵ_X rises as well.

For a given average markup, increases in e raise the coefficient on y_t while having no effect on the coefficient on q_t. The reason for this apparently anomalous result can be seen from the formula (3.6) giving our measure of markup variations. For a given μ^* [and hence $H/(H-\bar{H})$], changes in e affect markup variations only by affecting the influence of private output on the markup. In particular increases in e raise the weight of changes in output on the measured markup. These increases therefore raise the estimated effect of y_t on μ_t.

We now turn to estimation of the same equation but with q_t on the left-hand side. We again consider separately the estimation in levels with robust standard errors and the estimation in quasi-first differences. For our baseline series on markup variations, the estimation of such equations including both a constant and a trend yields

22. See Rotemberg and Woodford (1989).

$$q_t = -15.8 - 0.015t + 4.33y_t \quad + \quad 3.96\mu_t$$
$$\quad (4.8) \quad (0.006) \quad (0.78) \qquad (0.84)$$
$$R^2 = 0.952 \qquad DW = 1.81$$

$$q_t = -4.66 - 0.006t + 1.29y_t \quad + \quad 1.20\mu_t$$
$$\quad (2.5) \quad (0.006) \quad (0.52)\, (0.48)$$
$$\rho = 0.969 \qquad R^2 = 0.952 \qquad DW = 1.81.$$

The coefficient on the markup equals $X/K\epsilon_X$ and that on private value added equals $X\epsilon_Y/K\epsilon_X$. The estimates of both ϵ_Y and ϵ_X are positive. In addition, the ratio of the coefficient on $\hat{y}_t$ over that on $\hat{\mu}_t$ gives ϵ_Y, which is thus estimated to be near one in both specifications. What does differ between the levels and the quasi-differenced specification is the implied estimate of ϵ_X.

To obtain an estimate of ϵ_X we must multiply the inverse of the coefficient on μ by $3.75Y/K$. This gives estimates of ϵ_X of 0.09 in the levels form and 0.45 in the quasi-differenced one. Both are, once again, below $\mu^* - 1$. These conclusions are sensitive to our use of a δ equal to 0.9. If instead, one assumes that α/γ is one so that δ equals 0.987, our estimate of $X/K\epsilon_X$ rises to 3 (from 0.375). The result is that the implied levels of ϵ_X rise to 0.73 for the levels regression and 2.4 for the quasi-differenced one. Both, particularly the latter, are larger than $\mu^* - 1$.

In Table 2 we show how the coefficients on $\hat{y}_t$ and $\hat{\mu}_t$ vary in the quasi-differenced form as we vary μ^* and e. As we increase the average markup (and hence increase its variability) the correlation between the

Table 1 ESTIMATION OF QUASI-DIFFERENCED EQUATION (5.4) FOR DIFFERENT SPECIFICATIONS

Elasticity of substitution	*Coefficient on*	*Average Markup* 1.2	1.6	2
0.5	q_t	0.020	0.035	0.058
		(0.010)	(0.014)	(0.022)
	y_t	−0.364	−1.083	−2.099
		(0.06)	(0.69)	(0.132)
1	q_t	0.020	0.035	0.058
		(0.010)	(0.014)	(0.022)
	y_t	0.065	−0.416	−1.099
		(0.06)	(0.50)	(0.132)
2	q_t	0.020	0.035	0.058
		(0.010)	(0.014)	(0.022)
	y_t	0.279	−0.083	−0.599
		(0.06)	(0.49)	(0.132)

markup and stock prices falls so that the former falls. In contrast, the latter coefficient estimate rises as we increase the average markup.

For a given average markup, increases in e lower the estimated value of $X\epsilon_Y/K\epsilon_X$ while having no effect on the estimate of $X/K\epsilon_X$. The reason for this is, once again, that the increases in e raise the influence of y_t on μ_t. Increases in e therefore reduce the regressions' estimate of the independent effect of output on stock prices.

We now consider the sensitivity of our results to the addition of the Bils correction for the difference between the average and marginal wage. We obtain this correction using the method given in Appendix 2. The resulting correction is reasonably substantial. We estimate that the increased use of overtime implies that, when hours rise by 1% the average wage rises by 0.056 of 1%, while the marginal wage rises by 0.417 of 1%. Using the resulting markup series, estimation of the quasi-differenced form of (4.4) for our basic case yields

$$\underset{}{\mu_t} = \underset{(0.7)}{-0.56} - \underset{(0.0009)}{0.002t} - \underset{(0.10)}{0.66y_t} + \underset{(0.017)}{0.043q_t}$$

Period: 1952.II–1988.IV $\rho = 0.944$ $R^2 = 0.998$ $DW = 1.54.$

The reverse equation with q on the left-hand side yields instead

$$q_t = \underset{(2.5)}{-4.92} - \underset{(0.006)}{0.006t} + \underset{(0.55)}{1.49y_t} + \underset{(0.41)}{1.06\mu_t}$$

Period: 1952.II–1988.IV $\rho = 0.969$ $R^2 = 0.952$ $DW = 1.82.$

Table 2 ESTIMATION OF QUASI-DIFFERENCED EQUATION (5.4) WITH q AS THE DEPENDENT VARIABLE

Elasticity of substitution	*Coefficient on*	*Average markup* 1.2	1.6	2
0.5	μ_t	1.52	1.20	0.86
		(0.74)	(0.48)	(0.32)
	y_t	1.34	2.09	2.59
		(0.55)	(0.69)	(0.86)
1	μ_t	1.52	1.20	0.86
		(0.74)	(0.48)	(0.32)
	y_t	0.68	1.21	1.73
		(0.51)	(0.50)	(0.58)
2	μ_t	1.52	1.20	0.86
		(0.74)	(0.48)	(0.32)
	y_t	0.36	0.89	1.31
		(0.56)	(0.49)	(0.52)

In both cases, the estimate of ϵ_Y rises with the correction. This is not surprising since the correction makes marginal cost more procyclical. However, the estimates of ϵ_X are not very much affected by the correction.

4.3 INSTRUMENTAL VARIABLES ESTIMATES OF (4.8)

The estimation of (4.8) by instrumental variables offers several advantages over the procedures that rely on observations of q. First, the estimates are less affected by variations in either $\hat{\lambda}_t$ and $\hat{\phi}_t$. Second, the method does not require observations on the present discounted value of profits X. It does however require information on discount rates (or marginal rates of substitution). Given the inadequacies of various rates of return as discount rates, we experiment with the return on the stock market, the return on Treasury Bills, and the return on prime commercial paper. Third, it allows us to recover quantitative estimates for both ϵ_Y and ϵ_X more easily. Finally, this method might be somewhat less prone to endogeneity bias.

We include a constant and a trend as well as the logarithms of the markup, output, hours, the real wage, and the level of real returns in our estimation. As instruments we use a constant, a linear trend, the current and one lagged value of the logarithms of output, the labor input, and the real wage as well as the ex post real return between $t - 1$ and t.

The results of estimating (4.8) for the period 1947.III to 1988.IV using our baseline markup series and the return on the stock market are presented in Table 3. We show estimates and summary statistics for both the case where $\epsilon_Y = \epsilon_X = \epsilon$, and for the case where ϵ_Y and ϵ_X are allowed to differ.

Table 3 THE BASIC INSTRUMENTAL VARIABLES SPECIFICATIONS: U.S. DATA 1947.III–1988.IV

Parameter	*Separate coefficients*	*Constrained coefficients*
Constant	0.538	−0.028
	(0.18)	(0.10)
Coefficient on trend	$0.32x10^{-3}$	$-0.33x10^{-3}$
	$(0.2x10^{-3})$	$(0.6x10^{-4})$
ϵ_Y	0.994	
	(0.21)	
ϵ_X	0.243	
	(0.07)	
ϵ		0.207
		(0.06)
DW	2.21	1.62
J	1.51	2.52

The summary statistics reported in Table 3 concerning the fit of the two equations are encouraging. The Durbin–Watson statistic reveals that little serial correlation remains in the errors. Because we use more instruments than there are coefficients, the two equations are overidentified. The test statistic proposed by Hansen (1982) to test these overidentifying restrictions is reported in the row marked J, and is distributed χ^2 with 5 and 6 degrees of freedom under the null hypothesis that the restrictions are valid. The actual values of this statistic are very small, which probably indicates that the instruments are quite collinear.

Turning to the estimates, consider first the case where ϵ_Y and ϵ_X are not constrained to be equal. A 1% increase in X is then estimated to raise the markup by about a fifth of a percentage point. A 1% increase in Y by contrast lowers the markup by about 1%. Both these coefficients are statistically significantly different from zero.

The estimates of ϵ_Y and ϵ_X are inconsistent with the homothetic versions of both dynamic models because they are statistically significantly different from each other. Once homotheticity is dropped, ϵ_Y can be larger than ϵ_X as long as the elasticity of demand is higher when Y is large. Then increases in Y raise disproportionately the number of customers that a deviator gets for a given change in his markup. This disproportionate increase implies that deviations become much more attractive when Y increases. They thus require relatively large reductions in the markup.

Measurement difficulties provide an alternative explanation for the difference between the two coefficients. To gain some intuition into the source of this discrepancy imagine first that δ equals one. Then, (4.8) makes the expected change in the logarithm of the markup between t and $t + 1$ a linear function of the expected change in the logarithm of private value added (with coefficient ϵ_Y) and of the expected real rate of return between t and $t + 1$ (with coefficient ϵ_X).

Since we set δ equal to 0.9, the finding that ϵ_Y exceeds ϵ_X probably reflects that the expected change in private value added is more correlated with the change in the markup than is the expected discount rate. This could well be due to the fact that the relevant discount rate for firms differs from the expected return on stocks, so that the measurement error in $\hat{r}_t$ biases the estimate of ϵ_X downward. One piece of evidence that lends credence to this interpretation is that, as we show below, the estimates of ϵ_X rise substantially when we use other rates of return.

An additional prediction of the implicit collusion model is that ϵ_X should be less than $\mu^* - 1$. This restriction is satisfied whether ϵ_Y and ϵ_X are allowed to differ as in the first column, or whether they are constrained to be equal, as in the second column. In the latter column, the estimate of the elasticity of the markup with respect to X/Y, ϵ, is 0.21, which is well below 0.6 while remaining significantly positive.

The difference between the J statistics reported in the two columns can be used to test whether the restriction that the two elasticities are the same is valid. This is the analogue of the likelihood ratio test proposed by Gallant and Jorgenson (1979), and it sometimes produces inferences that are at variance with those from Wald tests based on the standard errors of the coefficients. Indeed, in this case, the Wald test rejects the equality of the two coefficients, but the difference between the two J statistics is 1.01, which is well below the critical value for the χ^2 distribution with one degree of freedom.

In Tables 4, 5, and 6 we report variations on the model that are designed to gauge the robustness of our results. Tables 4 and 5 are devoted to obtaining estimates for different values of the average markup and for different values of the elasticity of substitution. We again consider in particular elasticities of substitution equal to 0.5, 1, and 2, and average markups of 1.2, 1.6, and 2. Table 4 is devoted to estimates when the two elasticities are equal, while the estimates of Table 5 are obtained without imposing this restriction.

The two parameters μ^* and e affect the results. As explained in Section 3, increases in μ^* and reductions in e both increase the tendency of the markup to be countercyclical. It is thus not surprising that our estimates of ϵ in Table 4 and those of ϵ_Y in Table 5 tend to rise with μ^* and fall with e. What is once again more surprising is that the estimates of ϵ_X in Table 5, which correspond to estimates of the effect of expected rates of return on the markup, also increase with μ^* and fall with e. With the exception of the estimates corresponding to an e of 0.5 and an average markup of 1.2, the estimates of ϵ and ϵ_X in Tables 4 and 5 are lower than the corresponding $\mu^* - 1$ as required by the implicit collusion model.

Table 4 INSTRUMENTAL VARIABLES METHOD: ELASTICITY OF THE MARKUP WITH RESPECT TO X/Y

Elasticity of substitution	*Average markup*		
	1.2	*1.6*	*2*
0.5	0.310	0.240	0.399
	(0.10)	(0.09)	(0.16)
1	0.189	0.207	0.345
	(0.05)	(0.06)	(0.11)
2	0.144	0.210	0.346
	(0.04)	(0.06)	(0.09)

Table 5 INSTRUMENTAL VARIABLES METHOD: SEPARATE ELASTICITIES OF THE MARKUP WITH RESPECT TO Y AND X

Elasticity of substitution	*Coefficient on*	*Average markup* 1.2	1.6	2
0.5	ϵ_Y	0.235	1.592	2.882
		(0.32)	(0.21)	(0.35)
	ϵ_X	0.360	0.248	0.432
		(0.22)	(0.08)	(0.14)
1	ϵ_Y	0.183	0.994	1.987
		(0.14)	(0.21)	(0.32)
	ϵ_X	0.190	0.243	0.422
		(0.06)	(0.07)	(0.12)
2	ϵ_Y	0.042	0.689	1.530
		(0.12)	(0.21)	(0.35)
	ϵ_X	0.146	0.238	0.413
		(0.04)	(0.07)	(0.11)

Table 6 presents other variations while holding the average markup and elasticity of substitution fixed at our base levels of 1.6 and 1. Some of these have no material effect on the results. As can be seen in the first row, this is true in particular when we change our instruments by replacing the lagged return with the lagged dividend–price ratio. It is also true when we use hourly earnings in manufacturing instead of hourly compensation as our measure of the wage. This can be seen by comparing the results in the last three rows with the corresponding results using hourly compensation.

Somewhat more substantive differences emerge when we replace the stock return by returns on Treasury Bills and commercial paper.[23] In the second and third rows of Table 6, it is apparent that the resulting estimates of ϵ_X are larger (while those of ϵ_Y are smaller). The evidence against the homothetic versions of the models is now much weaker; the two coefficients ϵ_X and ϵ_Y are now not statistically different from each other. On the other hand, the estimates of ϵ_X now exceed $\mu^* - 1$, though not by a statistically significant amount.

The next three rows of Table 6 illustrate the effects of changing δ by changing α/γ. In particular, they present estimates from letting α/γ equal one. The resulting increase in δ raises the estimate of ϵ_Y and lowers that of

23. These estimates are constructed by assuming that there is a risk premium attached to these rates of return, so that the average interest rate r is equal to the average rate of return in the stock market. This adjustment has a negligible effect on the estimates. However, some adjustment of this form is needed when α/γ is one, to ensure that X remains bounded.

Table 6 INSTRUMENTAL VARIABLES METHOD: VARIATIONS WITH AVERAGE MARKUP EQUAL TO 1.6 AND ELASTICITY OF SUBSTITUTION EQUAL TO 1

	ϵ_Y	ϵ_X
Use of lagged dividend/price ratio instead of lagged return as an instrument	1.020 (0.20)	0.208 (0.06)
Use of return on Treasury Bills instead of stock return	0.550 (0.13)	0.713 (0.15)
Use of return on commercial paper instead of stock return	0.491 (0.14)	0.751 (0.14)
Use of stock return but δ=0.987 so that α/γ equals one	1.062 (0.20)	0.184 (0.06)
Use of return on Treasury Bills with δ=0.987	0.933 (0.17)	0.365 (0.25)
Use of return on commercial paper with δ=0.987	0.916 (0.19)	0.455 (0.24)
Use of stock returns and hourly earnings in manufacturing instead of hourly private compensation	1.270 (0.28)	0.354 (0.10)
Use of hourly earnings and return on Treasury Bills	0.670 (0.14)	0.706 (0.14)
Use of hourly earnings and return on commercial paper	0.570 (0.14)	0.803 (0.13)

ϵ_X. Note from (4.8) and (4.9) that a reduction in δ makes $\hat{x}$ more sensitive to near term changes in profitability. So the increase in ϵ_X as one lowers δ means that markups are relatively more correlated with changes in near term profitability than with interest rates.

4.4 THE ESTIMATES OF $\hat{x}$ BASED ON (4.9)

To obtain our last proxy for the level of $\hat{x}$ we run vector autoregressions that include $\hat{y}$, $\hat{\mu}$, $\hat{r}$, and $\hat{h}$ and the logarithmic deviation from trend of aggregate investment. We used the $\hat{\mu}$ series constructed assuming an elasticity of substitution of 1.0 and a μ^* equal to 1.6. These vector autoregressions explain each variable with two lags of itself and two lags of each of the other variables.[24] We then computed $\hat{x}$ for our two values of δ and for our three rates of return. The results are summarized in Table 7. In the first column we report the correlation of $\hat{\mu}$ with the relevant measure of $\hat{x}_t - \hat{y}_t$. As predicted by the implicit collusion model, these correlations are uniformly positive.

In the next two columns we report estimates of ϵ_X and ϵ_Y from running a regression of $\hat{\mu}$ on our proxy for $\hat{x}$ and on $\hat{y}$. The estimates are once again consistent with the implicit collusion model, and, at least when δ equals

24. We experimented with including three lags and the results were essentially identical.

Table 7 RESULTS BASED ON x_t CONSTRUCTED WITH VECTOR AUTOREGRESSION

	Correlation of μ and $x-y$	Regression based on (3.19)		Reverse regressions		Correlation of y and $x-y$
		ϵ_X	ϵ_Y	$1/\epsilon_X$	ϵ_Y/ϵ_X	
Stock returns						
$\delta=0.9$	0.115	0.170	0.919	1.54	3.67	0.613
		(0.02)	(0.09)	(0.2)	(0.2)	
$\delta=0.987$	0.127	0.028	0.737	8.80	15.7	0.580
		(0.004)	(0.076)	(1.8)	(1.8)	
Treasury Bill returns						
$\delta=0.9$	0.713	0.461	0.671	0.979	0.966	−0.324
		(0.05)	(0.08)	(0.13)	(0.13)	
$\delta=0.987$	0.209	0.102	0.669	1.90	3.31	0.376
		(0.02)	(0.11)	(0.4)	(0.5)	
Commercial paper returns						
$\delta=0.9$	0.646	0.344	0.604	1.03	0.95	−0.312
		(0.04)	(0.09)	(0.2)	(0.2)	
$\delta=0.987$	0.152	0.066	0.628	2.26	4.00	0.375
		(0.013)	(0.11)	(0.6)	(0.7)	

0.9, are similar to the estimates obtained from the instrumental variables procedure. In the case where δ equals 0.987, the estimate of ϵ_X is substantially smaller than that obtained from the differenced form (4.8).

Standard errors obtained using the Newey–West method to allow for both serial correlation and heteroscedasticity are reported below the estimates. These standard errors suggest that both ϵ_X and ϵ_Y are statistically significantly different from zero, so that the alternative models are once again rejected.

The next two columns present "reverse" regressions of our constructed proxy for $\hat{x}$ on $\hat{\mu}$ and $\hat{y}$. The coefficient on $\hat{y}$ measures ϵ_Y/ϵ_X. This coefficient is estimated to be much larger than one whenever δ equals 0.987. Here too, reducing δ raises the sensitivity of $\hat{x}$ to near term changes in profitability and, as a result, makes $\hat{\mu}$ more sensitive to $\hat{x}$. With stock returns, ϵ_Y/ϵ_X is above one even when δ is equal to 0.9. However, with the other returns, ϵ_Y/ϵ_X is very close to one (and not significantly different from it). Just as in the instrumental variables specification, the estimates with these rates of return are consistent with homothetic preferences.

One important reason for computing our proxy for $\hat{x}$ is to investigate whether business cycles might be due to changes in the markup induced by changes in X/Y. While a complete analysis of this question is beyond the scope of this paper, we ask at least whether our estimate of $\hat{x} - \hat{y}$ is pro- or countercyclical. In the homothetic version of the implicit collusion model, markups fall only if $\hat{x} - \hat{y}$ falls. If reductions in markups are to be a central force in business expansions and the implicit collusion model is to explain the timing of these expansions, $\hat{x} - \hat{y}$ must be countercyclical. For the same reason, the customer market model implies that $\hat{x} - \hat{y}$ should be procyclical.

The constructed $\hat{x}$ using stock market returns is so procyclical that $\hat{x} - \hat{y}$ is procyclical as well. The other measures of real returns, by contrast, give countercyclical $\hat{x} - \hat{y}$ in our baseline case where δ equals 0.9.

5. *Sectoral Evidence on Varying Markups*

This section will address three issues that will be dealt with in three subsections. The first is to see whether markups are more countercyclical in those sectors in which the implicit collusion story makes the most sense. That story would seem grossly inadequate if it describes markups in very unconcentrated industries better than it describes markups in more concentrated ones. We thus construct markups for different two-digit manufacturing sectors to see where markups are more countercyclical.

The sectoral data will also allow us to understand better the role of expectations of future sales. Expected future sales in an industry depend both on current sales in the industry and the current state of the economy in general. Thus we can use aggregate data to make inferences about future sales in an industry. This means that we have access to a richer set of proxies for X, and can expect to observe more independent variation in X and Y when studying industry data. We exploit these proxies in our second subsection.

The third subsection is devoted to industry case studies where we have specific information on the source of demand fluctuations and their effect on price. We provide evidence from the baby food industry and from the electric equipment industry that appears consistent with the model of implicit collusion. These industries would seem particularly relevant since they are very concentrated, and in the case of the electrical equipment industry, members of the industry were convicted of colluding.

5.1 MARKUP CYCLICALITY AND INDUSTRY CONCENTRATION

We study Department of Commerce data at the two-digit SIC level. This is the value-added data used by Hall to construct the average level of the markup in different industries. We address two related questions with these data. We investigate which sectors have more procyclical real product wages and which have more countercyclical markups. We are particularly interested in the question whether wages are more procyclical and markups more countercyclical in more concentrated sectors. Four-firm concentration ratios are hardly perfect as an indicator of whether collusion is possible. However, there are several reasons for doubting that collusion is possible in sectors with a large number of firms. First, small firms tend to have a great deal to gain and relatively little to lose from undercutting their rivals. Second, collusion requires a fair amount of coordination (so that defectors can be punished), and this would seem difficult when there are many firms.

In the first column of Table 8, we thus report the 1967 four-firm concentration ratios for each two-digit industry from Rotemberg and Saloner (1986). These concentration numbers are themselves sales-weighted averages of the concentrations of the four-digit industries that compose each two-digit sector. These concentration numbers are only weakly associated with Hall's (1988a) measures of average industry markups. In fact, they are slightly negatively correlated (Rotemberg and Summers, 1990). Some extension of our model is needed to account for this fact.[25]

25. One possible explanation of the lack of correlation between Hall's (1988a) measures of markups and concentration is provided in Rotemberg and Summers (1990).

One way of gauging the cross-sectional implications of the theory is to pretend initially that all industries have the same average markup, the same elasticity of substitution, and the same correlation between technology shocks and employment. One advantage of this approach is that it does not rely on Hall's (1988a) measures of average markups. Then Equation (4.5) implies that industries whose real wages are more positively correlated with employment have markups that are more negatively correlated with employment. This leads us to analyze the correlation between real product wages and employment in different industries. Correlations of this form are reported in Rotemberg and Saloner (1986). A related question is which sectors have real product wages that fall more in recessions. This is the question asked by Barsky and Solon (1989), who, for a small sample of industries, run regressions of the industry's wage divided by the industry's PPI on the overall unemployment rate. Their results suggest, as the implicit collusion model predicts, that more concentrated sectors have more procyclical real wages.

In Table 8a, we report analogous results using our yearly two-digit data for the period 1948–1985. We once again consider two measures for the nominal wage. The first is total employee compensation divided by total hours. The second is the industry's hourly earnings for production workers. We obtain real wages by dividing these by the industry's value added deflator. The second column in Table 8a reports the correlation between the detrended value of the logarithm of real hourly compensation and the detrended value of the logarithm of employment. The third reports the correlation for our earnings based measure.

The results in Table 8a are broadly consistent with those reported for the period 1948–1978 by Rotemberg and Saloner (1986). Concentrated industries and also durable goods industries are more likely to have a positive correlation between real product wages and employment. To gain a crude understanding of the importance of this effect we present at the bottom the cross-sectional correlation between concentration and the elements in each column. One might be concerned that these large correlations are due exclusively to the effect of durability. We thus also ran regressions of the correlation between earnings (or compensation) and employment on concentration and a dummy that took a value of one if the industry produces durable goods. In the earnings-based regressions both coefficients are significant at about the 20% level while in the compensation-based regressions they are both significant at under the 10% level.

The last two columns of the table present corelations between detrended GNP and industry wages. The results are similar to those obtained by Barksy and Solon in that, for both of our measures of wages,

Table 8 RESULTS BASED ON TWO-DIGIT DATA
a. THE BEHAVIOR OF PRODUCT WAGES

			Correlations of			
Industry	*SIC code*	*Four-firm concen.*	*Indus. empl. and hrly. comp.*	*Indus. empl. and hrly. earn.*	*GNP and hrly. comp.*	*GNP and hrly. earn.*
Food	20	0.345	−0.192	−0.155	−0.057	−0.214
Tobacco	21	0.736	−0.130	−0.099	−0.078	−0.012
Textiles	22	0.341	−0.174	−0.210	0.107	0.168
Apparel	23	0.197	−0.388	−0.273	0.440	0.538
Lumber	24	0.176	−0.383	−0.331	−0.340	−0.236
Furniture	25	0.216	0.103	0.339	0.206	0.350
Paper	26	0.312	−0.281	−0.101	0.393	0.267
Printing	27	0.189	−0.384	−0.353	−0.442	−0.362
Chemicals	28	0.499	0.260	0.332	0.120	0.038
Petroleum	29	0.329	−0.114	−0.217	−0.427	−0.422
Rubber	30	0.691	0.106	0.097	0.421	0.410
Leather	31	0.245	−0.021	0.141	0.165	0.110
Stone and Glass	32	0.374	0.439	0.362	0.023	0.148
Primary metals	33	0.429	0.039	−0.114	0.353	0.315
Fabricated metals	34	0.291	0.309	0.275	−0.423	−0.411
Non-Electrical Machinery	35	0.363	−0.268	−0.273	0.443	0.447
Electrical Machinery	36	0.450	0.060	0.154	0.407	0.639
Motor Vehicles	371	0.808	0.530	0.512	0.489	0.603
Other transportation eqp.	372–9	0.501	0.185	0.040	−0.243	0.116
Instruments	38	0.478	−0.151	−0.072	0.458	0.636
Correlations with C4			0.523	0.420	0.343	0.372

b. THE BEHAVIOR OF MEASURED MARKUPS

		Correlations of			
Industry	*SIC code*	*Indus. empl. and comp.-based markup*	*Indus. empl. and earn.-based markup*	*GNP and comp.-based markup*	*GNP and earn.-based markup*
Food	20	−0.612	−0.551	−0.124	−0.019
Tobacco	21	−0.172	−0.256	−0.223	−0.365
Textiles	22	−0.868	−0.829	−0.613	−0.644
Apparel	23	−0.387	−0.561	0.336	0.122
Lumber	24	−0.325	−0.399	−0.420	−0.524
Furniture	25	−0.875	−0.904	−0.201	−0.265
Paper	26	−0.954	−0.942	0.395	0.417
Printing	27	−0.845	−0.845	−0.159	−0.167
Chemicals	28	−0.971	−0.968	0.175	0.178
Petroleum	29	−0.718	−0.657	0.108	0.148
Rubber	30	−0.454	−0.439	−0.448	−0.420
Leather	31	−0.878	−0.884	0.378	0.372
Stone and Glass	32	−0.878	−0.859	−0.210	−0.271
Primary metals	33	−0.767	−0.707	−0.208	−0.192
Fabricated metals	34	−0.822	−0.776	0.099	0.133
Non-Electrical Machinery	35	−0.687	−0.678	−0.221	−0.180
Electrical Machinery	36	−0.979	−0.981	0.120	0.050
Motor Vehicles	371	−0.319	−0.280	−0.579	−0.619
Other transportation eqp.	372–9	0.095	0.229	−0.085	−0.403
Instruments	38	−0.164	−0.232	−0.343	−0.534
Correlations with C4		0.434	0.477	−0.409	−0.416

concentrated sectors tend to have higher real product wages in booms. Moreover, some unconcentrated industries, such as lumber and wood products (SIC 24) and fabricated metals industries (SIC 34), actually have lower real product wages in booms.

We now consider sectoral markup variations. To construct series of markup variations, we need to have estimates of the average markup μ^* in each sector. We use Hall's (1988a) estimates for this purpose. These estimates are quite substantial in certain cases so that, sometimes, $\mu^* s_K$ exceeds 1. As is apparent from (3.3), this means that the functional form (2.1) cannot adequately capture the presence of fixed costs in these industries (more than all of employment would have to be devoted to fixed costs). This ceases to be an issue if we consider instead a production function given by

$$Y_t = F(K_t, z_t H_t) - \Phi_t \tag{5.1}$$

with the fixed costs Φ_t growing at rate of trend output. In this section we will assume that the elasticity of substitution of capital for labor equals one. Proceeding as in Section 3, the deviation of the markup from trend is then given by

$$\hat{\mu}_t^i = \hat{y}_t^i - \hat{w}_t^i - [1 + (\mu^{i*} - 1)s_H^i]\hat{h}_t^i \tag{5.2}$$

where the i superscript denotes that the variable corresponds to sector i. To construct these markups we used sectoral detrended data on value added, the value added deflator, total hours, and our two indices of nominal wages.

The first question we ask is whether concentrated industries have more variable markups than unconcentrated industries. This would seem to be suggested by our implicit collusion model, though, in its simplest form, that model does not account for the large observed average markups in certain unconcentrated sectors. We thus computed the variance of $\hat{\mu}$ for each sector. The correlations of these variances with concentration are 0.084 and 0.086 for the compensation and earnings based markups, respectively. These correlations are small, suggesting that our measurement technique makes even the markups in unconcentrated sectors quite volatile. However, concentrated sectors have at least slightly more volatility in the markups than unconcentrated ones.

More relevant than variability is how markups are related to changes in employment and GNP. Reductions in markups are associated with outward shifts in labor demand. So, ignoring variations in labor supply and in market real wages, one would expect large levels of employment to be associated with low markups whatever the source of markup variation. In the implicit collusion model, these increases in employment and reductions in markups would of course be attributable to low values of X/Y.

In the first two columns of Table 8b, it is apparent that the negative correlation of employment with markups is a feature of all sectors. Moreover, the numbers reported at the bottom suggest that this negative correlation is not more pronounced in concentrated sectors. There are various possible explanations for these correlations. They might result from the use of upwardly biased estimates, particularly for unconcentrated sectors, of μ^* in (5.2). The existence of such systematic biases is suggested by absence of any significant correlation between Hall's estimates of μ^* and concentration.

Another possibility is that there exist hours variations that are not due to markup variations or technology shocks. These could be due either to measurement error or to changes in labor demand due for instance to changes in distortionary taxation. Whatever the source of these changes in employment, Equation (5.2) implies that they will be negatively correlated with constructed markups. These considerations suggest that we should consider instead the correlation of constructed industry markups with aggregate GNP. As long as the measurement error in employment is industry specific, i.e., not correlated with GNP, measurement error should not pose a problem for the interpretation of correlations of markups with GNP.

Furthermore, even if there are other sources of markup variation (or more generally of sectoral labor demand shifts) we are mainly interested in whether the models describe the covariation of markups with the business cycle. Finally, according to the implicit collusion model increases in aggregate demand raise output by lowering markups in relatively concentrated sectors. Thus, they should have less effect on the output of unconcentrated sectors. This suggests that the correlations between markups and GNP are less affected by spuriously high estimates of μ^* in unconcentrated sectors.

We thus study whether markups in concentrated industries fall more in booms than do markups in less concentrated industries. It is apparent in the last two columns of Table 8a that they do. Indeed, the negative correlation of concentration with the correlation of markups and GNP is slightly stronger than the positive correlation between concentration and the correlation between real product wages and GNP.

5.2 MARKUP EQUATIONS FOR SECTORAL DATA

In this section we test the implicit collusion model more sharply by estimating markup equations for the various two-digit industries. The essence of this estimation procedure is the construction of sectoral proxies for $\hat{x}_t$. The proxies we construct are limited in that, for simplicity and given the data limitations, they hold expected rates of return constant. We also ignore the impact of expected future $\hat{\mu}$ on $\hat{x}$. We focus instead on the fact that different sectors expect their future sales to evolve differently.

We thus focus on only the first term of (4.9) and seek to construct an estimate for

$$\hat{x}_t^i = E_t \sum_{j=0}^{\infty} \delta^j \hat{y}_{t+j+1}^i. \tag{5.3}$$

To obtain this estimate we use the fact that current aggregate GNP contains different information about the future course of output (which we treat here as sales) in different sectors. We thus start by running regressions of an industry's future output on its current output and current aggregate GNP. In other words we run regressions of the form

$$\hat{y}_t^i = c_1^i \, \hat{y}_{t-1}^i + c_2^i y_t + \nu_t^i \tag{5.4}$$

where the unsuperscripted $\hat{y}$ represents aggregate GNP and ν_t^i is a residual. We also run a regression of the form

$$\hat{y}_t = c_3 \hat{y}_{t-1} + \nu_t \tag{5.5}$$

where ν_t is a residual. As long as δc_3 and δc_1 are less than one, the Hansen and Sargent (1980) prediction formulas then imply that $\hat{x}$ in (5.3) is approximately equal to

$$\hat{x}_t^i = \frac{\hat{y}_t^i}{1 - \delta c_1^i} + \frac{\delta c_2^i c_3 \hat{y}_t}{(1 - \delta c_1^i)(1 - \delta c_3)}. \tag{5.6}$$

We then run regressions of the industry's markup $\hat{\mu}_t^i$ on its $\hat{x}_t^i$ proxy and its output $\hat{y}_t^i$. The coefficients in these regressions are ϵ_X^i and ϵ_Y^i, respectively. We estimate these regressions for our 20 industries simultaneously by GLS. Rather than let each industry have its own coefficient we assume that ϵ_X^i and ϵ_Y^i are linear functions of concentration. Thus $\epsilon_X^i = \epsilon_X^1 + \epsilon_X^2 C4^i$ and $\epsilon_Y^i = \epsilon_Y^1 + \epsilon_Y^2 C4^i$ where $C4^i$ is the four-firm concentration ratio for industry i. We estimate these regressions for our two measures of wages and for δ equal to both 0.9 and 0.6. The results of estimating these equations are reported in the first four rows of Table 9, where we also report standard errors that are not explicitly corrected for the presence of serial correlation.

The results for our two measures of wages are essentially identical. In all cases, the coefficients on concentration have the sign predicted by the implicit collusion model. In more concentrated sectors, both higher values of $\hat{x}$ and higher values of $\hat{y}$ raise the markup more, so that both ϵ_X and ϵ_Y are more likely to be positive. When δ (which now applies to yearly data so that it should be lower) is 0.9, the estimate of ϵ_X^i is positive only if the concentration ratio exceeds 0.18, whereas ϵ_Y^i is negative whenever the concentration ratio exceeds one-half. A lower value of δ raises the absolute value of all coefficients. However, interestingly, the cutoff levels of concentration for which ϵ_X^i and ϵ_Y^i change sign do not change much.

Note that, in the context of these markup equations, a high sectoral output depresses markups more in concentrated sectors, while this was not true of sectoral employment in Table 8b. In common with the results in that table, regressions of markups on sectoral output leaving out our measure of $\hat{x}$ also have more positive coefficients in more concentrated sectors. This apparent discrepancy is due to the fact that $\hat{x}$ and $\hat{y}$ are more positively correlated in more concentrated sectors. In other words, when we run a regression of $\hat{x}^i$ on $\hat{y}^i$ and allow the coefficient to depend linearly on concentration, the coefficient is higher in more concentrated sectors.

5.3 DEMAND CONDITIONS AND PRICING: INDUSTRY CASE STUDIES

In this subsection, we briefly discuss two industry case studies that provide anecdotal evidence of possible use in distinguishing among the

Table 9 MARKUP EQUATIONS FOR CROSS SECTION OF SECTORS

Equations explaining markups with x and y		$\epsilon_X{}^1$	$\epsilon_X{}^2$	$\epsilon_Y{}^1$	$\epsilon_Y{}^2$
δ=0.9	Compensation data	−0.117	0.624	−0.357	0.733
		(0.0007)[a]	(0.017)	(0.030)	(0.064)
	Earnings data	−0.101	0.603	−0.394	0.839
		(0.0007)	(0.184)	(0.032)	(0.067)
δ=0.6	Compensation data	−0.317	1.458	−0.633	1.781
		(0.037)	(0.082)	(0.068)	(0.145)
	Earnings data	−0.287	1.571	−0.636	2.097
		(0.040)	(0.088)	(0.075)	(0.156)

[a]Standard errors in parentheses.

theories of pricing presented above.[26] One advantage of considering case studies of this kind is that they allow us to focus on more narrowly defined markets than in the analysis of industry data above; hence the market structure (clearly oligopolistic in both of the cases discussed here) can be better identified. Another is that a wider range of types of information about the possible determinants of industry pricing can be brought to bear.

5.3.1 The Baby Food Industry 1958–1974 The U.S. market for canned baby foods and cereals was a classic oligopoly; in 1972, the three largest producers (Gerber, Beech-Nut, and Heinz) accounted for 91% of industry sales.[27] Buyers' concern for quality control and nutritional expertise allowed this small group of producers, who spent large sums on their image of reliability, to dominate the market. Nonetheless, demand remained somewhat price sensitive; "Gerber brand baby foods [the industry leader] could sell for a penny or so more at retail than other brands (an 11% premium), differentials beyond this could shift customer purchases to other brands."[28]

This market provides an interesting case study of the effects of a large change in expectations regarding the future growth of demand. The U.S. birth rate grew sharply during the 1950s, reaching a peak number of births of 4.3 million in 1957. This resulted in corresponding strong sales of canned baby food throughout the decade. During the late 1950s and early 1960s, producers expected this growth trend to continue. In 1958, the marketing department of Gerber Products had forecast that births would increase to 4.4 million per year by 1965, and to 5.1 million per year by 1970.[29] Instead, the rate of births fell throughout the 1960s and early 1970s, to only 3.75 million by 1965, and to fewer than 3.5 million by the early 1970s. Hence by the late 1960s, not only had sales of baby food declined from those of the early 1960s, but it had become evident that demand would continue to contract for several more years.

Under such circumstances, the alternative models of markup determination imply different responses. According to the static model, future sales expectations have no effect on current markups. If the effect of current sales on markups is countercyclical, then the declining sales of the late 1960s should have meant increased markups. According to the customer market model, firms should cease to hold prices down for the

26. We would like to thank Rob Gertner for drawing our attention to these materials, and for helpful discussions of the cases.
27. Harrigan and Porter (1982, p. 7 and Exhibit 4).
28. *Ibid.* (p. 4).
29. *Ibid.* (p. 3 and Exhibit 1), citing Gerber's 1958 *Annual Report.*

sake of maintaining market share, and instead raise prices to increase the revenues obtained from current customers. According to the implicit collusion model, instead, collusion should become more difficult to sustain, so that markups should fall. In fact, price competition intensified in the industry, with price wars breaking out in 1968 and continuing, separated by periods of temporary truce, until July 1974. H.J. Heinz (the third largest producer) took the lead in cutting prices, with the other firms forced to respond; Heinz succeeded by this means in raising its market share.[30]

There are, of course, several ways of accounting for the price cuts. One might simply postulate a static model of procyclical markups, or even argue that marginal costs fell sharply due to low utilization. Still, the revision of producers' expectations of future sales growth would seem to have been a more dramatic change than the decline in current sales itself, and the effect seems not to have been at all consistent with the prediction of the customer market model, while it looks very much like a breakdown of oligopolistic collusion.

5.3.2 The Electrical Equipment Conspiracy 1948–1962 The U.S. market for large turbine generators of electricity was dominated by two large producers, General Electric with an average market share of 61% over this period, and Westinghouse with an average market share of 32%. A third producer, Allis-Chalmers, that left the market at the end of 1962, accounted for most of the rest.[31] There exists considerable evidence suggesting collusive pricing in this industry. An antitrust suit concluded in 1962 led to the imprisonment of seven industry executives for fixing prices in this and other markets.

Despite this, collusion was far from perfect. Government-owned utilities bought through sealed bids. Investor-owned utilities negotiated with manufacturers' salesmen but without revealing to one producer what its rivals had bid. The diffusion of information about pricing policies was also hampered by the customization of generators to particular specifications and by the inclusion of spare parts and accessories in the bid. Each manufacturer had a "price book" that allowed a "book price" to be computed for a given generator, and these books were public information. But, the computation often allowed room for interpretation, due to the many possible options, and the price quoted could involve a discount that varied from customer to customer. GE acted as "price leader," with competitors matching its book prices. The discount-

30. *Ibid.* (pp. 9–10).
31. Porter and Ghemawat (1986, p. 6).

ing policies of the three producers differed, however, with GE tending to negotiate a more consistent discount from the book price, while the discounts of Westinghouse and Allis-Chalmers varied more with competitive circumstances.[32]

Both quantities ordered and average prices varied greatly from year to year (Fig. 5).[33] The cyclical variation in orders was apparently due to variation in utilities' forecasts of peak electricity demand. Forecasts had to be made far in advance, due to the delay involved in engineering and construction of the customized generator, and in installation by the utility (a total of 2.5 to 3 years on average, between the order and the unit's coming on-line), while purchasing capacity before it was needed was costly due both to the large capital outlay involved and to rapid technological progress in generator design. Utilities' expectations moved together, both because of common dependence on the national economy and the attention that utility executives paid to each others' forecasts.[34]

Average prices clearly move countercyclically with respect to orders: they fall in 1950 (a peak year for orders), rise to a peak in 1953–1954 (a cyclical trough in orders), fall in 1955 (the beginning of a new period of high demand), rise until a new peak in late 1957 and in 1958 (the next cyclical trough in orders), and then fall until late 1960 (the beginning of another high-demand period). This suggests increased competition in periods of temporarily high demand, consistent with the static model (with procyclical elasticity of demand) or the implicit collusion model, but not with the customer market model. Because one observes such countercyclical pricing in an industry with unusually cyclical demand, it is tempting to conclude that the temporary character of the variations in orders plays an important role in generating the variations in the degree of price competition, in which case the implicit collusion model would seem to fit the case best.

Because of the long time involved in engineering and construction (a year to 18 months, even without delays due to order backlogs), the periods of high demand were followed by periods of 1 to 2 years in which order backlogs were large even if few new orders were taken. It was during these periods of large order backlogs and hence high rates of capacity utilization that prices rose.[35] One might thus argue that prices rise and fall with marginal cost of production, which, in turn, varies with the degree of capacity utilization. Such an interpretation of the industry

32. *Ibid.* (pp. 1, 3, 4, 6).
33. *Ibid.* (Exhibit 3, taken from briefs filed in connection with a subsequent lawsuit by one of GE's customers).
34. *Ibid.* (p. 3).
35. *Ibid.* (Exhibit 4).

Figure 5 TURBINE GENERATOR ORDERS AND BACKLOGS (KILOWATTS) AND INDEX OF ORDER PRICES.

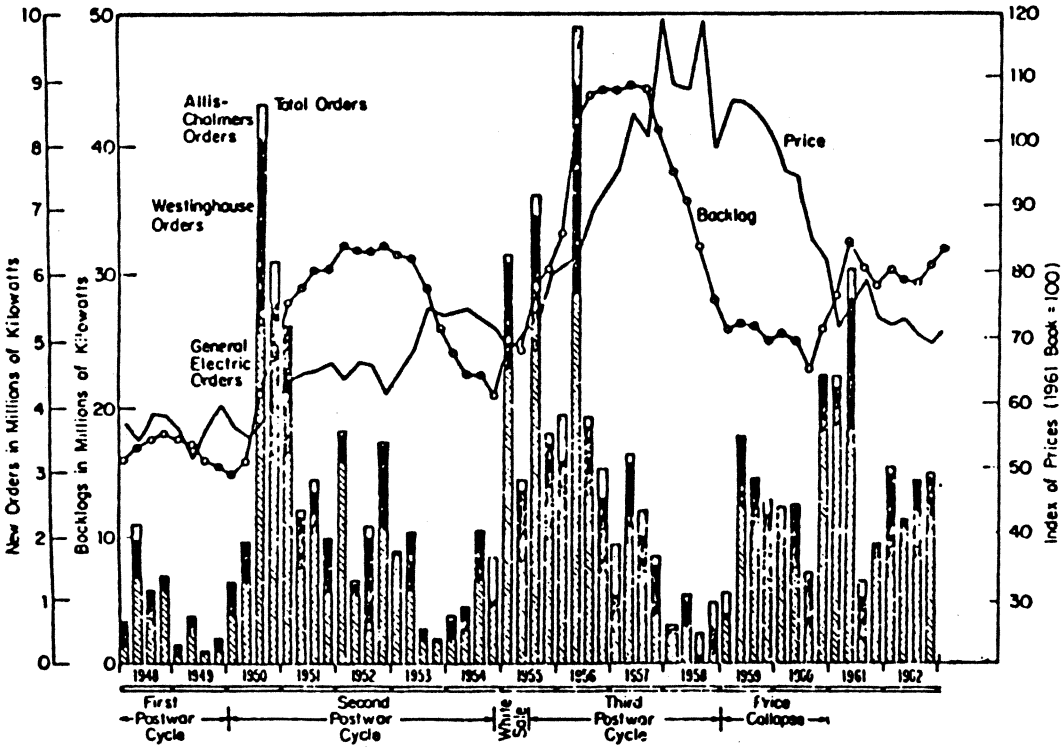


Exhibit material, *Ohio Valley Electric v. General Electric,* Civil Action 62 Civ. 695 (S.D.N.Y. 1965). (S.D.N.Y. is the Second U.S. District Court of New York.) Reprinted in Porter and Gemawhat (1986).

cycles, however, requires considerable myopia on the part of producers. When competing for orders, producers should calculate their marginal cost on the basis of the anticipated level of capacity utilization in production when the orders are to be filled, not at the time that the orders are taken. Hence the high-demand periods should have been periods in which firms, foreseeing a high level of capacity utilization in the following 2 years, would have charged high prices (had markups not been cut).

6. *Conclusions*

We have presented several sources of evidence that suggest that markup variations at cyclical frequencies might be due to changes in the ability to collude over time. These markup variations are partially responsible for fluctuations in activity because they affect the demand for labor. However, we have not measured the extent to which shocks that affect the degree of implicit collusion are responsible for fluctuations in economic activity. For that, a structural model with an explicit identification of the source of all disturbances is required. Such a structural model would include all the equilibrium conditions involved in the determination of markups, employment, output, investment, asset prices, wages, and so on. Our attempt here to estimate markup equations has repeatedly had to face issues of simultaneity and of the possible existence of various unobserved disturbances, and a satisfactory resolution of these problems requires a complete structural model. For example, the implications for the markup equation of observed comovements of markups and stock prices depend, among other things, on how adjustment costs (captured in our model by $\hat{\lambda}_t$) respond to shocks that move markups and stock prices. This can be analyzed only in the context of a joint model of investment and markup determination (like that considered by Chirinko and Fazzari, 1990).

The construction of a structural model will allow us to assess which demand disturbances affect the markup (and labor demand) through X/Y, the ratio of expected future profitability to current sales. One set of demand variables that appears to affect output is that associated with changes in the stocks of certain liquid assets. There are several possible mechanisms through which changes in these assets might affect the economy. One of these is the existence of nominal rigidities. The existence of such rigidities is compatible with the models presented above.

Consider first models in which labor contracts are imperfectly indexed with firms free to chose employment ex post as in Fischer (1977) and Taylor (1980). These can easily be accommodated by our model. With these contracts, Equation (1.3) for the effect of the markup on labor demand and the markup Equation (2.19) would continue to apply. Furthermore, endogenous countercyclical variation in markups would improve the empirical adequacy of the model in at least two respects. First, it would allow monetary surprises to expand output without reducing real wages. Second, it would increase the elasticity of output response to monetary surprises for any given indexation of contracts and any given fraction of the work force covered by the contracts. A general equilibrium model with nominal contracting and endogenous markup determination should provide firmer foundations for the sort of specification used by Taylor (1980) (which involves an ad hoc "markup pricing" rule). It should also improve the empirical adequacy of the kind of general equilibrium model with nominal wage contracts considered by King (1990) and Cho and Cooley (1990).

Now consider models with nominal price rigidity. Impediments to price flexibility such as costs of changing prices affect the markup equation directly, so that they require a bigger modification of the models we have considered. However, as we suggested in the introduction, those models are broadly complementary to the implicit collusion model. Prices may be low in booms both because raising prices would raise the temptation to cheat too much and because firms are reluctant to change prices. Combining the two mechanisms may be desirable for the reasons stressed in Ball and Romer (1990). Countercyclical markups act as a real rigidity, which may magnify the importance of relatively small costs of changing prices.

APPENDIX 1: THE EFFECT OF AGGREGATE DEMAND ON THE LABOR MARKET IN COMPETITIVE MODELS

In this Appendix we examine some possible competitive explanations for procyclical real wages in response to demand shocks. The first candidate applies only to the case of military purchases discussed in Rotemberg and Woodford (1989) and in Section 2. In this case, real wages could rise because labor supply to the private sector falls as a result of conscription. GNP could nonetheless rise as a result of the increase in the value added produced by the government sector. For the period after 1929, we know that this is not the explanation; private value added and private employment both rise together with the increase in military purchases. Thus the increased real wage must be reconciled with an increase in private labor demand.

The next two candidates rely on the assumption that changes in aggregate demand lead to changes in the sectoral composition of demand.

The first variant has labor supply increasing and expansions due to real wage declines in every sector. However, these reductions might be masked in the average real earnings series because high-wage sectors expand more. This lack of proportionality in the expansion of different sectors might be due to greater elasticity of factor substitution in high-wage sectors with the same degree of real wage decline in all sectors. This explanation does not seem sufficient because, as shown by Barsky and Solon (1989) as well as by some of our empirical work reported in Section 7, there are many sectors where real product wages expand together with output.

The second sectoral story has workers increasing their effort because the wage deflated by the consumer price index rises. On the other hand, not all sectors expand. Certain sectors face an increase in the relative price for the good that they sell so that their real product wage falls. By contrast, other sectors face increased real product wages and their output and employment fall. As long as the sectors whose real product wages fall are very labor intensive so that they expand their employment substantially, the net effect can be an increase in aggregate employment. One can check some of the explanatory power of this variant by seeing how relative prices respond to what are arguably changes in aggregate demand, and how this is related to the differential effect on output and employment in different sectors. This is something we hope to address in future research. However, the explanation does not seem a promising one, because, as in the previous case, there are not too many important sectors where the real product wage is countercyclical. Nor are there many sectors where sectoral output and employment are countercyclical.

A third category of competitive explanations is based on the idea that capital utilization varies with aggregate demand. Thus the production function is

$$Y_t = \tilde{F}[u_t K_t, z_t(H_t - \bar{H}_t)] \qquad (A.1)$$

where Y_t, K_t, H_t, and u_t represent output, capital input, hours worked, and capital utilization at t, respectively. The variables z_t and $\bar{H}_t$ represent the state of labor augmenting technical progress and fixed costs at t, as in (2.1). It is then argued that, while K_t is predetermined at t, $u_t K_t$ may vary. However, such a model is incomplete unless it also explains why capital is not always fully utilized. Moreover, the cyclical behavior of real wages depends critically on the particular explanation that is chosen for the partial utilization of capital.

One variant based on Lucas (1970) has a longer "workweek of capital" in booms as firms employ additional shifts and more overtime hours. According to this variant, the defect of (2.1) is not so much that capital is predetermined but that different hours (straight time, overtime, second shift, etc.) are not perfect substitutes. What is needed is a production function like (A.5) in Appendix 2, where different hours are imperfect substitutes because they use capital at different times. With such a production function, there are separate demand curves for the different types of labor. However, firms are induced to hire more hours of any one type only if the real wage for that type of labor input falls. This does not explain how real wages can rise together with output when demand increases. In fact, as Appendix 2 shows, this type of production function actually *increases* the extent to which countercyclical markups are needed to account for the observed cyclical behavior of average real wages.

An alternative capital utilization story assumes that capital utilization is choice variable, which is independent from hours worked. In this story, full utilization is costly because it implies more rapid depreciation of the capital stock. Suppose that capital accumulation obeys the relation

$$K_{t+1} = [1 - \delta(u_t)]K_t + f(Y_t - C_t) \qquad (A.2)$$

where δ is the depreciation rate which is increasing and convex in u and $Y_t - C_t$ represent investable resources at time t. In the absence of adjustment costs, $f(x) = x$, so that gross capital accumulation is simply equal to the difference between output and consumption. In the presence of adjustment costs, f is increasing and concave. Substituting for Y_t in (A.2) and differentiating with respect to u_t, we obtain

$$f' \tilde{F}_1 - \delta' = 0. \qquad (A.3)$$

This equation simply says that firms must end up with the same capital stock if they marginally increase current utilization and use the resulting increased output for investment purposes. If such a modification of utilization raised future capital, it would be strictly profitable; if it lowered it, the firm would gain from lowering its utilization.

We consider first the case without adjustment costs. Then (A.3) does not depend on the state of aggregate demand. An increase in aggregate demand does not, by itself, change u so that it has no direct effect on labor demand. Insofar as, for other reasons, the increase in aggregate demand raises employment, it does raise the marginal product of capital so that equilibrium utilization does rise. If the production function F is homogeneous of degree one, F_1 depends on the ratio of $z_t(H_t - \overline{H}_t)$ to u_tK_t.

Then, (A.3) implies that utilization is a function of $z_t(H_t - \bar{H}_t)/K_t$. This gives rise to the reduced-form production function

$$Y_t = F[K_t, z_t(H_t - \bar{H}_t)] = \tilde{F}\left[u\left(\frac{z_t(H_t - \bar{H}_t)}{K_t}\right)K_t, z_t(H_t - \bar{H}_t)\right]. \tag{A.4}$$

Our analysis in the text uses this reduced-form production function F, which inherits homogeneity of degree one from $\tilde{F}$. Thus our analysis is consistent with variations in capital utilization.

We now turn to the case where there are adjustment costs so that f is concave and f' falls when investment is large. This means that those conditions that raise investment must lower δ', so that they must lower the equilibrium value of capital utilization. In this case, the forces that raise investment also lower labor demand for any given real wage. The close link between utilization and investment is easily understood. A low rate of capital utilization is a form of investment, so it should occur whenever the firm is generally trying to increase its future capital stock.

When the government increases its spending, real interest rates should rise and investment fall. As we show in Rotemberg and Woodford (1989), this is indeed what seems to happen following increases in military purchases. In the presence of adjustment costs, this raises f' so that capital utilization and labor demand rise as well. So this model can explain why real wages rise with increases in military spending. But this model has a very strong implication. It says that, for fixed f' and fixed technology F, labor demand moves inversely with investment. This would seem to be somewhat problematic, since investment is procyclical. Moreover, the change in real wages is positively correlated with the change in real investment spending (in general, though not following increases in military purchases). Of course, one might want to know how investment can be procyclical in a competitive model with fixed technologies F and f. In the presence of adjustment costs, investment can rise when firms expect conditions to warrant high capital stocks in the future. Thus, investment demand should rise if firms anticipate a high marginal product of capital in the future or high labor supply in the future. But the model with varying utilization that we have presented would not allow these increases in investment demand to translate into increases in labor demand.

However, there could also be technological reasons why investment is procyclical. Investment would rise whenever the marginal product of capital F_1 rises and when the cost of adjustment f' falls. Greenwood, Hercowitz and Huffman (1988) present a model with exogenous variations in f'. These are not treated as changes in adjustment costs but, equivalently, as changes in the productivity of new capital relative to old capital. As they show, these changes in f' induce sympathetic movements in output, investment, and labor demand.

The model of varying capital utilization thus is able to resolve our empirical puzzles only in the case of a rather special form of the model (involving both depreciation-in-use and significant costs of adjusting the capital stock) and a very special type of aggregate demand shock. Neither the importance of depreciation-in-use nor the particular type of investment demand shocks needed can be directly measured in any very obvious way; hence credence in this particular resolution will depend on further empirical study of a rather subtle kind.

APPENDIX 2: OVERTIME AND MARGINAL COST

In this Appendix, we consider the effects of variation in the use of overtime hours on our calculation of marginal cost. We first consider the effect of letting straight time and overtime hours be imperfect substitutes (as in Hansen and Sargent, 1988) and then the effect of an allocative overtime premium (as in Bils, 1987).

Hansen and Sargent assume that the two types of hours are used with the same capital stock at different times with the same Cobb–Douglas production function so that total output is

$$Y_t = K_t^{\alpha}[z_t(H_{1t} - \bar{H}_t)]^{1-\alpha} + K_t^{\alpha}[z_t H_{2t}]^{1-\alpha}$$

where H_{1t} and H_{2t} represent straight-time and overtime hours at t, respectively, while overhead hours are assumed to all be straight-time hours. More generally we may suppose that

$$Y_t = F\left\{K_t, z_t Q[(H_{1t} - \bar{H}_t), H_{2t}]\right\} \tag{A.5}$$

where each function F and Q is homogeneous of degree one, concave, and increasing in both arguments. This allows us to nest both the Hansen–Sargent specification (in which Q is a CES function with elasticity of substitution $1/\alpha$) and the case of perfect substitutes (in which Q simply adds its two arguments).

In this case, corresponding to (3.1) we have

$$\mu_t = \frac{z_t F_2\{K_t, z_t Q[(H_{1t} - \bar{H}_t), H_{2t}]\} Q_2[(H_{1t} - \bar{H}_t), H_{2t}]}{w_{2t}} \tag{A.6}$$

where w_{2t} represents the real wage paid to overtime hours. We might equivalently write μ_t as the ratio of the marginal product of straight-time hours to their cost. However, in this case it seems likely that the appropriate costs include costs of varying the level of employment. By contrast there probably exist no appreciable adjustment costs for overtime hours.[36]

Defining the average wage as

$$w_t = \frac{w_{1t}H_{1t} + w_{2t}H_{2t}}{H_{1t} + H_{2t}}$$

and assuming that the overtime premium (w_{tt}/w_{1t}) does not vary over time, one obtains

$$\hat{w}_{2t} = \hat{w}_t - \frac{(w_1 - w)H_1}{wH}\hat{h}_{1t} - \frac{(w_2 - w)H_2}{wH}\hat{h}_{2t}. \tag{A.7}$$

Also, letting $\hat{h}_t$ denote the percentage deviation in total hours,

$$\hat{h}_{1t} = \frac{H}{H_1}\hat{h}_t - \frac{H_2}{H_1}\hat{h}_{2t}. \tag{A.8}$$

To obtain an expression for the deviation of the markup we proceed as before. We log-linearize the production function (A.5) and the first-order condition (A.6) and combine the two to eliminate $\hat{z}_t$. Using (A.7) and (A.8) this yields

$$\hat{\mu}_t = \frac{e - \mu^* s_K}{e - e\mu^* s_K}\hat{y}_t + \frac{(1-e)\mu^* s_K}{e - e\mu^* s_K}\hat{k}_t - \left\{1 + \frac{w_1}{w}\left[\frac{\mu^* - 1 - \mu^*(1 - s_K)/\epsilon_{12}}{1 - \mu^* s_k}\right]\right\}\hat{h}_t$$

$$- \left\{\frac{1}{\epsilon_{12}} + \frac{(w_2 - w_1)H_2}{wH}\left[\frac{\mu^* - 1 - \mu^*(1 - s_K)/\epsilon_{12}}{1 - \mu^* s_K}\right]\right\}\hat{h}_{2t}. \tag{A.9}$$

This is the analogue of (3.6) and reduces to (3.6) if $w_1 = w_2$ and $\epsilon_{12} = \infty$.

If these conditions do not obtain, (A.9) may result in more countercyclical markups than those implied by (3.6). For example, setting $\mu^* = 1$, $e = 1$, and $w_1 = w_2$ leads to (3.9) so that, if $\epsilon_{12} < \infty$ and H_{2t}/H_t is procyclical, then $\hat{\mu}_t$ is more countercyclical than the inverse of the labor share.

36. The existence of adjustment costs for straight-time but not for overtime hours is the most plausible explanation for the greater use of overtime hours in booms; see, e.g., Bils (1987) and Hansen and Sargent (1988).

To gauge the empirical magnitude of this effect we ran regressions of $\hat{h}_{2t}$ on $\hat{h}_t$ and $\hat{h}_t^2$ where the hatted variables are detrended logarithms. Using data for the U.S. manufacturing sector (the only overtime data available) and allowing for an error with both first and second order serial correlation, we obtain

$$\hat{h}_{2t} = \underset{(0.59)}{7.01}\hat{h}_t + \underset{(8.11)}{2.69}\hat{h}_t^2$$

Period: 1956.III–1989.I

Ignoring the statistically insignificant quadratic term, we see that overtime hours increase by 7% whenever total hours increase by 1%. So, as explained in the text, we can obtain our baseline series assuming constant returns as long as we also assume that $\epsilon_{12} = 6$.

If one follows Bils (1987) and assumes that $\epsilon_{12} = \infty$ one has to provide an alternative explanation for the use of overtime workers. Bils (1987) simply assumes that overtime hours are a determinate function of total hours $H_2 = V(H)$. Then, while the average wage is

$$w_{1t} + (w_{2t} - w_{1t})\frac{V(H_t)}{H_t}$$

the marginal wage (the increase in the wage bill when total hours rise by one unit) is

$$w_{1t} + (w_{2t} - w_{1t})V'(H_t).$$

Assuming again a constant overtime premium, $w_{2t} = (1 + p)w_{1t}$, the percent change in the marginal wage for a 1% increase in employment is

$$\gamma_M = \frac{pV''H^2}{H + pV'H}$$

while the corresponding percent change in the average wage is

$$\gamma_A = \frac{p(V'H - V)}{H + pV}.$$

The logarithmic deviation of the ratio of marginal to average wage, $\hat{u}_t$ is then equal to $(\gamma_M - \gamma_A)\hat{h}_t$. Bils then argues that if $\gamma_M > \gamma_A$, the constructed markup series is more countercyclical than the labor share. This method actually understates the degree to which $\hat{\mu}_t$ is countercyclical by assuming that the cost of an additional straight time hour is w_{1t}. It thereby neglects the costs of adding employees, which, presumably, lies behind the use of overtime hours. We adopt Bils's method (i.e., we simply subtract $\hat{\mu}_t$ from the right-hand side of (3.6)) in constructing the markup series used in the regressions reported at the end of subsection 4.2.

If we interpret the above regression as a second-order logarithmic expansion of $V(H)$, the coefficient on $\hat{h}_t$ equals $V'H/V$ while the coefficient on $\hat{h}_t^2$ equals one-half of

$$\frac{H^2V''}{V} + \frac{V'H}{V} - \left(\frac{V'H}{V}\right)^2.$$

Using these facts, together with knowledge that in our data V/H equals 0.0187, gives a value for γ_M of 0.417 and one for γ_A of 0.056. As in Bils's analysis, the former is about eight times larger than the latter. Bils's estimates are both somewhat larger because his index of total hours covers only production hours in manufacturing, so that his average V/H is higher.

REFERENCES

Abreu, Dilip. (1986). Extremal equilibria of oligopolistic supergames. *Journal of Economic Theory* 39:191–225.

Ball, Lawrence, and David Romer. (1990). Real rigidities and the non-neutrality of money. *Review of Economic Studies* 57(April):183–204.

Barro, Robert J., and Robert G. King. (1984). Time separable preferences and intertemporal substitution models of business cycles. *Quarterly Journal of Economics* 99:817–839.

Barsky, Robert, and Gary Solon. (1989). Real wages over the business cycle. *NBER Working Paper 2888*, March.

Baxter, Marianne, and Robert G. King. (1990). Productive externalities and cyclical volatility. Working Paper 245 Center for Economic Research, University of Rochester, September.

Bils, Mark. (1987). The cyclical behavior of marginal cost and price. *American Economic Review* 77(December):838–857.

———. (1989). Pricing in a customer market. *Quarterly Journal of Economics* 104 (November):699–718.

Blanchard, Olivier, and Peter Diamond. (1989). The Beveridge curve. *Brookings Papers on Economic Activity* 1:1–75.

Blanchard, Olivier, Changyong Rhee, and Lawrence Summers. (1990). The stock market, profit and investment. *NBER Working Paper 3370*, May.

Burnside, Craig, Martin Eichenbaum, and Sergio Rebelo. (1990). Labor hoarding and the business cycle. Mimeo, Northwestern University, October.

Campbell, John Y., and Robert J. Shiller. (1988). Dividend price ratios and expectations of future dividends and discount factors. *Review of Financial Studies* 1:195–228.

Cho, Jang-Ok, and Thomas F. Cooley. (1990). The business cycle with nominal contracts. *Rochester Center for Economic Research Working Paper 260,* December.
Chirinko, Robert S., and Steven M. Fazzari. (1990). Market power, returns to scale and Q: Evidence from a panel of firms. Mimeo, University of Chicago.
Dixit, Avinash, and Joseph Stiglitz. (1977). Monopolistic competition and optimum product diversity. *American Economic Review* 67:297–308.
Domowitz, Ian, R. Glenn Hubbard, and Bruce C. Petersen. (1988). Market structure and cyclical fluctuations in U.S. manufacturing. *Review of Economic Statistics* 70(February): 55–66.
Dunlop, John T. (1938). The movement in real and money wage rates. *Economic Journal* 18:413–434.
Evans, Charles L. (1990). Productivity shocks and real business cycles. Mimeo, University of South Carolina.
Fama, Eugene F., and Kenneth R. French. (1988). Dividend yields and expected stock returns. *Journal of Financial Economics* 22(October):3–25.
Farrell, Joseph, and Carl Shapiro. (1988). Dynamic competition with switching costs. *Rand Journal of Economics* 19(Spring):123–137.
Fischer, Stanley. (1977). Long term contracts, rational expectations and the optimal monetary supply rule. *Journal of Political Economy* 85(February):191–206.
Froot, Kenneth A., and Paul D. Klemperer. (1989). Exchange rate pass-through when market share matters. *American Economic Review* 79(September):637–654.
Gallant, Ronald A., and Dale W. Jorgenson. (1979). Statistical inference for a system of simultaneous, non-linear, implicit equations in the context of instrumental variables estimation. *Journal of Econometrics* 11(October/December): 275–302.
Garcia-Milà, Teresa. (1987). Government purchases and real output: An empirical analysis and equilibrium model with public capital. W.P. 93.88, Institut d'Analisi Econòmica, CSIC, Universidad Autonoma de Barcelona, August.
Gordon, Robert J. (1990). Are procyclical productivity fluctuations a figment of measurement error? Mimeo, Northwestern University.
Gottfries, Nils. (1986). Price dynamics of exporting and import-competing firms. *Scandinavian Journal of Economics* 88:417–436.
———. (1988). Market shares and pricing behavior in the tradeables industry: An examination of Swedish manufacturing. Institute for International Economic Studies Paper No. 412, August.
———. (1990). Customer markets, credit market imperfections and real price rigidity. *Economica* (forthcoming).
Greenwald, Bruce, and Joseph Stiglitz. (1988). Financial market imperfections and business cycles. *NBER Working Paper 2494.*
Greenwood, Jeremy, Zvi Hercowitz, and Gregory W. Huffman. (1988). Investment, capacity utilization, and the real business cycle. *American Economic Review* 78:402–417.
Hall, Robert E. (1987). Investment under uncertainty: Theory and tests with industry data. *NBER Working Paper 2264,* May.
———. (1988a). The relation between price and marginal cost in U.S. industry. *Journal of Political Economy* 96(October):921–948.
———. (1988b). Substitution over time in work and consumption. *NBER Working Paper 2789,* December.
Hansen, Gary D., and Thomas J. Sargent. (1988). Straight time and overtime in equilibrium. *Journal of Monetary Economics* 21:281–308.

Hansen, Lars P. (1982). Large sample properties of method of moments estimators. *Econometrica* 50(July):1029–1054.
Hansen, Lars Peter, and Thomas Sargent. (1980). Formulating and estimating dynamic linear rational expectations models. *Journal of Economic Dynamics and Control* 2(February):7–46.
Harrigan, Kathryn R., and Michael E. Porter. (1982). Heinz U.S.A. (baby foods). Harvard Business School Case no. 9-381-102, Boston: HBS Case Services.
Jullien, Bruno, and Pierre Picard. (1989). Efficiency wage and macroeconomic policy. CEPREMAP w.p. no. 8918, July.
Kalecki, Michael. (1938). The determinants of the distribution of national income. *Econometrica* 6:97–112.
Keynes, John Maynard. (1939). Relative movements of real wages and output. *Economic Journal* 49:34–51.
King, Robert C. (1990). Money and business cycles. Mimeo, University of Rochester, July.
Klemperer, Paul D. (1987). Markets with consumer switching costs. *Quarterly Journal of Economics* 102(May):375–394.
Kydland, Finn E., and Edward C. Prescott. (1988). Cyclical movements of the labor input and its real wage. *Federal Reserve Bank of Minneapolis Working Paper 413*, November.
Lindbeck, Assar, and Dennis J. Snower. (1987). Transmission mechanisms from the goods to the labor market. Seminar Paper 403, Institute for International Economic Studies, Stockholm, December.
Lucas, Robert E. Jr. (1970). Capacity, overtime and empirical production functions. *American Economic Review* 60:23–27.
———. (1989). The effects of monetary shocks when prices are set in advance. Mimeo, University of Chicago.
Morrison, Catherine J. (1990). Market power, economic profitability and productivity growth measurement: An integrated structural approach. *NBER Working Paper 3355*, May.
Murphy, Kevin M., Andrei Shleifer, and Robert W. Vishny. (1989). Building blocks of market clearing models. *NBER Macroeconomics Annual 1989*, 247–286.
Newey, Whitney K., and Kenneth D. West. (1987). A simple, positive semidefinite heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55:703–708.
Parsons, Donald O. (1973). Quit rates over time, a search and information approach, *American Economic Review* 63(June):390–401.
Phelps, Edmund S. (1989). Aggregate demand in a non-monetary model of labor-market and product-market equilibrium. Mimeo, Columbia University, June.
Phelps, Edmund S., and Sidney G. Winter. (1970). Optimal price policy under atomistic competition. In *Microeconomic Foundations of Employment and Inflation Theory.* E. Phelps (ed.). New York: W. W. Norton.
Porter, Michael E., and Pankaj Ghemawat. (1986). General Electric vs. Westinghouse in large turbine generators (A). Harvard Business School Case no. 9-380-128, Boston: HBS Case Services, August.
Robinson, Joan. (1932). *The Economics of Imperfect Competition.* London: Macmillan.
Rotemberg, Julio J. (1987). The new Keynesian microfoundations. *NBER Macroeconomics Annual 1987*, 69–104.

Rotemberg, Julio J., and Garth Saloner. (1986). A supergame-theoretic model of price wars during booms. *American Economic Review* 76(June):390–407.
Rotemberg, Julio J., and Lawrence H. Summers. (1990). Inflexible prices and procyclical productivity. *Quarterly Journal of Economics* 105(November):851–874.
Rotemberg, Julio J., and Michael Woodford. (1989). Oligopolistic pricing and the effects of aggregate demand on economic activity. NBER Working Paper *3206*, Cambridge, Massachusetts, December.
Shapiro, Carl, and Joseph E. Stiglitz. (1984). Equilibrium unemployment as a worker discipline device. *American Economic Review* 74:433–444.
Solow, Robert M. (1957). "Technical change and the aggregate production function. *Review of Economics and Statistics* 39(August):312–320.
———.(1964). On the short-run relation of employment and output. Draft of Presidential Address, Econometric Society, mimeo.
Stiglitz, Joseph E. (1984). Price rigidities and market structure. *American Economic Review* Papers and Proceedings 74:350–355.
Stockman, Alan C. (1986). Aggregation bias and the cyclical behavior of real wages. Mimeo, University of Rochester.
Summers, Lawrence H. (1986). Some skeptical observations on real business cycle theory. *Quarterly Review,* Federal Reserve Bank of Minneapolis, Fall:23–27.
Tarshis, Lorie. (1939). Changes in real and money wage rates. *Economic Journal* 19:150–154.
Taylor, John B. (1980). Aggregate dynamics and staggered contracts. *Journal of Political Economy* 88(February):1–23.
Woodford, Michael. (1990). Self fulfilling expectations and fluctuations in aggregate demand. NBER Working Paper *3361,* May.
Zink, Helmut. (1989). Increasing returns, quality uncertainty, product differentiation, and countercyclical pricing. Mimeo, Dept. Economics, Univ. Bern, September.

Comment

N. GREGORY MANKIW
Harvard University and NBER

Reading this paper by Rotemberg and Woodford reminded me of the time I decided to give up macroeconomics. It was about a dozen years ago. I was an undergraduate at Princeton, and I had just been taught God's truth about how the economy works. The IS-LM model explains the downward-sloping aggregate demand curve. The upward-sloping aggregate supply comes from the assumption of a flexible price level and a nominal wage that is fixed in the short run.

Armed with these powerful tools of analysis, I reached the conclusion (completely on my own) that recessions must be quite popular. Sure, a few people get laid off. But most people get to enjoy the higher real wages that result when prices fall and their nominal wages do not.

So I went to one of my professors—Alan Blinder I think it was—to ask him about this. I had the vague recollection that recessions were, in fact, politically unpopular, but this just did not make any sense to me. If high real wages accompanied low employment, as *The General Theory* and my professors had taught me, then most households should welcome economic downturns.

Well, Professor Blinder admitted to me that real wages do not move countercyclically. My conclusion did follow logically from the theory I had been taught as God's truth, but it just did not fit the facts. It was at that point that I decided to abandon macroeconomics. After all, how could I trust my macro textbook again? If (as a mere undergraduate) I had managed to uncover this big lie, how many more big lies remained undetected? I decided to stick to microeconomics.

As one can see, my resolve weakened over time. Yet I have never stopped being disturbed by the cyclical behavior of the real wage. Over the years, I have kept my eye on the various ways this real-wage puzzle can be resolved. In this paper, Rotemberg and Woodford explore one possible resolution, suggested by Rotemberg and Saloner's (1986) earlier work on supergame models of price wars during booms. I must admit that this Rotemberg–Saloner–Woodford explanation of the real wage at first seems somewhat unlikely. But it starts to seem more appealing when compared with the alternatives.

As far as I know, there are six ways to explain the failure of real wages to move countercyclically over the business cycle. They are summarized in Table 1. None of these explanations commands a consensus among macroeconomists, and none leaves me completely satisfied.

One explanation is that the business cycle is driven by technology shocks, as in real-business-cycle models. When the available technology worsens, the economy goes into a recession, and the marginal product of labor and the real wage fall. We all have our views about the plausibility of

Table 1 WHY AREN'T REAL WAGES COUNTERCYCLICAL? SIX ANSWERS

1. Technology shocks are the source of economic fluctuations.
2. Implicit contracts smooth the real wage over time.
3. The marginal product of labor does not diminish as employment rises.
4. Firms set prices based on long-run average cost.
5. Prices are about as sticky as nominal wages.
6. Desired markups over marginal cost are countercyclical, perhaps because oligopolistic collusion is harder to maintain in booms.

this story, so I will not discuss it anymore here. I will say, however, that this explanation may suffer from the opposite problem from the traditional Keynesian story. Real-business-cycle models tend to imply a strongly procyclical real wage. Although the real wage is not countercyclical, it is also not strongly procyclical. Real-business-cycle models appear to have as much trouble explaining the real wage as traditional models.

A second hypothesis is that the cyclical behavior of the real wage is meaningless, because the real wage does not reflect the true shadow price of labor. Instead, implicit contracts between workers and firms keep the real wage steady while the shadow price fluctuates. Although I find this view somewhat appealing, I do not know of any direct evidence to support it. If this explanation were right, we should observe more cyclical real wages in industries with weak attachment between workers and firms. Yet I suspect that this prediction is probably just not true.

A third explanation of the real wage is that the marginal product of labor does not diminish as employment rises, so that the labor demand curve is horizontal. As Robert Hall puts it, marginal cost is "as flat as a pancake." In his paper for this conference, Hall points out that Dornbusch and Fischer (1990) take this approach in Chapter 13 of their textbook. Certainly, if one is willing to abandon the law of diminishing returns, then it is easy to explain a real wage that is roughly acyclical. It also becomes easier to explain a variety of other phenomena, such as the failure of the production-smoothing model of inventories.

I personally find this view a bit implausible. A flat labor demand schedule implies that if the government (or a national union) were to raise real wages by only a few percent, employment would fall by huge amounts, yet I think few economists would predict that outcome. The law of diminishing returns seems as sound an axiom as any on which we economists rely, even though it has been out of fashion recently. Diminishing returns, of course, leads to downward-sloping labor demand and upward-sloping marginal cost. Moreover, as Mark Bils has emphasized, the tendency of marginal cost to increase during booms is strengthened by firms' increased reliance on more expensive overtime labor.

I am somewhat inclined to believe the view expressed in the *first* chapter of Dornbusch and Fischer, which says, "A key fact about aggregate supply is that it is nonlinear. At low levels of output, prices do not change much. . . . But as the economy gets close to full employment . . . , further increases in output will be accompanied by increased prices." Here Dornbusch and Fischer suggest that marginal cost is not only increasing, but it is increasing at an increasing rate. This conclusion would follow naturally if firms have fixed capacity in the short run, and if more firms hit their capacity in booms than in recessions. In other words, marginal cost is not as flat as a pancake; it is as curved as a croissant.

A fourth explanation of the real wage, which Jim Tobin suggested to me several years ago, is that firms do not set prices based on marginal cost. Instead, they set prices based on long-run average cost, which in turn is proportional to nominal wages. If one followed Alan Blinder (1991) and asked firms why they do what they do, Tobin's hypothesis would likely fare quite well. The argument against this hypothesis is that it seems to require that firms not maximize profit. The assumption of profit maximization is, I suspect, a fetish that few economists are willing to give up easily.

A fifth explanation of the cyclical behavior of the real wage is that prices are about as sticky as nominal wages. As one might guess, this is the explanation that I tend to favor. In fact, it was thinking about the real-wage puzzle that originally got me interested in thinking about imperfections in goods markets and, eventually, about monopolistically competitive firms facing menu costs (Blinder and Mankiw, 1984; Mankiw, 1985). Alan Blinder's survey evidence indicates that the typical firm in the U.S. economy changes its prices about once a year. This is roughly the time span over which many nominal wages are fixed. So, as I read the evidence, the hypothesis of equally sticky wages and prices seems fairly attractive.

The sixth and final hypothesis, which is examined in this paper by Rotemberg and Woodford, is that the desired markup of prices over marginal cost is countercyclical. This could happen simply because the elasticity of demand changes over the business cycle. But I share their judgment that this seems unlikely. Instead, if markups fall during booms, it must be that industries in some way become more competitive.

Deciding among these six explanations is, of course, an empirical issue, and it is exactly the issue that Rotemberg and Woodford take up. Their paper is impressive in the way it brings to bear a variety of evidence—time series, cross sectional, and case study—to provide support for their model of countercyclical markups. Yet I am left with an uneasy feeling about their interpretation of the evidence.

I have two reservations. First, if we were to follow Blinder's strategy and ask businessmen if they behaved in this way, they would probably tell us that we were crazy. The level of sophistication in these supergame models seems just too great to describe realistically how firms set prices. I must admit that the more I talk to real businessmen—such as textbook publishers—about how they set their prices, the less compelling I find the assumption of highly sophisticated, fully rational, profit maximization.

My second and perhaps more important reservation is that the evidence that Rotemberg and Woodford present can, I suspect, be explained in other ways. They present many pieces of evidence, all pointing in the same direction. Yet each piece of evidence on its own seems potentially controvertible.

For example, they present an ingenious test in which they examine how Tobin's q affects the markup. Yet, as they point out, their method for computing the change in the markup is closely related to calculations of the Solow residual. Therefore, many of the standard problems with interpreting Solow residuals arise here as well. In particular, their calculations would seem to be affected by labor hoarding, by which I mean unmeasured variation in workers' effort.

Similarly, Rotemberg and Woodford report that there are systematic cross-industry differences in the cyclical behavior of the real product wage. Those industries that are more concentrated have more procyclical real product wages. This is an important fact, and their model can explain it. But so can other models. In their 1987 paper, Rotemberg and Saloner examine the relative rigidity of monopoly pricing. They use a menu–cost model to show that greater concentration should lead to stickier prices. Naturally, if prices are stickier in highly concentrated industries, real product wages will tend to be more procyclical.

In the end, I remain skeptical of the supergame model of markups. Yet I find it more appealing than I did before reading this paper. Unlike most papers that I read, this one actually moved my priors. Countercyclical markups may actually be part of the explanation for why real wages are not countercyclical. As long as I get to keep reading papers like this one, I will not give up on macroeconomics.

REFERENCES

Blinder, Alan S. (1991). Why are prices sticky? Preliminary evidence from an interview survey. *American Economic Review Papers and Proceedings* 81(May): 89–96.

Blinder, Alan S., and N. Gregory Mankiw. (1984). Aggregation and stabilization policy in multi-contract economy. *Journal of Monetary Economics* 13:67–86.

Dornbusch, Rudiger, and Stanley Fischer. (1990). *Macroeconomics*, 5th ed. New York: McGraw-Hill.

Mankiw, N. Gregory. (1985). Small menu costs and large business cycles: A macroeconomic model of monopoly. *Quarterly Journal of Economics* 100(May): 529–539; reprinted in *New Keynesian Economics*, Vol. 1. N.G. Mankiw and D. Romer (eds.). Cambridge, MA: MIT Press, 1991.

Rotemberg, Julio J., and Garth Saloner. (1986). A supergame-theoretic model of price wars during booms. *American Economic Review* 76(June):390–407; reprinted in *New Keynesian Economics*, Vol. 2. N.G. Mankiw and D. Romer (eds.). Cambridge, MA: MIT Press, 1991.

———. (1987). The relative rigidity of monopoly pricing. *American Economic Review* 77(December):917–926.

Comment

VALERIE A. RAMEY
University of California, San Diego and NBER

1. Introduction

Rotemberg and Woodford present a thorough analysis of the case for a collusive model of the business cycle. Their paper makes two main points. First, they argue that in order to reproduce the effects of aggregate demand shocks one needs to introduce imperfect competition. Second, they suggest that the type of imperfect competition supported by the data is the implicit collusion model of Rotemberg and Saloner (1986). The result is a nice blend of macroeconomics and industrial organization, in both the theory and empirical work, with extraordinary attention to detail.

One of the real strengths of the paper is that instead of embroiling themselves in the question "are demand shocks or supply shocks more important," Rotemberg and Woodford, following Barro (1981) and Hall (1986), focus on military spending as a clear shift in aggregate demand. The authors argue that the competitive model has counterfactual predictions for the effect of increases in military spending. In the competitive model, a temporary increase in defense spending increases the real interest rate, which leads to increased hours and output, and decreased consumption. The increase in output occurs because of a shift in labor supply. Under the assumption of diminishing returns to labor, and absent effects on the production function, real wages should fall.

Rotemberg and Woodford suggest that these implications are at odds with the data. They argue that real wages and consumption, in fact, increase in response to an increase in defense spending, despite the increase in hours. Thus, defense spending cannot have its effect solely through shifts in labor supply. The key to Rotemberg and Woodford's alternative model is that collusive behavior that leads to countercyclical markups allows aggregate demand shocks to affect labor demand. The augment is simple. Consider the following equation from their paper:

$$F_H(K_t, H_t, z_t) = \mu_t w_t, \tag{1}$$

where the left-hand side is the marginal product of labor (H) and the right-hand side is the product of the markup μ and the real wage w. The competitive model assumes that μ is equal to one and invariant. Rotemberg and Woodford's collusive model implies that μ is greater than one, and is countercyclical. Thus, in their model, an increase in hours, which lowers the left-hand side does not mean that the real wage must fall. Rather, μ will fall. Hence, their model is consistent with a concurrent increase in military spending, hours, real wages, output, and consumption.

To support their arguments, Rotemberg and Woodford provide four types of empirical results: (1) the effect of military spending on hours, output, consumption, and real wages, (2) the cyclicality of the markup, (3) tests of three imperfect competition models, and (4) the relationship between markups and concentration in industries. I will argue that the regularities in the data that they cite are not regularities at all, and that their measure of the countercyclicality of the markup is based on implausible assumptions. I will address each of these in turn.

2. *The Effects of Military Spending*

Does military spending have the effects claimed by the authors? In fact, Rotemberg and Woodford's own estimates do not support their claims. First, Table 1 of their paper shows clearly that military spending depresses consumption, since the only significant coefficient on military spending lags is a negative coefficient and is much larger than the sum of the positive coefficients. This is the same result Hall (1986) found. Second, the coefficient estimates of the effect of military spending on hours and wages and compensation are generally not significant. Moreover, for a given lag, the values of the coefficients in the hours regression and the real compensation regression have opposite signs, meaning that military spending has opposite effects on the two variables.

Thus, the main empirical regularities the authors use to argue that imperfect competition is necessary are not regularities at all. A competitive model with more general functional forms for the utility and production functions could probably capture most of the movements.

3. *Countercyclical Markups*

Rotemberg and Woodford calculate markups using the Hall methodology, but also allow for overhead labor. The weakness of this approach is that the *cyclical* behavior of the markup depends crucially on the *average* level of the markup. When the elasticity of substitution between capital and variable labor is equal to one (their baseline case), the formula appears as follows:

$$\hat{\mu}_t = \hat{y}_t - f(\mu^*)\hat{h}_t - \hat{w}_t, \tag{2}$$

where $\hat{\mu}$ is the cyclical variation in the markup, $\hat{y}$ is the cyclical variation in value added, μ^* is the average level of the markup, $\hat{h}$ is the cyclical variation in hours, and $\hat{w}$ is the cyclical variation in the wage rate. f is an increasing function of μ^*. The authors' baseline case is $\mu^* = 1.6$, implying an average markup of 60%. When $\mu^* = 1.6$, $f(\mu^*) = 2$. Thus, it is not surprising that they find very countercyclical markups, since the coefficient on the negative of hours is equal to two. When they allow for a markup of 10%, which substantially lowers the coefficient on the negative of hours, they find that the markup is actually procyclical.

Is it reasonable to suppose that the average markup is 60%? This number is consistent with Hall's (1988) findings, but micro studies, such as Bresnahan's (1981) study of the auto industry, find markups on the order of 10%. In fact, in Rotemberg and Woodford's model a markup of 1.6 implies an implausibly high ratio of overhead labor to total labor. By their Equation (4.3), an average markup of 1.6 implies that the ratio of overhead labor to total labor exceeds 50%! Simple evidence suggests this number is too high. Consider total employment in manufacturing minus the employment of production workers in manufacturing (from CITIBASE) as an upper bound on the number of overhead workers. (This number is an upper bound, because it shows clear cyclical variation, declining by 15% during recessions.) The ratio of nonproduction workers to total employment in manufacturing has increased over time, but has never exceeded 30%. If we consider 20% to be a reasonable ratio, then the average markup must be 1.16. Such a value corresponds to a value of $f(\mu^*)$ of 1.2, significantly reducing the countercyclicality of the markup.

Let us consider some simple alternative evidence on the cyclicality of the markup. From Equation (2.3) of the paper, we have the following relationship between profit rates and markups:

$$\pi_t = (1 - \frac{1}{\mu_t}) - \frac{FC_t}{p_t y_t}, \tag{3}$$

where π is the ratio of (after-tax) profits to the value of sales, μ is the markup, FC is fixed cost, and py is the value of sales. It is easy to see from Equation (3) that in the absence of fixed costs, procyclical profits rates imply that the markup should be procyclical. In fact, profit rates are procyclical. Figure 1 shows quarterly after-tax profit rates in manufacturing from 1947:1 to 1990:2 from CITIBASE. Note that this series is from reported profits, and is not based on any assumptions about market structure or production functions. Profit rates are clearly procyclical.

Thus, by Equation (3), the only way to reconcile countercyclical markups with procyclical profit rates is to allow for the presence of significant fixed costs.

To investigate the cyclicality of the markup in the presence of fixed costs, I estimate the following simple model. I assume that fixed costs are proportional to nonproduction worker employment in manufacturing times average hourly earnings in manufacturing. (All series are from CITIBASE.) This variable is divided by the value of sales in manufacturing to capture the last term in Equation (3). I then regress the profit rate in manufacturing on this variable as well as the economywide unemployment rate. The unemployment rate (U) is meant to capture the cyclicality of the markup. If the unemployment rate enters negatively in the regression, this is an indication that the markup is procyclical; if it enters positively, this is an indication that the markup is countercyclical. The results are given as follows:

$$\pi_t = \text{constant} + \text{trend} - \underset{(-2.99)}{12.105}\,\frac{FC_t}{p_t y_t} - \underset{(-2.67)}{0.217}\,U_t \tag{4}$$

$$R^2 = 0.570,\ DW = 0.952$$

Figure 1 PROFIT RATES IN MANUFACTURING

With an AR(1) correction:

$$\pi_t = \text{constant} + \text{trend} - \underset{(-4.80)}{39.645}\ \frac{FC_t}{p_t y_t} - \underset{(-0.10)}{0.009}\ U_t \tag{5}$$

$$R^2 = 0.609,\ DW = 2.002,\ \rho = 0.848$$

The simple OLS estimates given in Equation (4) indicate that the unemployment rate enters negatively, suggesting *procyclical* markups. There is, however, substantial serial correlation, so Equation (5) reports the estimates with an AR(1) correction. Here, the coefficient on unemployment is still negative, but indistinguishable from zero. Thus, these estimates imply an acyclical markup. These results are only suggestive, but coupled with the arguments above on the size of the average markup, they cast doubt on Rotemberg and Woodford's finding of pronounced countercyclicality of the markup.

4. *Tests of the Collusive Theory*

The main implication of the implicit collusion hypothesis is that the markup should increase when future profitability increases, and decrease when current output increases. Rotemberg and Woodford test their theory against two competing theories (monopolistic competition and customer markets) by estimating the relationship between their measure of markups on current output and a measure of future profitability. Because future profitability is unobservable, they must use a proxy. They use two methods: (1) the q method and (2) estimation of a difference equation. They are very careful in their estimation, using several different estimation procedures for each method. In general the results support the collusive model. However, these results are entirely dependent on their estimate of the markup. The negative effect of y is due to their finding that the markup is countercyclical.

It seems that there is a simpler way to estimate the impact of future changes in profitability on the markup. A known change in future corporate tax rates will affect future profitability. Therefore, according to their model if tax rates are expected to decrease in the future, the current markup should decrease. Such evidence would complement the evidence they present in their paper.

5. *Concluding Comments*

In sum, Rotemberg and Woodford have presented provocative evidence for a collusive model of the business cycle. However, neither their evidence against the competitive model nor their evidence for the collusive model is completely compelling. Other models, such as models with external increasing returns, can produce many of the same results. Thus, the evidence available thus far is not decisive.

REFERENCES

Bresnahan, Timothy. (1981). Departures from marginal-cost pricing in the American automobile industry: Estimates for 1977–78. *Journal of Econometrics* 17(November):201–227.

Barro, Robert. (1981). Output effects of government purchases. *Journal of Political Economy* 89(December):1086–1121.

Hall, Robert E. (1986). The role of consumption in economic fluctuations. In *The American Business Cycle.* Robert J. Gordon (ed.). Chicago, IL: The University of Chicago Press.

———. (1988). The relation between price and marginal cost in U.S. industry. *Journal of Political Economy* 96(October):921–948.

Rotemberg, Julio J., and Saloner, Garth. (1986). A supergame-theoretic model of price wars during booms. *American Economic Review* 76(June):390–407.

Discussion

In response to Valerie Ramey, Julio Rotemberg offered that one alternative model to accommodate various observations is declining marginal costs. Rotemberg and Woodford assume increasing returns through the presence of fixed costs, but they do not assume declining marginal costs. He doubts that internal declining marginal costs explains the puzzles because one does not observe firms shutting down and bunching production. Also, in recessions firms choose to shut down certain plants that likely have higher marginal costs. External declining marginal costs is problematic because no one has developed a convincing story of what they are.

Martin Eichenbaum raised two points concerning the acyclicality of real wages. First, he pointed out that labor hoarding and varying capital utilization rates in a competitive model with shocks to aggregate demand can easily accommodate the lack of a correlation between wages and hours worked. Second, he argued that many shocks hit the economy and that one need not have a model that provides a zero correlation for every shock. Gregory Mankiw, however, noted that real wages were not particularly high in the 1982 recession, which was plausibly caused by monetary authorities. Valerie Ramey also pointed out that the same is true when one uses military purchases to identify aggregate demand shocks. Rotemberg indicated that periods of high rates of capital utilization should be accompanied by low rates of investment. Investment, however, is procyclical, and therefore variations in capital utilization will not solve the puzzle.

Robert Hall illustrated the differences between his paper and Rotemberg and Woodford's. Whereas he suggests that the lack of movement in real wages results from a flat labor demand curve, Rotemberg and Woodford offer that the marginal cost curve is steep, and countercyclical markups lead to acyclical real wages. Their different conclusions result from different identifying assumptions. Hall assumes that variations in markups are uncorrelated with his aggregate demand instruments and subsequently measures a large elasticity of labor demand. Rotemberg and Woodford assume a smaller value for the elasticity of labor demand and find that markups are countercyclical. Otherwise, the regression equation both develop are functionally equivalent.

Robert Gordon indicated that he had initially thought that the acyclicality of real wages arose from a mixture of demand and supply shocks. He believed that real wages were procyclical in the 1970s and 1980s, and that once one removed supply shocks,. real wages were countercyclical. He presented some evidence, however, that proved his priors to be mistaken. After detrending real wages by more than a single trend, one does not find a significant negative correlation with capacity utilization rates over the 1960s, early 1970s and late 1980s.

Mark Bils suggested that straight-time wages are a poor measure of the marginal cost of an extra unit of labor. Robert Barsky argued that his work, as well as work by Kydland and Prescott, Stockman, and others cited in the paper, advises against the use of aggregate real wages. Composition effects dominate.

Olivier Blanchard asserted that simultaneity bias in the regressions involving q was more problematic than was treated in the paper. For example, the change in the share of capital in Europe has increased 5–10% over the last decade. This is likely independent of changes in collusion. Moreover, shocks such as this will move markups and stock market evaluations. Thus, errors in the regression equation will be correlated with all of the variables. Rotemberg agreed and stated that they plan to develop a more structural model allowing for exogenous variations in markups.