JOHN LEAHY

A Survey of New Keynesian Theories of Aggregate Supply and Their Relation to Industrial Organization

I survey the recent literature on the Phillips curve. Along the way, I will try to relate this literature to topics of interest to industrial organization. I will also point out the gaps in our understanding and places where more careful micro-economic analysis would be helpful to macroeconomists. In the conclusion, I summarize what an industrial organization economist might take away from this literature.

JEL codes: E3, L16
Keywords: New Keynesian, price stickiness, aggregate supply, strategic complementarity, industrial organization.

When a business cycle economist is asked about industrial organization, a typical response will be something like, “What can we learn about business cycles from the market for yogurt?” When an industrial organization economist attends a macro-economics seminar, a typical question will be, “You assume the same elasticity of demand for all goods?” I think that it is fair to say that the two groups tend to talk past each other. This is unfortunate since there is considerable overlap in topics that the two fields find important. Both are interested in the determinants of markups and marginal cost. Both are interested in the interaction between firms’ pricing strategies. The gap is more a reflection of the type of questions that the two groups attempt to answer. Industrial organization economists tend to be interested in the interactions among firms in a particular market setting, such as the auction of timber licenses or the pricing of electricity, cement, or automobiles, whereas macroeconomists are concerned with the evolution of aggregate quantities, such as the growth of output over time, fluctuations in employment over the business cycle, or the comovement of output and inflation.

This paper was written for the “Dialogue between Micro- and Macroeconomics Conference” held in Bern in 2010. I would like to thank Chris House, Virgiliu Midrigan, and Mark Gertler for helpful discussions, and one anonymous referee for valuable comments. This research was supported by the National Science Foundation under grant no. SES-0648545.

John Leahy is a Professor in the Department of Economics, New York University and N.B.E.R. (E-mail: john.leahy@nyu.edu).

Received September 2, 2010; and accepted in revised form November 3, 2010.

Journal of Money, Credit and Banking, Supplement to Vol. 43, No. 5 (August 2011) © 2011 The Ohio State University
These differences in focus lead to differences in model strategy. Because macroeconomists focus on aggregates, their models tend to be dynamic and general equilibrium. Following Lucas’ famous critique of policy evaluation, macroeconomists have also found it important that their models have solid microfoundations, since the effects of any policy action depend on how agents respond to their new environment. In addition, macro-economic models are often stochastic, either because they are interested in the effects of uncertainty *per se* or because they wish to consider the response of the economy to various economic disturbances.

These four elements—uncertainty, dynamics, general equilibrium, and microfoundations—put strong limits on the complexity of macro-economic models. The more complex the model, the more difficult it is to solve. The more complex the model, the more difficult it is to understand.

For this reason, macroeconomists try to make their models complex enough to answer interesting questions, while simple enough that they can be solved and understood. These simplifications, such as a representative consumer, symmetric firms, or a constant elasticity of demand, may seem crude or even laughable to an industrial organization economist used to analyzing a given industry or set of firms, but in macroeconomics, every new model element must pass a simple test: does it matter for the question at hand? If a question does not appear to depend on the distribution of wealth or consumption across agents, then it makes sense to assume a representative consumer. If a question does not appear to depend on the distribution of markups across firms, then it makes sense to assume that all firms have the same elasticity of demand. As we get better at solving and testing more complex models, we revisit these assumptions and amend the models where needed. Because industrial organization looks with a microscope at particular markets and particular interactions, details become much more important. Many of these details wash out in the aggregate.

Given that macroeconomics is such a broad field, I will focus on one area of macroeconomics: New Keynesian models of aggregate supply. These models attempt to explain two general empirical regularities linking real and nominal variables: the apparent real effects of monetary policy and the correlation between output and inflation over the business cycle, also known as the Phillips curve. These models appeal to sticky prices to explain the link between real and nominal variables. If prices are sticky, then increases in nominal demand will lead to increases in real output. Increases in real output may then pressure firms to overcome price inertia leading to inflation.

This area of macroeconomics is of particular relevance to industrial organization, because in order to model price rigidity, one needs firms that actually set prices. This takes us out of the realm of perfect competition that characterizes many real models of business cycles. It also opens the door for the consideration of interesting strategic interactions in firms’ price setting behavior.

Firms set prices in New Keynesian models to achieve desired markups over marginal cost over some specified time horizon. This characterization highlights

1. For a more in-depth treatment of this subject, see Woodford (2003).
the three ingredients of the typical model: a model of price frictions, a model of the markup, and a model of marginal cost. Models differ on each of these margins. Some models simply assume that prices are fixed for some given period of time. Other models endogenize the firm’s choice of when to change its price. Some models assume desired markups are fixed. Others assume that the elasticity of demand depends on the actions of the firm. Some models assume constant marginal cost. Others assume that marginal cost depends in complex ways on a firm’s production, firms’ interactions, or adjustment frictions. Still others assume that marginal cost is subject to the same sort of nominal frictions that affect prices.

It is safe to say that the profession has been quite creative in constructing potential models of the interaction between prices and costs. We have been less successful at identifying which of these interactions are important in practice. In what follows, I survey some of the recent literature on the Phillips curve. Along the way, I will try to relate this literature to topics of interest to industrial organization. I will also point out gaps in our understanding and places where more careful micro-economic analysis would be helpful to macroeconomists.

Four themes dominate this discussion. The first is that price inertia is pervasive. Price rigidity plays a central role in macroeconomics, but only a minor role in industrial organization.

The second theme is that measurement is difficult, and careful measurement is very important. The properties of business cycles depend on many concepts, such as markups and marginal cost that are not directly observable. Instead they must be deduced from a model, and different models give different answers. Even price, which is a variable on which we have a lot of data, is not a simple variable to measure. Questions arise as to which prices are important. Do we care about prices of all firms equally, or are the prices of some firms more important? For a given firm, do we care about all prices equally, or are some prices, such as nonsale prices, more important?

The third theme is that some of the model elements, such as diminishing returns, countercyclical markups, or wage rigidity, that macroeconomists introduce into their models in order to increase the real effects of nominal disturbances, have subtle roles to play. Introduced in one way, they have the desired results. Introduced in another, they may produce the opposite. For example, diminishing returns, when introduced at the firm level, amplify the real effects of nominal shocks, whereas diminishing returns introduced at the level of the economy as a whole tend to mitigate the effects of these shocks. In order to understand the relationship between prices and costs over the cycle, we need to understand the precise mechanism underlying these movements and the scale upon which they operate.

The final theme is the importance of generalization. Addressing all of these issues requires careful study of firms and their environment, a strength of industrial organization. At the same time, we need conclusions that generalize to the economy as a whole, if we are going to import the results into macro-economic models.
1. THE NEW KEYNESIAN AGGREGATE SUPPLY CURVE

1.1 Price Setting with Nominal Rigidities

I focus on the supply side of the economy, where nominal rigidities lead to deviations from the frictionless optimum. Time is discrete and indexed by \( t \). There is a continuum of firms of unit mass indexed by \( i \). Each firm produces a single consumption good and chooses its nominal price to maximize the present value of profits.

The central idea behind New Keynesian economics, much like old Keynesian economics, is that money matters because there are frictions in price setting. It is important that these are nominal frictions, since real frictions would not impose any difficulties in responding to nominal shocks. Real frictions, for example, do not prevent all nominal prices from doubling in response to a doubling of the money supply. If there are frictions in nominal price setting, then firms will tend to choose prices so that they hit their desired markups over nominal marginal cost on average. The typical New Keynesian model, therefore, has three ingredients: a pricing friction, a markup equation, and a cost curve.

The first ingredient of the model is the pricing friction. Any model of price inertia must confront the fact that many prices are not just sluggish, they are fixed in nominal terms for periods at a time. The median price in the U.S. economy is fixed for about 3.3 months (Klenow and Malin 2010). It is very difficult to match this fact without imposing some restrictions on firms’ ability to adjust prices. There are two popular approaches in the literature. Models with time-dependent pricing rules assume that the timing of nominal price changes is exogenous. The two most popular time-dependent rules are “Taylor pricing” (Taylor 1980), in which a firm resets its price at fixed intervals, and “Calvo pricing” (Calvo 1983), in which firms adjust their price in any given period with some fixed probability \( 1 - \alpha \) (writing this probability as \( 1 - \alpha \) simplifies certain mathematical expressions). The second approach is to allow firms to choose when to alter their prices and to assume that some fixed cost of price adjustment prevents continuous adjustment. These rules are often referred to as “state-dependent rules,” since the decision whether or not to adjust depends on the state in which the firm finds itself, or “Ss rules” since the trigger-target character of the optimal policy is reminiscent of the Ss model of inventory management.

It may seem that state-dependent rules are inherently better than time-dependent rules, since state-dependent rules arise out of a well-formulated decision problem, whereas time-dependent rules impose \textit{ad hoc} restrictions on firms’ pricing strategies. Allowing for state dependence, however, has significant costs. The economy-wide cross-sectional distribution of prices often becomes a state variable in state-dependent pricing models. For example, the greater the number of firms that have prices below their optimum the more likely is inflation. Given the complexity of the optimal strategies, it becomes difficult to obtain simple an-

---

2. Levy and Young (2004) report an extreme example: the price of a 6.5-oz Coca Cola was fixed at 5 cents for over 70 years between 1886 and 1959.
alytic characterizations. Often these costs are not balanced by any great benefit. It turns out that in many applications the distinction between state and time dependence does not matter very much. When firms adjust their prices, they tend to adjust them by large amounts—in the neighborhood of 10% on average (Klenow and Malin 2010). Since the average of these price changes is nowhere near this large, it follows that firms are hit by large idiosyncratic shocks to their desired prices. The timing of price adjustment in many state-dependent models that include such idiosyncratic shocks often depends more on the idiosyncratic shocks than on the aggregate state of the economy. From a macro-economic perspective, the timing of price adjustment is effectively exogenous. The result is that many state-dependent models have dynamic properties similar to time-dependent models (Gertler and Leahy 2008).

It is also not clear that the distinction between state and time dependence matters for the relationship between these models and industrial organization. Here, the exact form of price stickiness is less important than the fact that there is some horizon over which prices are fixed. I will, therefore, begin by assuming Calvo pricing, since it leads to the simplest mathematical formulations. I will assume that each firm sells a single good and is allowed to alter the price of its good with a fixed probability $1 - \alpha$ each period. In a later section, I will return to state-dependent pricing models and discuss where and when state-dependence matters.

The second ingredient is the markup. It is common to assume a demand function of the form

$$Y_i = D \left( \frac{P_i}{P}, Y \right) = \left( \frac{P_i}{P} \right)^{-\varepsilon} Y.$$  

(1)

In this formulation, the demand for firm $i$ depends on aggregate demand $Y$ and its price $P_i$ relative to a price index $P$. The elasticity of demand $\varepsilon$ is constant. This demand curve is what would come out of a model with two-stage budgeting, such as Dixit and Stiglitz (1977), in which consumers first decide how much to consume and then how to allocate that consumption across goods.

The constant elasticity of demand implies that a firm’s desired markup of price over marginal cost is constant and equal to $\varepsilon / (\varepsilon - 1)$. The industrial organization literature is full of models of time-varying markups. Few of these models have been incorporated into the macro-economic literature. In the typical New Keynesian model, markups vary over the business cycle, not because the desired markup varies, but because firms’ prices are fixed and costs vary. I will begin with the constant elasticity specification. Later, I will discuss theories of time-varying markups and how they might be incorporated into macro-economic models.

The third ingredient is a model of the cost of production. Here, I will simply assume that the real cost of production depends on the level of production, as well as the level of aggregate activity:

$$C(Y_i, Y) = C \left( D \left( \frac{P_i}{P}, Y \right), Y \right).$$
A firm’s marginal cost could depend on its own production through returns to scale or contractual relationships such as overtime pay. It could depend on the level of aggregate activity through factor prices. Greater use of a factor raises the price of that factor to all firms.3

Putting these elements together, the problem of a firm that adjusts its price at date \( t \) is

\[
\max_{P_t} E_t \sum_{k=0}^{\infty} \alpha^k \Lambda_{t,t+k} \left[ \left( \frac{P_t}{P_{t+k}} \right)^{-\varepsilon} Y_{t+k} \frac{P_t}{P_{t+k}} - C \left( D \left( \frac{P_t}{P_{t+k}}, Y_{t+k} \right), Y_{t+k} \right) \right],
\]

where \( E_t \) is the expectations operator conditional on date \( t \) information, and \( \Lambda_{t,t+k} \) is the stochastic discount factor, that is, the ratio of marginal utility at date \( t+k \) to marginal utility at date \( t \), all multiplied by a factor \( \beta^k \). The term in brackets is real revenue minus real costs at date \( t+k \), where equation (1) has been used to substitute for demand. Note that \( P_t \) is fixed across time and future profits are discounted by \( \alpha^k \), the cumulative probability that the firm has not had an opportunity to change its price. In this problem, the firm’s horizon is the period over which its price is fixed. This is a direct result of the envelope theorem and the absence of any state variable besides prices: adjustment today does not affect profits at any date beyond the date of the next price adjustment.4

Differentiating with respect to \( P_t \), the first-order condition for this problem is

\[
E_t \sum_{k=0}^{\infty} \alpha^k \beta^k \left[ \left( \frac{P_t}{P_{t+k}} \right)^{-\varepsilon-1} Y_{t+k} \left( 1 - \varepsilon \right) \frac{P_t}{P_{t+k}} - \varepsilon C_1 \left( D \left( \frac{P_t}{P_{t+k}}, Y_{t+k} \right), Y_{t+k} \right) \right] = 0.
\]

We get a particularly simple relationship if we log-linearize this equation about a steady state with no inflation. Let lower case letters represent log deviations from steady-state values. The first-order condition becomes5

\[
E_t \sum_{k=0}^{\infty} \alpha^k \beta^k \left[ p_t^* - p_{t+k} - m c_{t,t+k} \right] = 0,
\]

3. This cost specification assumes that firms sell all of their production and do not hold inventories. Most New Keynesian models assume that the firm produces to meet demand at the posted price. See Kryvstov and Midrigan (2010) for an analysis of a New Keynesian model with inventories.

4. If we introduce additional state variables such as a firm-specific capital stock subject to adjustment costs or a customer base that must be built up over time, then today’s pricing decision will influence tomorrow’s pricing decision through its effect on this state variable. The envelope condition will no longer hold.

5. To derive (3), note that in steady state, the stochastic discount factor \( \Lambda_{t,t+k} \) is simply the discount factor \( \beta^k \), and that with zero steady-state inflation, the term in brackets in equation (2) is equal to zero in steady state. Fluctuations in the terms outside of the brackets, therefore, have no first-order effects.
where $p_t^*$ denotes the optimal price chosen by a firm that adjusts at date $t$ and $mc_{t,t+k}$ denotes the log deviation at date $t + k$ of real marginal cost from its steady-state value for a firm adjusting its price at date $t$. Note that both $p_t^*$ and $mc_{t,t+k}$, while specific to firm $i$, are common to all firms that adjust their prices at date $t$. I have, therefore, indexed them by $t$. Rearranging,

$$p_t^* = (1 - \alpha \beta) E_t \sum_{k=0}^{\infty} \alpha^k \beta^k [p_{t+k} + mc_{t,t+k}].$$

(4)

The optimal price is a weighted average of the firm’s future nominal marginal cost, $p + mc$, where the weights reflect discounting and the probability that the price will remain unchanged. This equation may be written recursively as

$$p_t^* = (1 - \alpha \beta)[p_t + mc_{t,t}] + \alpha \beta E_t p_{t+1}^*. $$

(5)

1.2 A Phillips Curve

We derive two versions of the New Keynesian Phillips curve: one that relates inflation to average marginal cost and one that relates inflation to output. In order to derive these curves, we need to replace firm specific variables $p_t^*$ and $mc_{t,t}$ in equation (5) with aggregates. We use the definition of the aggregate price level to replace $p_t^*$. In the neighborhood of the zero-inflation steady state, the log price index is a geometric average of individual firm prices

$$p_t = \int p_{ti} di.$$

Given the assumption of Calvo pricing, a random fraction $\alpha$ continue to charge their period $t - 1$ price in period $t$, and a fraction $1 - \alpha$ choose the new optimal price $p_t^*$. Since the nonadjusters are randomly selected from the population their average price is just $p_{t-1}$. The evolution of the price index is, therefore,

$$p_t = (1 - \alpha)p_t^* + \alpha p_{t-1}.$$

(6)

In order to replace $mc_{t,t}$, we note that the deviation of marginal cost of firm $i$ from its steady-state value may be written as

$$mc_i = \omega y_i + \sigma y,$$

(7)

where $\omega$ is the elasticity of a firm’s marginal cost with respect to own output and $\sigma$ is the elasticity of marginal cost with respect to aggregate output. This implies that the economy-wide average marginal cost is equal to

$$mc = (\omega + \sigma) y.$$
Moreover, linearizing the demand curve (1) about steady state gives
\[ y_t = -\varepsilon(p_t - p) + y. \] (9)

Equations (5)–(9) may be combined to yield two versions of the New Keynesian Phillips curve\(^6\)
\[ \pi_t = \lambda \frac{1}{1 + \omega \varepsilon} m_{c_t} + \beta E_t \pi_{t+1}, \] (10)
\[ \pi_t = \lambda \frac{\omega + \sigma}{1 + \omega \varepsilon} y_t + \beta E_t \pi_{t+1}. \] (11)

Here, \( \pi_t = p_t - p_{t-1} \) and \( \lambda = \frac{(1-\alpha)(1-\alpha\beta)}{\alpha} \). These equations relate inflation to expected future inflation and to current marginal cost or output. The higher is the current marginal cost or output, the higher are the prices set by firms that adjust their prices and the higher is inflation. The higher is the expected future inflation, the higher is the expected future marginal cost, the higher are the prices set by firms adjusting their prices, and the higher is current inflation.\(^7\) Note that the New Keynesian Phillips curve is forward looking; it relates inflation today to inflation tomorrow. This contrasts with the old Keynesian Phillips curve that relates inflation today to past inflation.

1.3 Real Rigidity and Strategic Complementarity

It is useful to define a number of variables. Let \( \psi_1 = \frac{1}{1+\omega \varepsilon} \) and \( \psi_2 = \omega + \sigma \), and let \( \psi = \psi_1 \psi_2 \). \( \psi_1 \) reflect the sensitivity of a firm’s optimal price to average marginal cost. \( \psi_2 \) is the elasticity of average marginal cost with respect to output. \( \psi \) combines these two elasticities and determines the sensitivity of a firm’s optimal price to output. The larger is \( \psi \), the greater is the response of prices to a given change in output.

To see this, suppose that there were no pricing frictions so that \( p_t = p + mc_t \) in all periods. Using (7), (8), and (9), it follows that
\[ p_t = p + \psi y, \]
\[ p_t = p + \psi_1 mc, \] (12)
so that \( \psi \) is the elasticity of the frictionless optimal price with respect to output and \( \psi_1 \) is the elasticity with respect to marginal cost.

Ball and Romer (1990) use the term “real rigidity” for the idea that a firm’s desired real price responds little to changes in real output. Small values of \( \psi \) are,

---

6. First, solve (7), (8), and (9) for \( mc \) as a function of \( mc \) and \( p_t - p \) and substitute for \( mc \) in (5). Then solve (6) for \( p_t \) and use this to replace \( p_t \) and \( p_{t+1} \) in (5). Finally, replace \( p_{t-1} \) with \( p_t - \pi_t \) and \( p_{t+1} \) with \( \pi_{t+1} - p_t \), and collect the \( p_t \) terms.

7. Note that we can iterate (10) forward and write current inflation as the expected present value of current and future marginal cost.
therefore, associated with real rigidity. Ball and Romer also characterize $\psi$ in terms of derivatives of the profit function. It is easy to show that their characterization holds in this setting. Let $\Pi(p_i - p, y)$ denote the period profits of firm $i$ written as a function of $p_i - p$ and $y$. Then,

$$
\psi = - \frac{d^2 \Pi}{d(p_i - p)dy} \frac{d^2 \Pi}{d(p_i - p)^2},
$$

which is Ball and Romer’s definition of real rigidity.

If, in addition, we assume that the velocity of money is constant, then the quantity equation implies

$$
y = m - p.
$$

Combining (12) and (13)

$$
p_i = (1 - \psi)p + \psi m.
$$

In this formulation, $\psi < 1$ is associated with strategic complementarity in price setting: a firm’s desired price is increasing in the prices charged by other firms.

It should not be surprising that low values of $\psi$ are associated with both real rigidity and strategic complementarity. It is precisely a firm’s desire to align its price with other prices that allows for large responses of output with only a little inflation.

### 1.4 Output and Inflation

Let’s return to the Phillips curve (11) and look at how the dynamics of output and inflation depend on the coefficient $\lambda \psi$. It turns out that in order to generate large and persistent effects of money on output, it is important that this product be small. Low values of $\lambda$ and $\psi$ increase both the magnitude and persistence of the response of output to nominal shocks.

To see this, suppose that velocity is constant so that equation (13) holds. Suppose that nominal income, $m = p + y$, is exogenous and that changes in $m$ are the only shock to the economy. Suppose that we begin in a steady state associated with a fixed level of $m$ and that $m$ rises permanently by 1%. Equation (4) says that the firm wishes to set its price equal to the present value of the frictionless optimal price. In this setting, the firm targets the present value of $\psi m + (1 - \psi)p$.

We consider two cases. Suppose first that $\psi = 1$, then the target price moves one for one with $m$. Since $m$ has risen by 1%, any firm that has the opportunity to adjust its price adjusts by 1%. The real effects of money are governed solely by the nonadjustment probability $\alpha$. $k$ periods following the increase in $m$, a fraction $1 - \alpha^k$ firms have fully responded to the shock, whereas the remaining $\alpha^k$ of firms remain at their original preshock prices. A high value of $\alpha$ leads to less aggregate price adjustment and, hence, a greater output response according to (13). A high value of $\alpha$ is associated with a slower rate of convergence of prices to their new steady state,
and hence, a more persistent effect of the money shock on output. A high value of \( \alpha \) is associated with a low value of \( \lambda \).

If \( \psi < 1 \), then the firms that adjust their prices raise their prices by less than 1%. Their marginal costs rise by a weighted average of \( p \) and \( m \), and since some prices are sticky, this average rises by less than 1%. Since prices adjust less on impact, a low value of \( \psi \) implies a larger impact of the money shock on output. Since adjusting firms no longer fully incorporate the increase in \( m \) into their new prices, convergence may take much longer. It may now take a firm many adjustments before a firm’s price converges to the new steady state. The response of output is, therefore, more persistent. Again, this persistence is associated with a low value of \( \lambda \psi \).

What determines \( \lambda \)? \( \lambda \) depends inversely on the nonadjustment probability \( \alpha \). There have been many recent efforts to estimate \( \alpha \) from data on individual prices. Klenow and Malin (2010) survey this evidence. For the United States over the period 1988–2005, the median price changes about once every 3.3 months. If the time period is a quarter, this implies a value of \( \alpha \) in the neighborhood of 0.53 and a value of \( \lambda \) in the neighborhood of 0.4.8 By comparison, Gali and Gertler (1999) estimate the coefficient on marginal cost \( \lambda \psi \) to be between 0.01 and 0.05. This discrepancy could be closed either with a higher value of \( \alpha \) or a value of \( \psi \) less than 1.9 To get a value of \( \lambda \) equal to 0.05, we would need \( \alpha \) to be just above 0.8 or a median duration of prices in the neighborhood of 10 months.

There are several reasons to believe that it may not be appropriate to calibrate \( \alpha \) to the median price change. First, as argued by Kehoe and Midrigan (2007) and Guimaraes and Sheedy (2008), we may wish to exclude sales from the calculation. Firms often return to the presale price after the sale is over. While it remains a puzzle why firms would find it optimal to return to the presale price, the consequence is that the economy has not responded to the monetary shock until the presale price has responded to the monetary shock. If we exclude sales from our calculations, the median duration of prices in the United States rises to 6.9 months, and if we take a broader view of sales and define a sale to be any deviation from a modal or reference price (Eichenbaum, Jaimovich, and Rebelo 2008), the median duration rises to 10.6 months (Klenow and Malin 2010). Clearly, the treatment of sales matters. Determining the extent to which sales respond to aggregate information is an area of active research, one to which the field of industrial organization has potentially much to offer.

Second, there is great heterogeneity across firms and across sectors in the average duration of prices. For example, the median duration for durable goods is 1.8 months. Prices for nondurable goods change every 3.3 months. Prices for services change every 7.7 months. Carvahlo (2006) argues that when this heterogeneity is

---

8. Given that adjustment is a binomial random variable, the time to next adjustment can be approximated as a continuous time-exponential random variable, implying a median waiting time equal to \(-\ln(2)/\ln(\alpha)\). At a quarterly frequency, \( \beta = 0.99 \).

9. Gali and Gertler implicitly assume \( \psi_1 = 1 \) when they assume that all firms have the same marginal cost.
taken into account, the best single choice for $\alpha$ is greater than the mean, resulting in a $\lambda$ that is one-third of the value associated with the mean. The intuition is simple: convergence to steady state is governed by the firms that have the stickiest prices. The firms with flexible prices adjust quickly and quickly become irrelevant for convergence. Carvahlo also argues that this downward adjustment in $\lambda$ should be greater if $\psi < 1$ because strategic complementarity causes the flexible price firms to mimic the behavior of the sticky price firms.

Little work has been done attempting to explain the differences in adjustment frequency across firms beyond correlating frequency with firm characteristics. Why are durable goods prices more flexible? Is it the way in which they are sold, or is it that buyers search more when faced with large infrequent purchases? Why are the prices of services sticky? Are these prices linked to the price of labor or do they involve long-lasting relationships between buyer and seller?

Not all considerations argue for higher $\alpha$. Below, we will show that state-dependent pricing models suggest using a smaller value of $\alpha$. Gertler and Leahy (2008) argue that this state dependence may argue for using a value for $\lambda$ that is three times the value calculated from the observed frequency of price adjustment. This adjustment roughly balances the effect of heterogeneity.

Given this evidence, most of the literature has taken the view that the average duration of prices in the United States is significantly less than 1 year, so that some form of real rigidity is necessary in order to generate large and persistent effects of money on output. Clearly, the final word on this issue remains to be heard.

What is $\psi$? Why $\psi$ should be small is a major gap in the literature. We have little direct empirical evidence on $\psi$. Rather, we have a collection of theories that put potential structure on $\psi$.

The approaches taken in the literature fall into two camps: mechanisms that work through $y_1$ relative to $y$ and mechanisms that work through $y$ alone (Kimball 1995). The former are incorporated in $\psi_1$ and the latter in $\psi_2$. A low $\psi$ can result either from a low elasticity of optimal price with respect to marginal cost, a low $\psi_1$, or a low elasticity of real marginal cost with respect to $y$, a low $\psi_2$. What follows is a brief review of the major theories and the structure that they place on $\psi$.

A low value of $\psi_2$ is associated with elastic factor supplies in the aggregate. For example, consider a world in which labor is the only factor, $Y_i = N_i$; consumption is the only output good, $Y = C = \sum N_i$; and labor is supplied in a single economy-wide labor market. In such a world, real marginal cost is equal to the aggregate real wage. If in addition, we assume a representative consumer with the following utility function

$$U(C, N) = \frac{C^{1-\theta} - 1}{1 - \theta} - \frac{1}{1 + \gamma} \left( \sum N_i \right)^{1+\gamma},$$

(14)

where $\theta$ is the coefficient of relative risk aversion and $\gamma$ is the inverse Frisch elasticity of labor supply, then the labor–leisure choice implies...
\[
\frac{W}{P} = C^\theta \left( \sum N_i \right)^\gamma.
\]

Using the production function and aggregate supply and demand, this first-order condition becomes
\[
\frac{W}{P} = Y^{\theta+\gamma}.
\]

(15)

In this world, real marginal cost is simply the real wage, and the elasticity of real marginal cost with respect to output is \( \sigma = \theta + \gamma \). Since the labor market is competitive, the elasticity of marginal cost with respect to own output \( \omega \) is zero. It follows that \( \psi_1 = 1 \) and \( \psi = \psi_2 = \sigma \).

Either a high Frisch elasticity of labor supply (low \( \gamma \)) or a low coefficient of relative risk aversion (low \( \theta \)) will deliver a low value of \( \psi_2 \). A high Frisch elasticity means that large changes in the labor supply requires only small changes in marginal cost. Low risk aversion implies that large changes in consumption generate small changes in the utility value of wages.

There are other ways to get elastic factor supplies. Each has its proponents in the literature. Some authors such as Christiano, Eichenbaum, and Evans (2005) and Smets and Wouters (2003) include monopsonistic labor markets, in which wages are sticky. In these models, sellers of labor services must meet demand at posted prices, thereby flattening marginal cost. Christiano, Eichenbaum, and Evans also include variable capital utilization. Basu (1995) and Huang and Liu (2001) consider models with intermediate inputs whose prices are sticky. One could also imagine constructing a model with external increasing returns to scale along the lines of Baxter and King (1991).

A low value \( \psi_1 \), on the other hand, requires some form of diminishing returns to scale at the firm level. This may be assumed directly by assuming that the production function has diminishing returns, or it may be assumed indirectly, by assuming that some factor such as land or capital is fixed in the short run and that variable costs depend on the other factors that are themselves subject to diminishing returns. Diminishing marginal utility performs much the same function. One popular choice is local labor markets plus diminishing marginal utility of leisure, so that when a firm expands production, it must pay a higher wage and its costs rise.

To see how diminishing returns works, we alter the production function in the example above. Suppose that \( Y_i = N_i^\eta \), where \( \eta < 1 \). It is easy to show that real marginal cost is now
\[
\frac{1}{\eta} \frac{W}{P} Y_i^{\frac{1}{\eta}-1}.
\]

In this example, the elasticity of marginal cost with respect to own output is \( \omega = \frac{1}{\eta} - 1 \). Since \( \psi_1 = \frac{1}{1+\epsilon_0} \), we need \( \eta \) to be small, if we want \( \psi_1 \) to be small.

To see how diminishing marginal utility works, suppose again that \( Y_i = N_i \) but alter the utility function so that utility depends separately on the supply of labor to each
firm rather than on the total quantity of labor supplied

\[ U(C, N) = \frac{C^{1-\theta} - 1}{1-\theta} - \frac{1}{1+\gamma} \sum N_i^{1+\gamma}. \]  \hspace{1cm} (16)

With this amendment, the first-order condition for labor becomes

\[ \frac{W}{P} = C^\theta N_i^{\gamma} = Y^\theta Y_i^{\gamma}. \]  \hspace{1cm} (17)

In this example, \( \omega = \gamma \). If we want \( \psi_1 \) to be small, we need \( \gamma \) to be large.

Note that \( \omega \) also affects \( \psi_2 \). In most macro-economic models, however, \( \varepsilon \) is a number larger than four and \( \sigma \) is a number greater than one. So, \( \omega \) has a greater impact on \( \psi_1 \).

It may seem counterintuitive that diminishing returns promotes price rigidity. It seems more natural that firms might want to respond to rapidly rising costs by raising their prices, and that this desire might promote flexibility. It is, therefore, useful to go through the intuition behind this form of strategic complementarity. Consider a firm that is considering an increase in the price of its output. The price increase will lead to a reduction in demand and, hence, production. Because of the diminishing returns, this reduction in production will reduce costs. The reduced costs in turn reduce the firm’s desire to raise its price. The more responsive costs, the less the firm raises its price.

A large elasticity of demand \( \varepsilon \) has effects very similar to increases in \( \omega \). A high elasticity of demand means that when a firm contemplates increasing its price, demand falls more and costs fall more. A high elasticity, therefore, mitigates price increases.

One of our themes was that the role played by a modeling element depends on exactly how that element is introduced. We see this with the inverse Frisch elasticity of labor supply \( \gamma \). When utility depended on the aggregate labor supply, as in (14), low values of \( \gamma \) were associated with low values of \( \psi \). When utility depended on the supply of labor in each market, as in (16), high values of \( \gamma \) were associated with low values of \( \psi \).

This idea generalizes. Pretty much any modification of the model can produce inertia if introduced in the right way. The general rule is that decreasing returns reduce \( \psi \), if introduced at the firm level, whereas increasing returns reduce \( \psi \), if introduced at the level of the economy as a whole. This is why it is important not only to find evidence for positive and negative spillovers but also to determine exactly how these mechanisms operate.

*How big is \( \psi \)?* There are at least two reasons that it is difficult to put a number on \( \psi \). First, as should be abundantly clear from the previous discussion, \( \psi \) is very sensitive to the precise structure of the economy. Second, \( \psi \) depends in some way on every controversial parameter in macroeconomics: the elasticity of intertemporal substitution, the elasticity of labor supply, and the elasticity of demand.

The standard Real Business Cycle model provides a natural benchmark and a useful place to start. Real Business Cycle models typically assume common economy-wide
factor markets, so that marginal cost is equalized across firms. If the firms are small, \( \omega = 0 \) and \( \psi = \sigma \). In the example given by equation (15), we get \( \psi = \theta + \gamma \). Since the Frisch elasticity is positive, we know that \( \gamma \geq 0 \). In addition, we assume log utility \( \theta = 1 \), then we conclude that \( \psi \geq 1 \). Increasing risk aversion or introducing capital into the production function only serve to increase \( \psi \). This parameterization does not generate large or persistent effects of money on output and points to the need to include additional modeling elements.

On the other extreme, if we assume that labor markets are segmented as in the example given by equation (17), then \( \psi = \frac{\gamma + \sigma}{1 + \epsilon} \). Now if \( \gamma = \sigma = 1 \), and \( \epsilon = 10 \), conclude that \( \psi = 0.2 \).

Note that this last example is very dependent on the choice of the elasticity of demand \( \epsilon \). This will be the case in any theory in which marginal cost depends on a firm's output, so that \( \omega > 0 \). \( \epsilon \) translates the change in the firms price into the change in the firms output that then affects marginal cost. The elasticity of demand is, therefore, an important parameter. Unfortunately, the calibration of this parameter differs widely in the literature. Macro-economic estimates tend to be fairly large. Basu and Fernald (1997) conclude that markups are in the neighborhood of 10%, which implies \( \epsilon = 11 \). Micro-economic estimates tend to be much smaller. Broda and Weinstein (2006), for example, estimate elasticities in the neighborhood of four. This implies a markup of 25%. Barsky et al (2003) and Chevalier, Kashyap, and Rossi (2003) estimate markups in the neighborhood of 50%.^10

It is clear that more work is needed to pin down the value of \( \epsilon \). One calculation that tends to favor high elasticities and low markups is the relationship between markups, returns to scale, and profit shares (see Hall 1990, Basu and Fernald 1997)^11

\[
\text{returns to scale} = \mu(1 - s) \cdot
\]

Given that profits share \( s \) is less than 3% and that there is scant evidence for large increasing returns to scale, it is difficult to believe that markups are very large on average.

**Why does this matter?** One could argue that we do not care about the microfoundations of \( \lambda \) and \( \psi \). All that matters for aggregate dynamics is the product \( \lambda \psi \). We can solve and estimate our macro-economic models without delving into the details of their construction. Interested parties may back out a set of micro-economic parameters using whichever micro-economic structure they deem appropriate. This has been the approach followed in most recent research.

This view misses two important points. Solid microfoundations help us test and validate our models. It takes a long time series to estimate macro-economic models with

10. Barsky et al. (2003) pursue a novel strategy. They compare the prices of brand name to generic products under the assumption that marginal cost is the same.

11. Returns to scale are equal to the ratio of average costs to marginal costs. The markup is the ratio of price to marginal cost. The ratio of average cost to price is \( 1 - s \).
precision. Micro-economic evidence can help determine the appropriate functional forms and provide additional information in estimating parameters.

More importantly, even if the macro-economic dynamics of various microfounded models are similar, their welfare implications may differ. The welfare costs of inflation in New Keynesian models come from the distribution of prices and the resulting misallocation of resources. Both the amount of price dispersion and the costs of resource misallocation depend on the frequency of price adjustment and the sources of strategic complementarity. More frequent price adjustment leads directly to less price dispersion. The effects of complementarity are more subtle. Greater complementarities may reduce price dispersion, by reducing the incentives to change prices by large amounts. Greater complementarities, however, are often associated with greater curvature in utility or production, so that any given level of price dispersion is more costly. Levin, Lopez-Salido, and Yun (2007) show that these considerations may affect the optimal trade-off between output and inflation. Observationally equivalent models from the macro-economic perspective—but not necessarily the micro-economic perspective—may lead to different monetary policy prescriptions.

1.5 Inflation Inertia

A strange thing happened during the derivation of the New Keynesian Phillips curve. We began with a model of sticky prices, in which some firms are stuck with prices that they had chosen in the past, and we ended up with a model that is entirely forward looking. Inflation in equation (11) is a jump variable. It depends only on current output and expected future inflation. The past does not enter at all. There is no inflation inertia.

This characterization of inflation has had a dramatic impact on the way in which we view monetary policy. Issues such as credibility and commitment become much more important when you have a forward looking Phillips curve. Transparency and communication become tools for affecting expectations of future inflation and, hence, the management of inflation today (Woodford 2010).

A serious problem with this formulation of the Phillips curve, however, is that equation (11) does not appear to fit the well with the econometric evidence on the response of inflation to money supply shocks. A common feature of the data is that inflation responds to monetary policy shocks with a lag. The response of inflation to a monetary policy shock appears hump shaped: there is very little response for up to a year; inflation then rises and peaks after about 2 years (Christiano, Eichenbaum, and Evans 2005, Mankiw 2001). Hence, \( E_t \pi_{t+1} \) initially rises more than \( \pi_t \) when policy is loose. The problem is that output must then fall in order for equation (11) to hold. Output, however, tends to rise in response to a loosening of monetary policy. It is for this reason that estimates of (11) using measures of the output gap often get negative estimates of \( \lambda \psi \), and it is for this reason that such estimates have been considered evidence against the specification in equation (11) (Gali and Gertler 1999).

12. See Mankiw (2001) for a discussion of this point.
The typical response to this problem has been to assume that the firms that do not have the opportunity to choose a new price may index their prices to past inflation. Firms that do not alter prices in period $t$ may charge $p_{t,t-1} + \pi_{t-1}$. This alters both the first-order condition for price changers and the price index

$$p_t^* = (1 - \alpha\beta) E_t \sum_{k=0}^{\infty} \alpha^k \beta^k [p_{t+k} + mc_{t+k} - (p_{t+k-1} - p_{t-1})],$$

$$p_t = (1 - \alpha)p_t^* + \alpha(p_{t-1} + \pi_{t-1}).$$

The result is a Phillips curve in the change in inflation

$$\pi_t - \pi_{t-1} = \lambda \psi y_t + \beta (E_t \pi_{t+1} - \pi_t).$$

Now the change in inflation is a jump variable. Rising inflation is now perfectly consistent with an increase in output. There is also inflation inertia in (18). All else equal, inflation today is increasing in past inflation. Lagged inflation tends to enter significantly in estimates of New Keynesian Phillips curves that include lagged inflation (Gali and Gertler 1999).

The problem with this specification is that there is little evidence that firms index their prices to past inflation or to anything else. A firm that indexes its price, changes its price every period. Instead, the typical pattern of price adjustment that we observe in the data show multiple periods of constant prices and occasional price adjustments (Klenow and Malin 2010). How equation (18) can be made consistent with this evidence is not very clear. Indexing helps the model to better fit the macro-economic evidence but misses wildly on the micro-economic facts. Models that include indexing are microfounded in the sense that agents in these models solve well-formulated maximization problems, but they lack microfoundations in the sense that the term is usually used, namely that the problems agents solve in the model bear some resemblance to the problems agents face in the real world.

Another possibility is that the problem is more with equation (8), which relates marginal cost to output. Efforts to estimate the marginal cost Phillips curve (10) have generally met with greater success than efforts to estimate (11) (Gali and Gertler 1999, Sbordone 2002). Given the firm’s goal of pricing relative to marginal cost, this formulation would appear to be closer to the firm’s first-order condition. The reason that these marginal-cost-based Phillips curves perform better is that the measures of marginal cost that they use, labor’s share in particular, tend to be slightly countercyclical and tend to lag the cycle (Gali and Gertler 1999).

This raises the issue of the cyclicality of marginal cost, another area rich in theories but slim on answers. Rotemberg and Woodford (1999) survey the literature. They argue that while it is true that labor’s share is countercyclical and that the inverse of labor’s share would be the correct measure of marginal cost if the production function were Cobb–Douglas, there are a number of reasons to believe that marginal cost may be more procyclical than labor’s share and may in fact be procyclical. Overhead labor, overtime, fixed costs of production, and factor-adjustment costs all imply that
labor’s share understates marginal cost in recessions and overstates marginal costs in expansions. Understanding the evolution of marginal cost is clearly important for understanding the New Keynesian Phillips curve.

One avenue that has not been explored is the role of trend inflation. The New Keynesian Phillips curve in equation (11) is a linearization about a steady state with zero inflation. Trend inflation is not zero in the US economy, especially during the 1970s and early 1980s. Cogley and Sbordone (2008) estimate a New Keynesian Phillips curve with nonzero trend inflation. Nonzero trend inflation tends to add a large number of terms to the linearization since nominal marginal cost is no longer constant in steady state. Cogley and Sbordone argue that these terms are relatively small and that equation (11) still does a reasonable job of describing the dynamics about trend. They find that once a nonzero trend is included in the model, lags of inflation no longer help to explain inflation. Inflation inertia disappears. It is possible movements in trend inflation could also explain the delayed response of inflation to monetary shocks. If, for example, a monetary shock led to rising trend inflation, then rising inflation would be consistent with an increase in output. This is an area for future research.

2. VARIABLE MARKUPS

One way to view New Keynesian economics is that sticky prices are a mechanism for generating countercyclical markups, and that countercyclical markups are an important force generating comovement between consumption and labor supply. In a competitive model, it is difficult to generate such comovement without productivity shocks (Barro and King 1984). An increase in consumption lowers the marginal utility of consumption and hence the utility value of a given wage. The result is a negative correlation between consumption and labor, unless the shock generating the increase in consumption is a shock to the productivity of labor. With imperfect competition, changes in the markup provide some room for wages to adjust. A shock that increases consumption may also cause the markup to fall. This allows wages to rise, providing an incentive for labor supply to increase.

This raises two questions. How do sticky prices interact with other theories of cyclical markups? What is the evidence for this central mechanism?

2.1 Markups as a Source of Complementarity

Countercyclical markups work very much like strategic complementarity $\psi$. Consider again the maximization problem of the firm, but without imposing the particular functional form for demand

$$\max_{p_i} E_t \sum_{k=0}^{\infty} \alpha^k \Lambda_{t,t+k} \left[ D \left( \frac{P_t}{P_{t+k}}, Y_{t+k} \right) \frac{P_t}{P_{t+k}} - C \left( D \left( \frac{P_t}{P_{t+k}}, Y_{t+k} \right), Y_{t+k} \right) \right].$$
The first-order condition is
\[
E_t \sum_{k=0}^{\infty} \alpha k \Lambda_{t,t+k} \frac{D_1 \left( \frac{P_t}{P_{t+k}}, Y_{t+k} \right)}{P_{t+k}} \times \left[ \left( 1 - \frac{1}{\epsilon \left( \frac{P_t}{P_{t+k}}, Y_{t+k} \right)} \right) \frac{P_t}{P_{t+k}} - C_1 \left( D \left( \frac{P_t}{P_{t+k}}, Y_{t+k} \right), Y_{t+k} \right) \right] = 0,
\]
where \( \epsilon \left( \frac{P_t}{P_{t+k}}, Y_{t+k} \right) \) is the elasticity of demand with respect its first argument. Linearizing the first-order condition, we arrive at
\[
\sum_{k=0}^{\infty} \alpha^k \beta^k [p_t^* - \mu_{t+k} - p_{t+k} - mc_{t,t+k}] = 0,
\]
where \( \mu_{t+k} \) is the log deviation of the markup \( \epsilon/(\epsilon - 1) \) from its steady state value. Recall \( mc_{t,t+k} \) is the marginal cost at date \( t + k \) of a firm choosing \( p_t^* \) at date \( t \).

A simple formulation is due to Kimball (1995), in which \( \epsilon \) is a function solely of \( \frac{P_t}{P} \). Let \( \xi \) denote the elasticity of the desired markup with respect to the relative price. Then, the first-order condition becomes
\[
\sum_{k=0}^{\infty} \alpha^k \beta^k [(1 - \xi)(p_t^* - p_{t+k}) - mc_{t,t+k}] = 0.
\]

Dividing through by \( 1 - \xi \), we see that the effect of variable markups is to multiply marginal cost by \( 1/(1 - \xi) \). Hence, this formulation leads to a New Keynesian Phillips curve with a coefficient on output cost equal to \( \lambda \psi / (1 - \xi) \). If \( \xi \) is negative, which happens if raising \( P_t \) relative to \( P \) makes demand more elastic and reduces the desired markup, then variable markups work exactly the same way as do strategic complementarities that operate at the level of the firm: the opportunity to raise one’s price leads to a reduction in the desired markup, which reduces the extent of the price increase.

Variable markups involve the same subtleties as do strategic complementarities. As we saw above, complementarities have different effects depending on whether they operate on the level of the firm or on the level of the economy as a whole. The same is true for markups. In the Kimball model above, one wants the markup to fall as a firm’s price rises above the aggregate price level. If firms are identical except for the prices that they charge, this is the same as having the desired markup fall as a firm’s output falls below average output. Therefore, markups need to be positively correlated with the firm’s relative output.

At the aggregate level, however, we need the desired markup to be negatively correlated with aggregate output in order to amplify and propagate nominal disturbances.
To see this, assume that $\mu$ is a function only of $y$ and that $\eta$ is the elasticity of the markup with respect to output, so that $\mu = \eta y$. Then, combining equations (3), (7), and (8) yields

$$\pi_t = \lambda mc + \eta y + \beta E_t \pi_{t+1}$$

$$= (\lambda \psi + \eta) y + \beta E_t \pi_{t+1}.$$ 

If $\eta$ is negative, so that an increase in output is associated with a reduction in the desired markup, then the coefficient on output in the New Keynesian Phillips curve is lower, and nominal shocks have larger, more persistent effects on output.

There are several models that generate negative values of $\eta$. Some generate countercyclical desired markups through cyclical variations in the composition of demand. In Bils (1989) and Gali (1994), the idea is that buyers of durable goods and investment goods are more price sensitive than buyers of nondurable goods, perhaps because these are big ticket items so that buyers spend more time searching for the best deals. Since purchases of investment and durable goods are very procyclical, these goods make up a larger share of aggregate demand during booms. In this way, demand for the typical good becomes more elastic during booms, and desired markups fall. In Edmonds and Veldkamp (2009), countercyclical variation in the dispersion of income leads to countercyclical variation in markups. Phelps and Winter (1970) construct a model in which there is both an extensive and intensive margin to a firm’s demand. Lowering the price not only makes a firm’s current customers purchase more, it attracts new customers. Customers are a stock variable, and the elasticity of demand is greater in the long run than in the short run. Markups in this model depend on the ratio of current to future demand. If demand is rising, then the firm may find it in its interest to lower markups and build a customer base. Markups should be low during booms and high during busts.

Another class of model looks to competition among firms. In the implicit collusion model of Rotemberg and Solon (1986), high prices are supported by the threat of price wars. These threats are more credible when future profits are high relative to current profits. Given the procyclicality of profits, this theory predicts that markups are low during booms. Bilbiie, Ghironi, and Melitz (2007) construct a model in which procyclical profits lead to procyclical entry. Entry increases competition and reduces markups during booms.

### 2.2 Evidence on Cyclical Markups

There appears to be a great deal of evidence that business cycles are associated with significant movements in the gap between the marginal product of labor and the marginal rate of substitution between leisure and consumption (Chari, Kehoe, and McGrattan 2007, Gali, Gertler, and Lopez-Salido 2007). There is less agreement on how much of this gap is due to gap between the marginal product of labor and marginal cost (i.e., the markup) and how much is due to the gap between marginal cost and the marginal rate of substitution between leisure and consumption. In the
former case, we should look to the behavior of firms for the source of fluctuations in this gap. In the latter case, we should look to the behavior of labor markets. The main problem is none of the variables in this calculation are observable. They all must be inferred through some model.

If one assumes a Cobb–Douglas production function, then the markup is inversely related to labor’s share. As labor’s share is slightly countercyclical, one would conclude that markups are slightly procyclical and that the bulk of the above mentioned gap is due to labor markets. Rotemberg and Woodford (1999) provide a large number of arguments why this calculation may be biased and markups might in fact be countercyclical. These arguments include overhead labor, fixed costs, overtime, adjustment costs, and labor hoarding. Bils (1987) finds that markups are countercyclical in a model with overtime labor. Recently, however, Nekarda and Ramey (2009) find that markups are procyclical in a very similar model but with a richer data set. Clearly there is more work to be done.

3. STATE DEPENDENCE

Up to this point, I have assumed that the probability of adjustment was constant. The main alternative is the Ss model, in which there are fixed costs of price adjustment and firms balance the costs and benefits of price adjustment. Such models are often referred to as state-dependent models since the timing of price adjustment depends on the state in which the firm finds itself. Early models such as Caplin and Leahy (1991, 1997) and Caballero and Engel (1991) held out promise that endogenizing the timing of price adjustment would lead to aggregate dynamics that were significantly different from time-dependent pricing models. These models show the potential for significant nonlinearities in aggregate dynamics. In these models, the response of inflation to a shock depends on the entire distribution of prices in the economy; the more firms are near their price adjustment triggers, the more prices respond to shocks.

Recent efforts, however, show more modest differences between Ss models and the Calvo model. The main reason is that idiosyncratic shocks tend to be much larger than aggregate shocks. Hence, most adjustment in these models is triggered by shocks that are idiosyncratic to the firm. While the timing of adjustment may be endogenous to the firm’s state, it is effectively exogenous to the aggregate state of the economy.

This does not mean that state dependence has no effect. There are at least three important practical implications of state dependence. First, the firms that adjust their prices are not a random subset of the firms in the economy as they are in the Calvo model. The firms that adjust their prices at any given point in time tend to be those who most want to adjust. This “selection effect” tends to make the economy look more flexible than one would assume, given the observed frequency of price adjustment (Golosov and Lucas 2007). The reason is that the firms that choose not to adjust their prices choose not to adjust because their prices are already near their optimum.

This observation has two consequences. First, if one is parameterizing a Calvo model, one should choose a probability of price adjustment that is greater than the
one that is observed in the data. Second, in order to match the persistence of real
variable to nominal shocks, one needs stronger strategic complementarities.\textsuperscript{13}

The second practical implication is that state dependence may in some cases affect
the way in which future marginal costs are discounted in the first-order condition (5).
With Calvo pricing future marginal costs are given exponentially declining weights.
This can lead to significant frontloading of price changes in cases in which marginal
costs are expected to rise or fall in the future. This frontloading of future movements
in marginal costs lies at the heart of the model’s inability to explain hump-shaped
impulse responses described above: if inflation is expected to rise in the future, it
must be because marginal cost is expected to rise in the future, which means prices
must rise today unless marginal cost is very low today. Frontloading also lies at the
heart of some puzzling results regarding the effects of credible disinflations. Ball
(1994) shows that a credible commitment to future deflation is expansionary: since
prices are expected to fall tomorrow, they must fall today, which reduces markups
and hence raises output today.

There is much less need for frontloading in Ss models. The reason is simple: firms
can change their prices whenever they want to. If there is an expected future change
in marginal cost, firms can wait and change their prices when costs change or they can
change their prices today and change them again in the future. Dotsey and King (2005)
show that this feature of Ss models can help to explain why the impulse response
function of inflation is hump-shaped in response to a nominal shock: if strategic
complementarities are strong, firms may choose to delay their price adjustment until
other firms are ready to adjust as well, thereby delaying the response of inflation.
Midrigan (2006) shows that this feature of Ss models can reduce the responsiveness
of prices to a future increase in the money supply: in the Calvo model, a firm has
to respond to a future increase in the money supply when it can, whereas in the Ss
model, the firm can delay adjustment until it is warranted.

The third practical implication of state dependence is that it alters the effects
of certain strategic complementarities. With Calvo pricing, both complementarities
that work through relative price $\psi_1$ and complementarities that work through ag-
gregate supply $\psi_2$, amplify and extend the real effects of nominal shocks. With
state-dependent pricing, complementarities that work through relative price have am-
biguous effects. Reductions in $\psi_1$ reduce the size of price changes as in the Calvo
model, but they also increase a firm’s incentive to change its price and, thereby,
increase the frequency of price adjustment. Complementarities that work through
relative prices tend to make costs more sensitive to relative price, and this makes
the profit function more concave in relative price. Since Ss firms compare the costs
and benefits of price adjustment, this increased concavity raises the benefits of price
adjustment and leads more firms to adjust. This is like an increase in $\alpha$. Which ef-
fect dominates, the dampening of price changes or the increase in the adjustment
frequency, depends on the parameterization of the model. Dotsey and King (2005)

\textsuperscript{13} Gertler and Leahy (2008) show how to build an Ss model with strategic complementarities.
present examples in which low values of $\psi_1$ lead to greater price flexibility rather than greater rigidity.

The incorporation of state-dependent pricing in macro-economic models remains an active area of research.

4. CONCLUSION

I have attempted to draw connections between the recent literature on money and output and topics of interest to industrial organization.

The following are the main themes:

(i) Price rigidity is a pervasive phenomenon that should not be ignored.
(ii) The cyclicality of the markup and marginal cost remain important topics for study.
(iii) We need to understand better how prices of various firms interact. What are the sources of strategic complementarity that reign in idiosyncratic price adjustment? How are costs passed through to prices? How and when are aggregate shocks passed through to prices?
(iv) We need general studies that attempt to characterize the economy as a whole rather than studies of particular, and possibly idiosyncratic, industries or firms.

LITERATURE CITED


