NBER WORKING PAPER SERIES

INVENTORIES AND STICKY PRICES: MORE ON THE MICROFOUNDATIONS OF MACROECONOMICS

Alan S. Blinder

Working Paper No. 620

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge MA 02138

January 1981

I am indebted to Yakov Amihud, Olivier Blanchard, Stanley Fischer, Wolfgang Franz, Arnold Kling, Julio Rotemberg, Robert Shiller, and Lawrence Summers for useful comments on an earlier draft; to Danny Quah for research assistance; and to the National Science Foundation for financial support. The research reported here is part of the NBER's research program in Economic Fluctuations. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

NBER Working Paper #620
January 1981

Inventories and Sticky Prices:
More on the Microfoundations of Macroeconomics

ABSTRACT

The role of inventories in making prices "sticky" is studied by analyzing a dynamic linear-quadratic model of a monopoly firm facing stochastic demand, but able to store its finished goods in inventory. It is shown that, in contrast to the usual presumption, firms that exhibit the smallest output responses to demand fluctuations may also exhibit the smallest price fluctuations. Specifically, firms which have very flexible inventory storage facilities or are subjected to very transitory demand shocks will rely on inventories as buffers, and will change neither production nor price very much. On the other hand, firms which have very inflexible storage facilities or whose demand shocks are quite permanent will display large swings in both price and output.

The standard assumption about inventory carrying costs that has been used in the literature (that they are linear) is shown to imply that production is impervious to fluctuations in demand. It is also established that prices may respond more strongly to positive demand shocks than to negative ones if it is impossible to hold negative inventories (i.e., to have unfilled orders).

The model offers an explanation for "stickiness" in relative prices. However, under certain circumstances, it may help explain the persistence of inflation.

Alan S. Blinder
National Bureau of Economic Research
1050 Massachusetts Avenue
Cambridge, Massachusetts 02138

(617) 868-3929

1. Introduction

The phenomenon of "sticky prices," that is, the apparent insensitivity of prices to fluctuations in demand, has long intrigued both microeconomists and macroeconomists.[1] Normally, increasing production and raising prices are thought of as alternative ways for a firm to respond to an increase in the demand for its product. Thus firms that opt for large quantity adjustments will display small price adjustments, while firms that make small quantity adjustments will be forced to make large price adjustments.

The logic behind this common conclusion is clear from Figure 1. Here y is output, $C'(y)$ is the (rising) marginal cost curve, $R'(y) + \varepsilon$ is the (falling) MR curve (with ε a random demand shock), and equilibrium is where MR = MC. It follows quite clearly that, among firms facing the same structure of demand, those with steep MC schedules will display strong price responses and weak output responses, while those with flat MC schedules will display weak price responses and strong output responses.

The main point of this paper is that this simple conclusion may well be reversed when output is storable. Specifically, it may be that some firms exhibit large output <u>and</u> large price responses to demand shocks while other firms respond rather little in either dimension. The central result is a theorem characterizing precisely what types of firms tend to fall in each category. Loosely speaking, the principal conclusion of the paper (stated more precisely as Theorem 1 below) is that <u>both</u> price <u>and</u> output responses become smaller as demand shocks become less persistent and output becomes more "inventoriable." That is, for given MC and MR curves, "sticky" prices will tend to emerge when it is not very costly to vary inventories and when demand shocks are very transitory.

---

[1] For a recent comprehensive survey from a macroeconomic perspective, see Gordon (forthcoming).

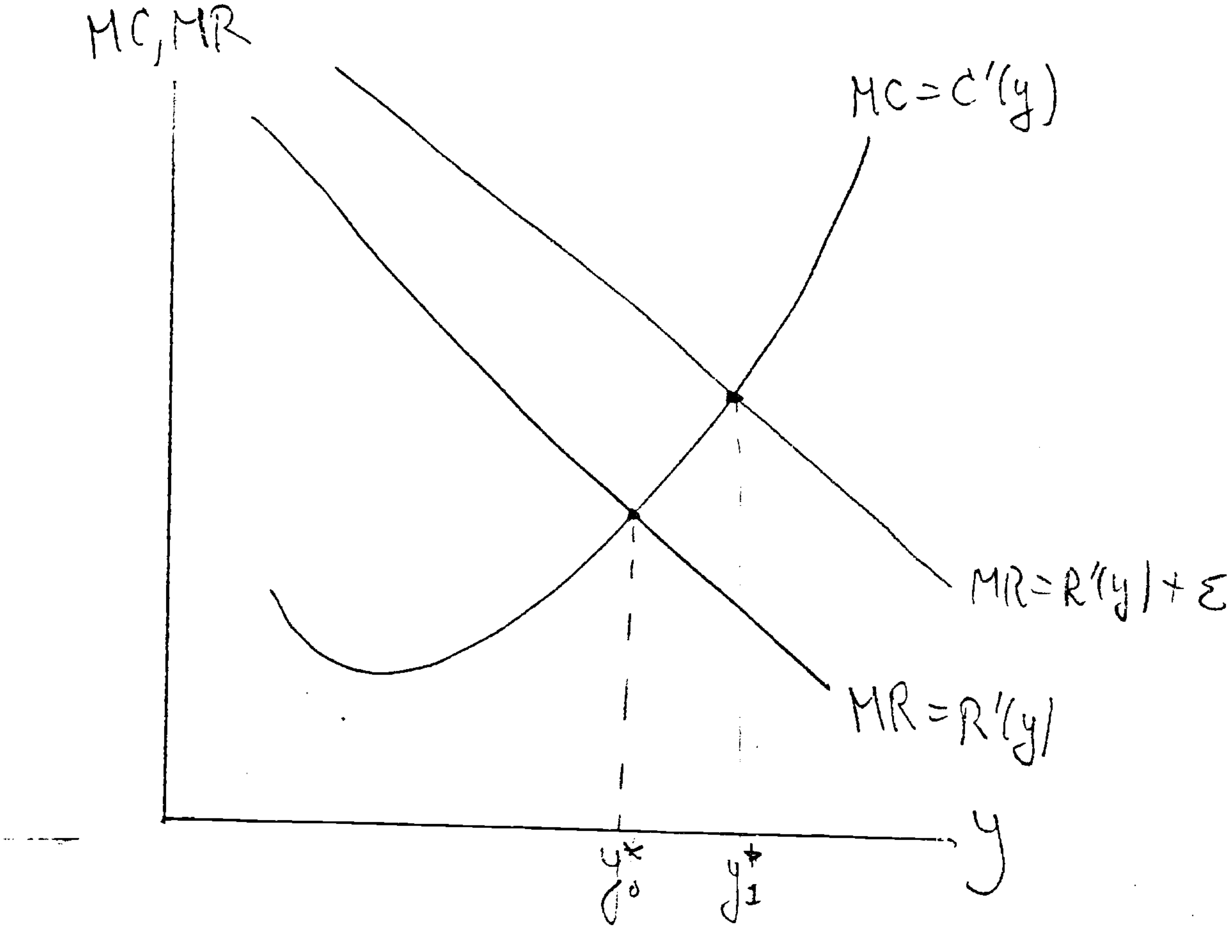

Figure 1

The paper is closely related to, and generalizes the results of, recent papers by Amihud and Mendelson (1980) and Reagan (1980), and a well-known earlier paper by Zabel (1972).[1] The two recent papers emphasize an asymmetry in pricing behavior that results when stockouts occur by showing that prices respond more strongly to demand shocks when there are stockouts than when there are not. This asymmetry, while it does occur in my model as well, is not the focus of this paper. Rather, for the most part I ignore stockouts and stress that the existence of inventories reduces the flexibility of prices in all states of demand and in both directions.[2]

The paper is organized as follows. This section concludes with an intuitive explanation of the main result, which both conveys the flavor of the proof and, more importantly, suggests that the result is quite a bit more general than the model used to derive it. Section 2 presents the formal model, compares it to the earlier literature, and states the main theorem precisely. Section 3 consists of a formal proof, and Section 4 collects some interesting related results on inventory behavior that flow from the model. Section 5 discusses some macroeconomic phenomena for which the model may provide microfoundations, and Section 6 is a brief summary.

Why does the existence of inventories of finished goods invalidate the simple story told by Figure 1? The answer is that a firm with storable output is operating simultaneously on two margins. It must decide how much output to produce for inventory, and it must decide how much inventory to sell.

---

[1] For other related work, see Maccini (1976), Blinder and Fischer (1981), Phlips (1980), and Reagan and Weitzman (1980).

[2] Such a result is hinted at by Zabel (1972, p. 535).

While profit maximization continues to require MC=MR each period, it is no longer necessary that this be done with output (y_t) equal to sales (x_t). Indeed, it will quite often be the case that optimal y_t and x_t differ, with changes in the stock of inventories (N_t) taking up the slack. Figure 2 illustrates the basic idea. The MC curve is $C'(y)$, and optimal output is determined by equating MC to the shadow value of inventories, denoted $\lambda(N)$, at point B. The MR curve is $R'(x) + \varepsilon$, and optimal sales are determined by equating MR to $\lambda(N)$ at point A. It is obvious that x and y need not be equal.

Now consider what happens when ε changes. Trivial algebra establishes that:

$$\frac{dy}{d\varepsilon} = \frac{1}{C''(y)} \frac{d\lambda}{d\varepsilon}$$

$$\frac{dx}{d\varepsilon} = \frac{1}{-R''(x)} \left(1 - \frac{d\lambda}{d\varepsilon}\right)$$

so it is clear that everything hinges on the response of λ to ε. If $\frac{d\lambda}{d\varepsilon}$ is close to unity, demand shocks will be met by substantial changes in output but small changes in sales, i.e., large changes in prices. Conversely, if $\frac{d\lambda}{d\varepsilon}$ is close to zero, demand shocks will elicit small output responses and large sales responses, i.e., small price responses. Large (small) output responses and large (small) price responses thus go together if the cross-sectional differences among firms come mainly in $\frac{d\lambda}{d\varepsilon}$ (rather than in $C''(y)$ and $R''(x)$).

What factors are likely to govern the size of $\frac{d\lambda}{d\varepsilon}$? First, intuition tells us that the shadow value of inventories will respond more strongly when shocks are expected to persist longer. A very transitory shock will change λ very little; a permanent one will change it a lot. Second, it seems likely that $\frac{d\lambda}{d\varepsilon}$ will be larger when goods are more difficult to inventory. Finally, Figure 2 suggests that the slopes of the MC and MR curves remain relevant, though not in as obvious a way as in Figure 1. The essence

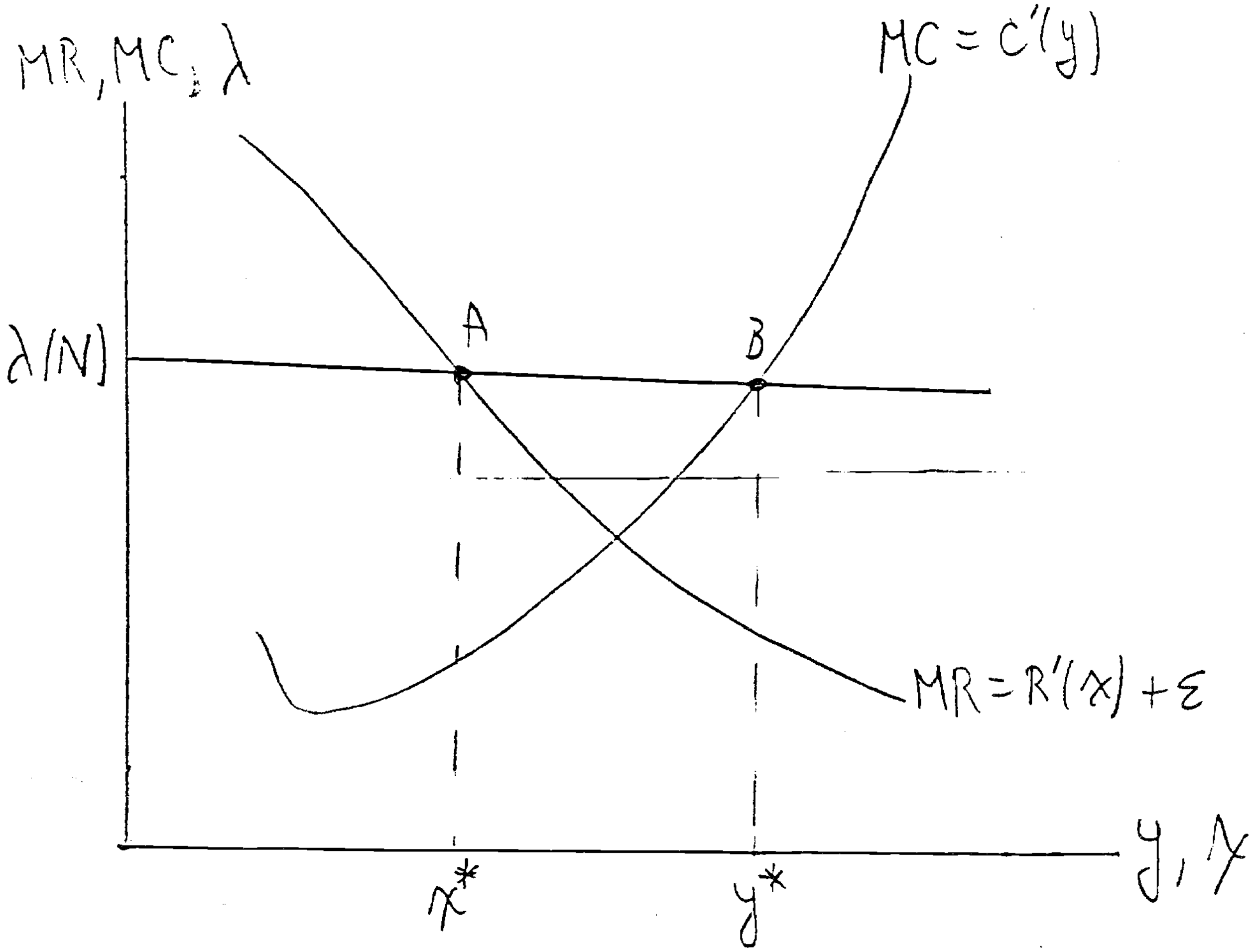

Figure 2

of Theorem 1 below is to show that $\frac{d\lambda}{d\varepsilon}$ is between zero and unity, and depends on the aforementioned parameters in the ways suggested by intuition.

2. The Model and Formal Statement of the Theorem

The firm, which I take to be a monopolist, chooses time paths for sales, x_t, and output, y_t, to maximize the expected discounted present value of profits. There is a random demand shock, η_t, each period which the firm sees <u>before</u> setting sales and output for that period. While the analysis could be conducted equally well in either discrete or continuous time, I choose the former.[1] Hence, the firm wishes to maximize:

$$E_0 \sum_{t=0}^{\infty} D^t \{ \tilde{R}(x_t, \eta_t) - C(y_t) - B(n_t)\}$$

subject to the constraints:

$$(2.1) \quad n_{t+1} - n_t = y_t - x_t \qquad t = 1, 2, \ldots$$

Here $\tilde{R}(\cdot)$ is the revenue function, $C(\cdot)$ is the production cost function, $B(\cdot)$ is the inventory carrying cost function, $D \equiv \frac{1}{(1+r)}$ is a discount factor, and the expectation is taken as of time zero. Each of the functions merits some discussion.

Revenue Function

The work of Zabel (1972) makes it quite clear that there is almost no hope of deriving results unless the random shock to the demand curve is additive. Hence I assume that price, p_t, is given by:

$$p_t = \pi(x_t) + \eta_t, \qquad \pi'(\cdot) < 0,$$

1 This is purely a matter of taste. The earlier version of this paper used continuous time.

where η_t is the random demand shock. Then define the revenue function as:

$$\tilde{R}(x_t, \eta_t) = x_t p_t = x_t \pi(x_t) + x_t \eta_t \equiv R(x_t) + x_t \eta_t,$$

where $R(\cdot)$, the deterministic part of the revenue function, is assumed to be strictly concave. Where explicit solutions are required, I assume a linear demand function:

$$(2.2) \quad \pi(x_t) = \frac{1}{\pi}(\pi_0 - \frac{1}{2} x_t) \qquad \pi_0, \quad \pi > 0,$$

so that $R(\cdot)$ is quadratic. π_0 is assumed to be so large that $\frac{\pi_0}{\pi} + \eta_t$ (which is marginal revenue at zero sales) is always positive and the constraint $x_t \geq 0$ is never binding. While this is a less general specification than that used by Zabel (1972), Amihud and Mendelson (1980), and Reagan (1980), its only role is to permit an explicit solution. All the results to follow hold approximately in the neighborhood of equilibrium for an arbitrary concave revenue function.

Production Cost Function

For the same reason, the production cost function is assumed to be quadratic:

$$(2.3) \quad C(y_t) = c_0 + \frac{1}{c}(c_1 y_t + \frac{1}{2} y_t^2)$$

where c_1/c (i.e., marginal cost at zero output) is assumed small enough so the constraint $y \geq 0$ is never binding. While this specification is less general than Zabel (1972), who worked with an arbitrary convex cost function, it is more general than Amihud-Mendelson (1980) and Reagan (1980), who restricted themselves to the case of constant marginal costs.

Inventory Cost Function

Here again, where explicit solutions are necessary, I assume that $B(\cdot)$ is quadratic:

$$(2.4) \quad B(n_t) = b_0 + b_1 n_t + (b/2)n_t^2, \qquad b > 0$$

where $n_t \equiv N_t - K$ is the deviation of inventories from some critical level, K. This is a more general specification than Zabel (1972) or Amihud-Mendelson (1980), both of which assumed a linear storage cost technology (b=0). Reagan (1980) assumed no explicit storage costs. It turns out that the parameter b is absolutely critical to the analysis and that b=0 is a very special case.

This cost function, which is sketched in Figure 3 for the case $b_1 = 0$ admits of two possible interpretations. First, K could be zero, in which case B(N) is storage costs if inventories are positive and represents the cost of holding a queue of unfilled orders if N is negative.[1] Alternatively, K could be some critical level of inventories below which production costs actually rise because it is difficult to schedule production efficiently, etc. The magnitude of K would then obviously influence the likelihood that the firm stocks out. For the most part, I will assume that negative values of N (interpreted as unfilled orders) are possible. But Section 5 will offer some brief remarks on what happens if negative N is impossible. Under either interpretation, it makes sense to assume $b_1 = 0$ since B(n) reaches its minimum at $n = -b_1/b$. But the parameter b_1 is not critical in what follows.

---

[1] The way the maximand is structured assumes that sales made at time t earn the firm p_t even though they are not delivered until some future date, that is, it assumes that unfilled orders are prepaid. If, in fact, unfilled orders are paid only upon delivery, then x_t yields the firm only $p_{t+s}(1+r)^{-s}$ where s is the delivery lag. Thus the cost to the firm of backordering the sale is: $p_t - p_{t+s}(1+r)^{-s} = p_t[1-(\frac{1+g}{1+r})^s]$ where $\frac{p_{t+s}}{p_t} = (1+g)^s$ defines g. For small g and r, this is approximately $-p_t(g-r)s$. If s is zero for positive N_t and rises in an approximately quadratic manner for negative N_t, then a quadratic function $B(N_t)$ over negative values of N_t is derived. Symmetry of $B(N_t)$ around zero is not essential.

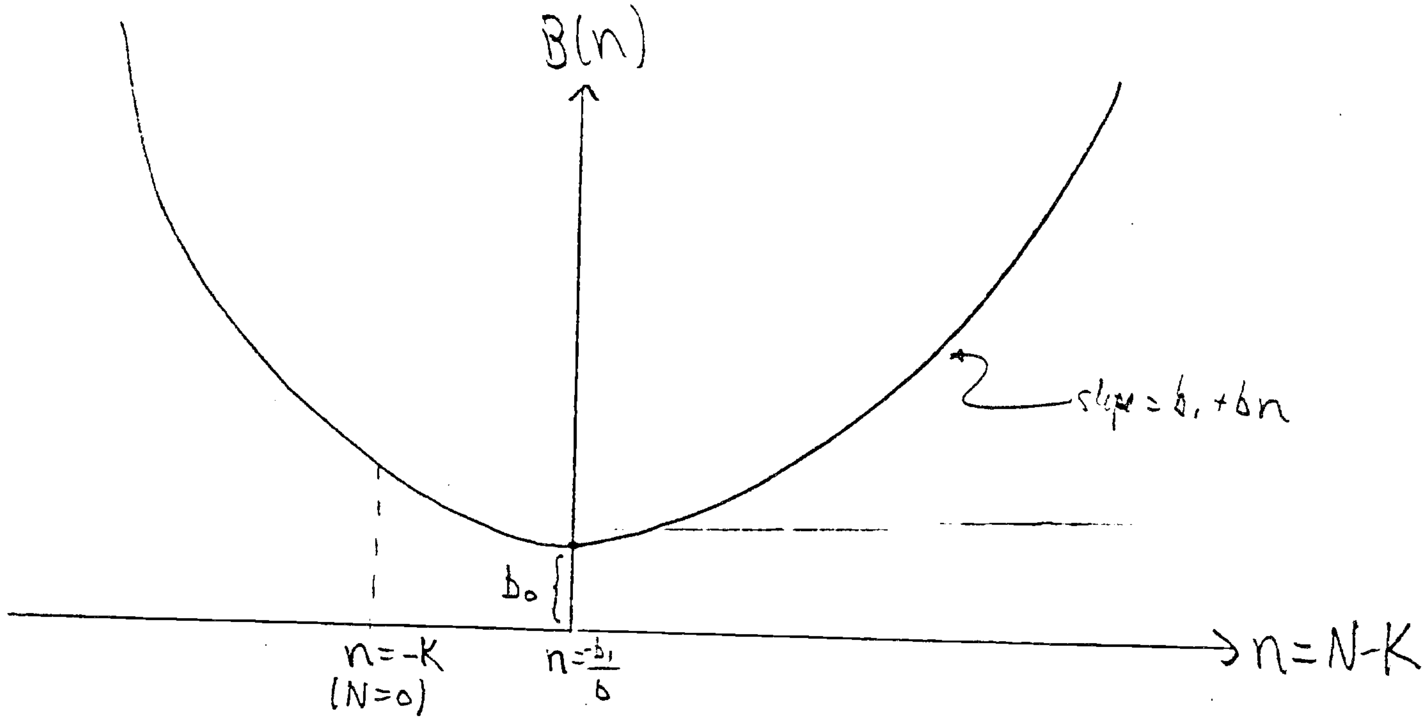

Figure 3

Distribution of Demand Shocks

Each of the papers referred to above assumes that demand shocks are independently and identically distributed (i.i.d.). This assumption, while it simplifies things greatly, is quite unsatisfactory. We know, for example, that disturbances at the macro level are highly serially correlated, and it would be surprising indeed if this serial correlation disappeared when we disaggregated to the industry of firm levels. In this paper, I assume that the stochastic structure of demand follows an AR(1) process:

$$(2.5) \quad \eta_t = \rho\eta_{t-1} + v_t, \quad 0 \le \rho \le 1$$

where v_t is a white noise disturbance term. Obviously, i.i.d. disturbances are a special case of (2.5) when $\rho = 0$. At the other extreme, $\rho = 1$ denotes a random walk.

I am now in a position to state the main theorem precisely.

THEOREM 1: The responses of the optimal price, p_t^*, and optimal production, y_t^*, to the contemporaneous demand shock, η_t, are both smaller when demand shocks are more transitory (i.e., ρ is smaller) and when marginal carrying costs of inventories are less sensitive to the level of inventories (i.e., when b is smaller).

I proceed now to the proof, whose basic idea was suggested by the intuitive analysis of Section 1. It will be clear that the linear quadratic structure of the problem is needed only to get an explicit solution and is unlikely to be "special" in any relevant sense.

Since the proof, while straightforward, is fairly long and involved, it will be useful to have a road map of where we are going. The proof proceeds in three stages:

1. The first-order conditions for the optimal control problem summarized

in Section 2 are derived.

2. These conditions are reduced to a second-order difference equation in a Lagrange multiplier which is interpreted as the shadow value of inventories. The homogeneous part of this difference equation is obtained.

3. The specific stochastic structure of demand is used to derive a particular solution, and the resulting difference equation is solved for its initial value. The results of step 1 are then used to derive the economic results of interest.

3.1 First-Order Conditions

Introducing the Lagrange multiplier q_t for constraint (2.1), the problem becomes to find a critical point of the Lagrangian:

$$E_0 \sum_{t=0}^{\infty} D^t \{R(x_t) + x_t \eta_t - C(y_t) - B(n_t) + q_t(-n_{t+1} + n_t + y_t - x_t)\}$$

for which the first-order conditions are:[1]

$$(3.1)\quad E_0 C'(y_t) = E_0 q_t \equiv \mu_t$$

$$(3.2)\quad E_0 R'(x_t) = E_0(q_t - \eta_t) \equiv \mu_t - \varepsilon_t$$

$$(3.3)\quad E_0 B'(n_t) = E_0(q_t - (1+r)q_{t-1}) = \mu_t - (1+r)\mu_{t-1}$$

$$(3.4)\quad n_t = n_{t-1} + y_{t-1} - x_{t-1},$$

where the following new symbols have been introduced to simplify the notation:

$$\mu_t = E_0 q_t$$

$$\varepsilon_t = E_0 \eta_t ;$$

that is, μ_t and ε_t are respectively the <u>actual values</u> of q_t and η_t for t=0, but are the <u>expectations</u> (based on information known at t=0) of these variables for t > 0. Consistent with this, the symbols y_t and x_t denote the output and sales for period t <u>as</u> planned at period 0. Hence only y_0 and x_0 correspond to

[1] The second order conditions obviously hold since $R''(x) < 0$, $C''(y) > 0$, and $B''(n) > 0$.

actual realized values.

Under the specific functional forms introduced in Section 2, (3.1) - (3.3) can be written as:

$$y_t = c\mu_t - c_1$$

$$x_t = -\pi(\mu_t - \varepsilon_t) + \pi_0$$

$$\mu_t = (1+r)\mu_{t-1} + bn_t + b_1.$$

These can be simplified somewhat by first computing the steady-state values of the nonstochastic part of the difference equation system, which are:

$$\bar{\mu} = \frac{\pi_0 + c_1}{\pi + c}$$

$$\bar{n} = \frac{r\bar{\mu} + b_1}{-b}$$

$$\bar{y} = c\bar{\mu} - c_1 = \bar{x} = -\pi\bar{\mu} + \pi_0 .$$

Then, if we let λ_t denote the deviation of μ_t from its steady state, our difference equation system can be written:

$$(3.5) \quad y_t = \bar{y} + c\lambda_t$$

$$(3.6) \quad x_t = \bar{x} - \pi\lambda_t + \pi\varepsilon_t$$

$$(3.7) \quad \lambda_t = (1+r)\lambda_{t-1} + b(n_t - \bar{n}) .$$

When (3.5) and (3.6) are substituted into the identity (3.4), equations (3.7) and (3.4) form a system of two first-order difference equations. The two stationaries of this system are given below:

$$0 = r\lambda_{t-1} + b(n_t - \bar{n})$$

$$0 = (c+\pi)\lambda_t - \pi\varepsilon_t,$$

and graphed in Figure 4, which clearly shows that for any fixed ε there is a unique equilibrium which is a saddle point. The firm must choose λ_0 to be on the unique stable arm, which immediately implies that λ_0 is a decreasing function of n_0.

3.2 Solving the Difference Equation System[1]

Since the firm computes a new optimal path each period, we are only interested in the solution for t=0. But to get this solution we must solve the entire system for its initial values. The easiest way to do this is to reduce it to a single second-order equation in λ_t. This is obtained by first-differencing (3.7) and using (3.5) and (3.6) to substitute out $n_t - n_{t-1}$. The result is:

$$(3.8) \quad \lambda_t - [2 + r + b(c+\pi)]\lambda_{t-1} + (1+r)\lambda_{t-2} = -b\pi\varepsilon_{t-1} \; .$$

Letting $\beta \equiv 2 + r + b(c+\pi)$, the characteristic roots of (3.8) are the solutions to the quadratic equation:

$$(3.9) \quad z^2 - \beta z + (1+r) = 0,$$

and the solution to the homogeneous part of (3.8) takes the form:

$$\lambda_t = K_1(z_1)^t + K_2(z_2)^t$$

where z_1, z_2 are the roots of (3.9) and K_1 and K_2 are constants to be determined by the initial conditions.

One of these constants follows immediately, however. By Descartes rule of signs, both roots of (3.9) are real and positive; furthermore, it is easy to show that one is less than unity while the other is greater than unity. Hence, if z_1 is the stable root and z_2 is the unstable root, picking λ_0 to be

---

[1] Readers whose tolerance for grinding is low and are willing to trust my derivation, can proceed immediately to equation (3.15), which is the desired solution.

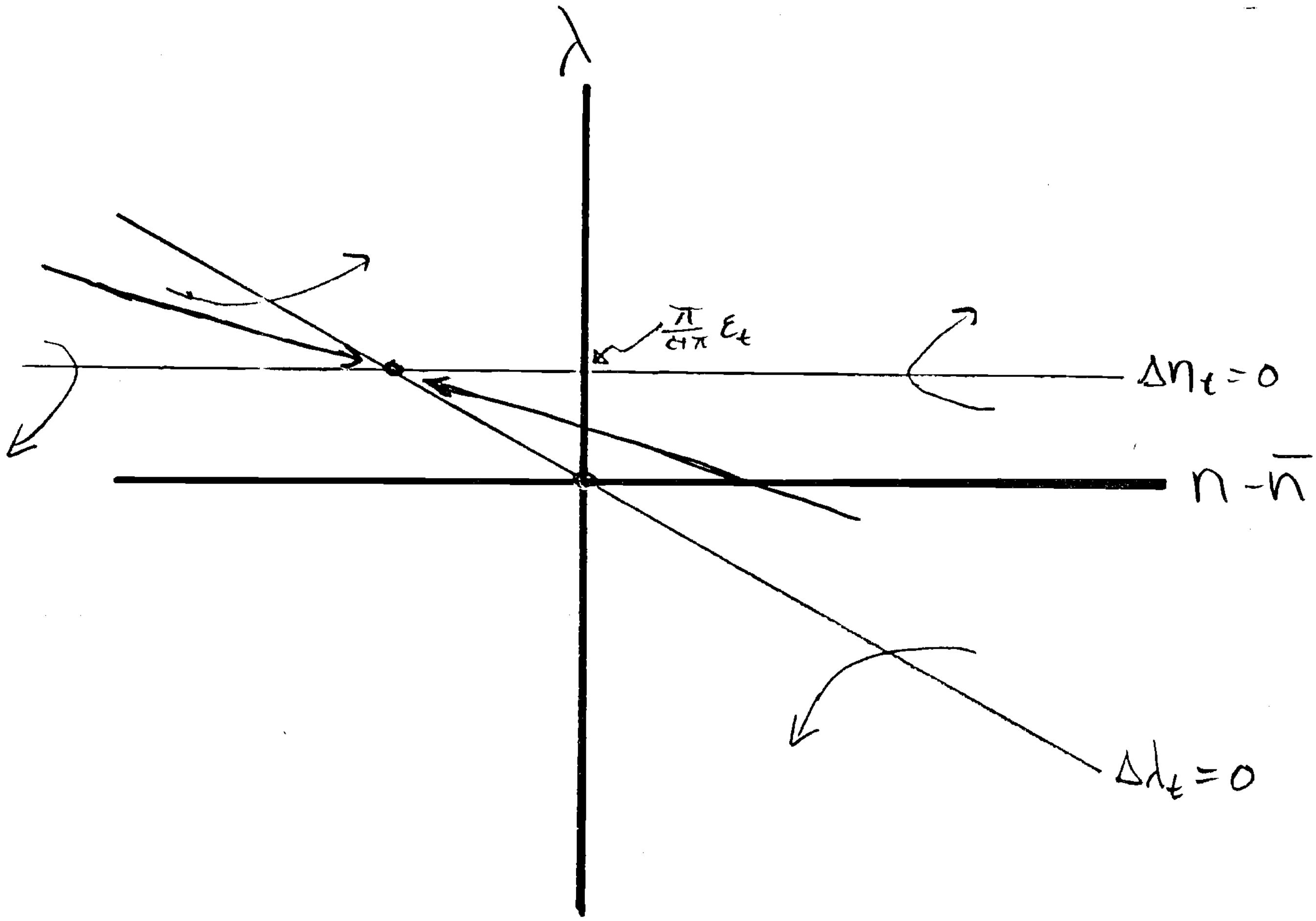
λ
$\frac{\pi}{1+\pi}\varepsilon_t$
$\Delta n_t = 0$
$n - \bar{n}$
$\Delta\lambda_t = 0$

Figure 4

on the stable arm in Figure 4 is tantamount to picking $K_2 = 0$. The full solution to (3.8) therefore takes the form:

$$(3.10) \quad \lambda_t = K_1 z_1^t + f_t$$

where f_t is a particular solution to the inhomogenous part of (3.8). That is, f_t must satisfy:

$$(3.11) \quad f_t - \beta f_{t-1} + (1+r) f_{t-2} = -b\pi\varepsilon_{t-1},$$

which makes it obvious that the form of f_t depends on the time structure of ε_t, which in turn depends on the nature of the stochastic process generating demand shocks.

One further result will prove useful in finding particular solutions. Setting t=1 in (3.7) gives:

$$(3.12) \quad \lambda_1 - \lambda_0 = r\lambda_0 + b(n_1 - \bar{n}) \quad .$$

By the usual identity:

$$n_1 - \bar{n} = (n_0 - \bar{n}) + (y_0 - x_0),$$

and substituting in from (3.5) and (3.6) leads to:

$$n_1 - \bar{n} = (n_0 - \bar{n}) + (c+\pi)\lambda_0 - \pi\varepsilon_0 .$$

Substituting this into (3.12) gives:

$$(3.13) \quad \lambda_1 - \lambda_0 = [r + b(c+\pi)]\lambda_0 + b(n_0 - \bar{n}) - b\pi\varepsilon_0 \quad .$$

3.3 Finding a Particular Solution

Under the AR(1) process described by (2.5), $\varepsilon_t = \rho^t \varepsilon_0$. (Recall that ε_t is the mean of the demand shock expected in period t.) The method is to try a solution of the form:

$$f_t = A\varepsilon_{t-1} = A\rho^{t-1}\varepsilon_0 \quad ,$$

and use equation (3.11) to determine the unknown coefficient A. Specifically, since A must satisfy:

$$A[\rho^{t-1}\varepsilon_0 - \beta\rho^{t-2}\varepsilon_0 + (1+r)\rho^{t-3}\varepsilon_0] = -b\pi\rho^{t-1}\varepsilon_0,$$

we obtain:

$$(3.14) \quad A = \frac{-b\pi\rho^2}{\rho^2 - \beta\rho + (1+r)} = \frac{-b\pi\rho^2}{Q(\rho)} \quad .$$

Notice that the quadratic form in (3.14), which I have abbreviated $Q(\rho)$, is identical to the quadratic in equation (3.9). Thus it factors into $(\rho - z_1)(\rho - z_2)$, a fact that will prove useful.

The last step in the derivation is to use this particular solution to solve for λ_0, the initial value of the shadow price of inventories. To do so, write (3.10) for t=0 and t=1:

$$\lambda_0 = K_1 + f_0$$

$$\lambda_1 = K_1 z_1 + f_1 = (\lambda_0 - f_0) z_1 + f_1$$

Subtracting yields:

$$\lambda_1 - \lambda_0 = \lambda_0(z_1 - 1) + f_1 - z_1 f_0 = \lambda_0(z_1 - 1) + A\varepsilon_0(1 - \frac{z_1}{\rho})$$

$$= \lambda_0(z_1 - 1) + \frac{b\pi\rho\varepsilon_0}{z_2 - \rho} \quad ,$$

where the factorization of $Q(\rho)$ in (3.14) has been used. Equating this to (3.13) gives a simple equation which can be solved for λ_0 to yield

$$(3.15) \quad \lambda_0 = b[1 + r - z_1 + b(c+\pi)]^{-1} \{\bar{n} - n_0 + \frac{\pi z_2}{z_2 - \rho} \varepsilon_0\}$$

Since I have earlier shown that y_0, x_0, and n_1 all depend on λ_0, everything we need to know about the economics of the problem is embedded in (3.15). I proceed now to ferret out the implications.

Proposition 1: In response to a positive (negative) shock to demand, the firm increases (decreases) output, sales, and price, and reduces (raises) inventory investment.

Proof: It follows from (3.4) - (3.6) that:

$$(3.16) \quad \frac{\partial y_0}{\partial \varepsilon_0} = c \frac{\partial \lambda_0}{\partial \varepsilon_0}$$

$$(3.17) \quad \frac{\partial x_0}{\partial \varepsilon_0} = \pi (1 - \frac{\partial \lambda_0}{\partial \varepsilon_0})$$

$$(3.18) \quad \frac{\partial (n_1 - n_0)}{\partial \varepsilon_0} = (c+\pi) \frac{\partial \lambda_0}{\partial \varepsilon_0} - \pi \quad ,$$

and from the demand curve that:

$$(3.19) \quad \frac{\partial p_0}{\partial \varepsilon_0} = \frac{1}{2} (1 + \frac{\partial \lambda_0}{\partial \varepsilon_0}).$$

Hence the proposition is proved if we can show that:

$$0 < \frac{\partial \lambda_0}{\partial \varepsilon_0} < \frac{\pi}{\pi + c} \quad .$$

The first inequality follows by inspection of (3.15) since z_1 is less than 1, and z_2 is greater than 1 (and thus surely greater than ρ).

The second inequality can be rewritten:

$$\frac{b(\pi+c)}{(1+r-z_1) + b(\pi+c)} \frac{z_2}{z_2 - \rho} < 1$$

To prove this, it is useful to establish first the following equivalence:

$$(3.20) \quad b(\pi+c) = \frac{1-z_1}{z_1}(1+r-z_1)$$

Proof: By (3.9),

$$z_1^2 - [2 + r + b(c+\pi)]\, z_1 + 1 + r = 0.$$

So:

$$z_1\, b(\pi+c) = z_1^2 - z_1 - (1+r)z_1 + (1+r)$$

$$= (1 + r - z_1)(1 - z_1),$$

from which (3.20) follows. Using (3.20) to eliminate $b(\pi+c)$ and simplifying, the inequality to be proven becomes:

$$\frac{z_2 - z_1 z_2}{z_2 - \rho} < 1 .$$

And this follows from the fact that $z_1 z_2$, the product of the roots of the characteristic equation (3.19), equals $1+r$.

QED

There is certainly nothing surprising about Proposition 1. Our interest, of course, centers on how the derivatives referred to in the propositon vary with changes in the parameters b and ρ. Consider first the parameter b, which indicates the degree of convexity of the inventory carrying cost function. Firms with high (low) b find it very costly (cheap) to vary inventory levels. Hence b measures (with sign reversed) the "flexibility" of inventory holdings. The following proposition sums up the effects of b on the firm's adjustment procedures.

Proposition 2: As b rises, i.e., as output gets less "inventoriable," demand shocks elicit:

(i) larger output responses;

(ii) smaller sales responses;

(iii) larger price responses;

(iv) smaller (in absolute value) responses of inventory investment.

Proof: By inspection of the derivatives of x_0, y_0, p_0 and $n_1 - n_0$ with respect to ε_0, it can be seen that the proposition follows if we can establish that

$$\frac{d}{db}\left(\frac{\partial \lambda_0}{\partial \varepsilon_0}\right) > 0.$$

Use (3.20) to rewrite (3.15) as:

$$(3.21)\quad \lambda_0 = \frac{1-z_1}{\pi+c}\left[\bar{n} - n_0 + \frac{\pi z_2}{z_2-\rho}\ \varepsilon_0\right],$$

from which it follows that:

$$(3.22)\quad \frac{\partial \lambda_0}{\partial \varepsilon_0} = \frac{\pi}{\pi+c}(1-z_1)\frac{z_2}{z_2-\rho} = \frac{\pi}{\pi+c}\ \frac{z_2-(1+r)}{z_2-\rho}$$

since $z_1 z_2 = 1+r$. Thus by direct computation:

$$\frac{d}{db}\left(\frac{\partial \lambda_0}{\partial \varepsilon_0}\right) = \left(\frac{\pi}{\pi+c}\right)\left(\frac{dz_2}{db}\right)\frac{z_2-\rho-[z_2-(1+r)]}{(z_2-\rho)^2}$$

$$= \frac{\pi}{\pi+c}\ \frac{(1+r-\rho)}{(z_2-\rho)^2}\ \frac{dz_2}{db}.$$

So the proposition follows if we can prove that dz_2/db is positive. This follows by differentiating (3.9) with respect to b, recalling that $dB/db = \pi + c$.

$$2z\frac{dz}{db} - \beta\frac{dz}{db} = z(\pi+c)$$

or

$$\frac{dz}{db} = \frac{z(\pi+c)}{2z - \beta}$$

Since $2z_2$ is, by the quadratic formula, $\beta + \sqrt{\beta^2-4(1+r)}$, the above expression is surely positive when evaluated at $z=z_2$.

QED

The intuition behind this result is really quite simple. Firms that can vary inventories rather painlessly (i.e., firms with low b) will use their inventories to absorb most shocks. They will therefore change output and price rather little, while making big sales responses via inventory changes. Firms that lack this flexibility because b is high must rely more on output and price responses.

Before leaving this topic, it is useful to examine two special cases. As $b \to \infty$, output becomes essentially non-storable because n cannot be changed except at prohibitive cost. In that case, it follows from the previous proof that $z_2 \to \beta \to \infty$ and therefore (3.22) implies that $\frac{\partial \lambda_0}{\partial \varepsilon_0} \to \frac{\pi}{\pi+c}$. By (3.16) and (3.17) the sales and output responses both approach $\frac{\pi c}{\pi+c}$, and therefore the inventory response to shocks goes to zero. (The price response in (3.19) goes to $\frac{\pi+\frac{1}{2}c}{\pi+c} < 1$.) As stated at the outset, as compared with a firm that <u>can</u> vary its inventories, a firm that <u>cannot</u> vary its inventories exhibits stronger price and output responses to a demand shock. Put differently, firms with flexible inventory storage technologies will smooth production and limit price fluctuations.

Now turn to the other extreme: a linear inventory cost function (b=0), the case treated by both Zabel (1972) and Amihud and Mendelson (1980). (Reagan (1980) assumed that carrying costs are zero for all levels of N, and hence b=0 for her, too.) A glance at (3.9) shows that $z_2 \to 1+r$ as $b \to 0$, so

(3.22) says that:

$$\lim_{b \to 0} \frac{\partial \lambda_0}{\partial \varepsilon_0} = 0.$$

This case is really quite special since it implies that as $b \to 0$

$$\frac{\partial y_0}{\partial \varepsilon_0} \to 0$$

$$\frac{\partial x_0}{\partial \varepsilon_0} \to \pi$$

$$\frac{\partial p_0}{\partial \varepsilon_0} \to \frac{1}{2} \quad .$$

A firm with such a linear cost structure does not change its production at all in the face of fluctuations in demand, even if these fluctuations are quite persistent. This is certainly not something we expect to be true intuitively.

Now turn to the effect of ρ on the way the firm adjusts to changes in demand. Remember that ρ measures the "temporariness" of shocks. $\rho = 1$ means that all shocks are permanent (a random walk), while $\rho = 0$ means that shocks are i.i.d. The following proposition summarizes the results.

Proposition 3: As ρ rises, i.e., as demand shocks become more persistent, these shocks elicit:

(i) larger output responses;

(ii) smaller sales responses;

(iii) larger price responses

(iv) smaller (in absolute value) responses of inventory investment.

Proof: Once again, the statements follow immediately if we can establish that $\frac{d}{d\rho} \left(\frac{\partial \lambda_0}{\partial \varepsilon_0} \right) > 0$. But now this is obvious from (3.22).

QED

The intuition here is really quite straightforward. If shocks are relatively permanent, it does not pay to use inventory fluctuations to buffer them. Output and price respond relatively strongly. On the other hand, inventories <u>do</u> play a major role in buffering relatively transitory shocks. Here price and output do not respond much, but sales do (as inventories are disgorged or accumulated).

Theorem 1 is simply the combination of Propositions 1-3, and hence is proven.

4. Further Results on Inventories

4.1 Reactions to Inventory Disequilibrium

The last section derived results on how the firm's output, sales, and price respond to changes in demand. This section uses the same apparatus to study how the firm responds to changes in its initial level of inventories, n_0. First we summarize the effects of n_0 on optimal y_0, x_0, p_0 and n_1.

<u>Proposition 4</u>:[1] Other things equal, a higher initial level of inventories leads the firm to produce less, sell more, charge a lower price, and reduce its inventory investment.

<u>Proof</u>: We saw in Section 3 that the only stable path to the saddle point in Figure 4 was negatively sloped. It therefore follows that:

$$(4.1) \qquad \frac{\partial \lambda_0}{\partial n_0} < 0.$$

Since optimal y_0 is defined by (3.5),

$$\frac{\partial y_0}{\partial n_0} = c \frac{\partial \lambda_0}{\partial n_0} < 0,$$

so that y_0 is a decreasing function of n_0. Since optimal x_0 is defined by (3.6),

$$\frac{\partial x_0}{\partial n_0} = -\pi \frac{\partial \lambda_0}{\partial n_0} > 0,$$

so that x_0 is increasing in n_0. Since the demand curve slopes down, this last

[1] This proposition provides the microfoundations for the Blinder-Fischer (1981) model.

finding also implies that p_0 is decreasing in n_0. The response of $n_1 - n_0 = y_0 - x_0$ follows by simple arithmetic.

QED

Once again, the proposition is intuitively appealing. Firms stuck with high inventories would normally be expected to cut production and "run a sale," which is just what the proposition says. However, the proposition is a bit more fragile than might be expected. For example, suppose that inventory carrying costs are linear, as the previous studies have all assumed. Then $b = 0$, the $\Delta\lambda = 0$ locus in Figure 4 is horizontal, and λ_0 is independent of n_0.[1] Quite contrary to intuition, the shadow value of inventories is independent of the amount of inventory on hand. This is one reason why extending the results to a nonlinear $B(n)$ function is crucial.

4.2 The Concept of Desired Inventories

The concept of "desired inventories" has proven to be an elusive one for students of empirical inventory behavior. Typically it is assumed, for no good reason, that desired inventories are an increasing linear function of current sales:

$$(4.2) \quad n_t^* = a_0 + a_1 x_t , \quad a_1 > 0,$$

or perhaps of expected sales. I propose the following definition of "desired inventories" which seems natural in the context of this model.

Definition: Desired inventory n_t^*, is the level of inventory that would make desired inventory change exactly zero, once the current demand shock is realized.

The linear-quadratic structure of the problem enables us to put some empirical teeth on this concept. Define n_0^* as the value of n_0 that would make optimal n_1 equal to n_0, that is, would make optimal inventory investment equal

[1] In taking the limit as $b \to 0$, it is important that the ratio b_1/b approaches some finite constant.

to zero. The exact solution can be worked out by using (3.5) and (3.6) to find the value of λ_0 that equates y_0 and x_0, and then using (3.15) to find the value of n_0 that produces this value of λ_0. The answer is:

$$(4.3) \quad n_0^* = \bar{n} - \frac{\pi(1+r-\rho)}{(z_2-\rho)(1-z_1)} \varepsilon_0$$

Notice that even when shocks are completely transitory ($\rho=0$), desired inventories still respond to the current demand shock. Specifically, when $\rho = 0$:

$$\frac{\partial n_0^*}{\partial \varepsilon_0} = -\pi \frac{1+r}{z_2-(1+r)} < 0.$$

At the other extreme, if all shocks are permanent ($\rho=1$), (4.3) becomes:

$$n_0^* = \bar{n} - \frac{\pi r \varepsilon_0}{b(c+\pi)}$$

In this case, we have what amounts to complete admustment of $\bar{n}$ to any observed demand shock because, as can be verified from the definition of $\bar{n}$, the response of n_0^* to ε_0 is identical to the response of $\bar{n}$ to π_0/π (which is the intercept of the demand curve).

The two optimal inventory concepts, n_0^* and $\bar{n}$, provide a rigorous basis for Feldstein and Auerbach's (1976) intuitive "target adjustment" model. The long-run target, $\bar{n}$, depends on production costs, inventory carrying costs, the rate of interest, and the long-run position of the firm's demand curve. It would not be expected to change very often or very quickly. This period's desired inventory stock, however, will deviate from the long-run target to reflect the current state of demand. The two concepts are far apart when shocks are transitory ($\rho \simeq 0$), close together when shocks are permanent ($\rho \simeq 1$).

Finally, a bit more algebra gives us the "inventory investment equation" implied by the model:

$$(4.4)\quad n_1^* - n_0 = \left[\frac{b(c+\pi)}{1+r-z_1+b(c+\pi)}\right](n_0^* - n_0) = (1-z_1)(n_0^* - n_0),$$

which clearly brings out the partial adjustment nature of inventory investment. The positive adjustment coefficient, $1-z_1$, depends on the convexity of the inventory carrying cost function, b, the rate of interest, r, the convexity of the production cost function, c, and the slope of the demand curve, π (see equation (3.9)). In particular, it is worth noting that since $\lim_{b\to 0} z_1 = 1$, the partial adjustment feature of inventory change is entirely lost when marginal inventory costs are constant--the case dealt with in the earlier literature.

There is, however, one feature of the model that is troublesome empirically. Since $0 \leq \rho \leq 1$, $z_2 > 1$, and $z_1 < 1$, the coefficient of ε_0 in equation (4.3) is negative; the current demand shock reduces desired inventory holdings. In contrast, when equations like (4.2) are embedded in an empirical model and estimated, the parameter a_1 is invariably estimated to be positive. This seems to be the one instance in which reality contradicts the model. I address this problem in a subsequent paper.

4.3 What Kinds of Firms React Strongly to Inventory Disequilibrium?

Finally, I present and prove a theorem analogous to Theorem 1 that summarizes how the sensitivity of a firm's decisions (y_0, x_0, p_0 and n_1) to its initial holdings of inventories (n_0) varies with its characteristics (b and ρ).

Theorem 2: Firms whose outputs are more "inventoriable," i.e., which have lower b, change their optimal levels of output, sales, price, and inventory investment less in response to inventory disequilibrium than firms whose outputs are less inventoriable. The degree of persistence of demand shocks, however, is irrelevant to the magnitude of the firm's reactions to inventory disequilibrium.

Proof: Given the apparatus already developed, the proof is almost immediate. Start by observing that (3.5), (3.6) and (3.21) imply:

$$(4.7)\quad \frac{\partial x_0}{\partial n_0} = -\pi \frac{\partial \lambda_0}{\partial n_0} = \frac{\pi}{\pi+c}\ (1-z_1)$$

$$(4.8)\quad \frac{\partial y_0}{\partial n_0} = c\,\frac{\partial \lambda_0}{\partial n_0} = -\ \frac{c}{\pi+c}\ (1-z_1)$$

$$(4.9)\quad \frac{\partial (n_1-n_0)}{\partial n_0} = \frac{\partial y_0}{\partial n_0} - \frac{\partial x_0}{\partial n_0} = -(1-z_1)$$

$$(4.10)\quad \frac{\partial p_0}{\partial n_0} = -\frac{1}{2\pi}\ \frac{\partial x_0}{\partial n_0} = -\frac{1}{2(\pi+c)}\ (1-z_1)\ .$$

I have already shown in an earlier proof that:

$$\frac{dz_1}{db} = \frac{z_1(c+\pi)}{2z_1-\beta}$$

which is negative since $0 < z_1 < 1$ and $\beta > 2$. Thus as b falls, z_1 rises toward unity, and by (4.7) - (4.10) the sales, production, price, and inventory-investment responses all fall in absolute value (without changing sign). The rest of the proposition follows by noting that ρ appears nowhere in any of these expressions.

QED

The intuition behind this theorem is again clear. Firms whose inventory carrying costs rise rapidly when they have excessive (or deficient) inventories will have to take decisive actions to correct their inventory disequilibrium whereas firms whose inventory cost functions are relatively linear can afford the luxury of waiting things out.

5. Macroeconomic Implications

While the results presented here stand on their own as part of the theory of monopoly under uncertainty, my own interest in them is from the point of view of providing microfoundations for macroeconomics. Viewed from this perspective, the model has several interesting implications.

5.1 Price Rigidity

On the surface, it would appear that the analysis provides an explanation for the macroeconomic phenomenon of price rigidity which is consistent with maximizing behavior: prices tend to move sluggishly in industries whose outputs are inventoriable. Thus industries with perishable output (e.g. agriculture) are more likely to be "flexprice" industries while those with easily storable output are more likely to be "fixprice" industries.

However, such a conclusion would be a bit hasty. The macroeconomic phenomenon that needs explanation is the stickiness of absolute prices, while the model presented here--like those of Zabel (1972), Amihud-Mendelson (1980) and Reagan (1980)--provides an explanation for rigidity in relative prices.[1] The firm's nominal price in this model will respond less than completely to the general price level only if its nominal cost functions respond less than completely. Thus, it would appear, price ridigity is "explained" only by assuming cost (i.e., wage) rigidity. This is a feature shared by all attempts to provide microfoundations for nominal price rigidity because demand and supply curves derived from maximizing behavior are always homogenous of degree zero in all nominal magnitudes.

Let us consider what relative price rigidity means in an inflationary environment. A firm that behaves according to the model presented here, but that finds itself in an economy with a persistently rising price level,

[1] This is easily seen by inspecting the solutions. If all nominal parameters doubled, say, the firm's price would also double and nothing real would change.

wants its relative price to move sluggishly. To accomplish this, it must continually raise its absolute price more or less in line with the overall rate of inflation. In such a world, we would not observe stickiness of nominal prices; they would change quite frequently. (In fact, if they did not change frequently enough, relative prices would change more than firms desire.) Instead, we would find a tendency for the actual inflation rate of particular products to gravitate toward the expected aggregate inflation rate. Then, if it is the case that expected inflation rate can be nudged downward only by decreasing the actual inflation rate (a point which is vigorously disputed by the rational expectationists), we will have a very sluggish inflation rate which stubbornly resists disinflationary policy, but which responds quickly to permanent changes in the rate of increase of costs.

In any case, the general lesson seems to be that microfoundations of price rigidity seem only to push the question back one stage. Instead of asking: why is the inflation rate so persistent?, we must ask instead: why is it so hard to reduce inflationary expectations? Looking across industries, however, the model does seem to imply that the rate of relative price change will be more sluggish where output is most "inventoriable" (durable goods producers?) and less sluggish where output is least "inventoriable" (agriculture?).

5.2 Stockouts and Downward Price Rigidity

It is often supposed--with little supporting evidence, I believe--that prices are more rigid downward than upward. If the model presented here is altered to prohibit negative inventories ("unfilled orders"), then it will generate asymmetrical price responses; prices will react more strongly to increases in demand than to decreases in demand. Since this asymmetry is the major point of the papers by Amihud-Mendelson (1980) and Reagan (1980), and since I am far from convinced that it is of great empirical importance, I develop it only briefly here.

For this purpose, it is natural to set K, the critical level of inventories, equal to zero (so that n=N) and to impose the constraint $N(t) \geq 0$ for all t. When the firm does not stock out, the constraint is not binding, and the dynamic system follows the same equations as before. However, if the firm stocks out ($N_0 = 0$), and the unconstrained solution would be calling for $N_1 < N_0$, the constraint $N_1 \geq 0$ becomes binding. As we know, $N_1 = N_0$ only when the shadow value of inventories takes on the value $\lambda_0 = \frac{\pi}{\pi+c}\varepsilon_0$, so that by (3.17):

$$\frac{\partial x_0}{\partial \varepsilon_0} = \pi(1 - \frac{\partial \lambda_0}{\partial \varepsilon_0}) = \frac{\pi c}{\pi+c} .$$

By comparison, in the unconstrained case it was proven earlier (see the proof of Proposition 1) that: $\frac{\partial \lambda_0}{\partial \varepsilon_0} < \frac{\pi}{\pi+c}$ and hence that $\frac{\partial x_0}{\partial \varepsilon_0} > \frac{\pi c}{\pi+c}$. In a word, the sales response is smaller when the firm has stocked out, and hence the price response must be greater. This is the asymmetry result.

At the macro level, it appears likely that the number of firms experiencing a stockout is greater at higher levels of macroeconomic activity. Hence price responses to demand shocks should be greater at high levels of activity than at low levels.

5.3 Investing in Flexibility

The findings of the model can be placed in a broader context. How does a firm plan for and react to fluctuations in demand? The conventional view is that there is little or no advance planning and there are two ways to react when a shift in demand occurs: change price or change output. If these are the only avenues of response, then they clearly must be alternatives: the more one is used the less the other is needed.

The model presented here shows that once a third method of adjustment --building up or drawing down inventories--is allowed, sales and production can, and normally will, respond to shocks differently, and strong price responses may be associated with strong output responses rather than with weak ones. Thinking in a somewhat longer time frame, of course, firms understand the greater flexibility in dealing with unanticipated events that variable inventory stocks can buy for them, and should plan their inventory storage facilities accordingly. They should be willing to invest in acquiring flexibility (i.e., reducing b) by organizing their production and inventory procedures to make their outputs more inventoriable. How much they are willing to invest in lowering b depends, of course, on the costs and benefits of such investments.

Furthermore, inventories are not the only possible vehicle for enhancing flexibility. The phenomenon of labor hoarding has often been noted in empirical macroeconomics, especially in the context of explaining the procyclical pattern of labor productivity. Holding excess supplies of labor in reserve is another way that the firm can achieve flexibility. That is, inventories of labor may be partial substitutes for inventories of goods.[1] Similar remarks may be made about plant and equipment. It may well be rational for firms facing stochastic demand to invest in more capacity than they expect to use in normal times so they are better positioned to take advantage of periods in which demand is higher than normal.

6. Summary

1. When output is not storable, firms can react to increases in demand only by raising prices or by boosting production. The more they do of one, the less they have to do of the other.

2. However, when output is storable, the <u>same</u> firms that raise prices a lot may also raise production a lot, while other firms may raise both price and output rather little. Firms in the latter category will take up

[1] This idea is developed more fully in some work in preparation by Robert Topel.

the slack by selling a lot out of inventories.

3. Firms whose marginal costs of inventory holding are relatively constant and whose demand shocks are very transitory will rely heavily on inventory changes to absorb shocks, and will vary price and output rather little. Conversely, firms with sharply rising marginal inventory costs and/or rather permanent demand shocks will rely less on inventories as buffer stocks and will exhibit larger price and output fluctuations.

4. The previous literature has assumed that inventory carrying costs are linear (or zero). As long as stockouts are avoided, this extreme assumption leads to the conclusion that production is totally unresponsive to fluctuations in demand. In macroeconomic terms, the critical link between fluctuations in demand and fluctuations in employment is missing in such models.

5. Other things equal, firms with higher inventories will produce less, charge lower prices (i.e., sell more), and accumulate fewer additional inventories. Thus, at the macro level, production, employment, and prices should all respond negatively to high levels of inventories.

6. The reactions of output, price, and sales to inventory stocks are strongest when the firm's inventory carrying cost function is most convex, weakest when it is close to linear.

7. Inventory investment is characterized by the "partial adjustment" specification that is so popular in empirical work: inventory change is proportional to the gap between desired and actual inventory holdings. Desired inventories in the model are a <u>decreasing</u> function of the current demand shock.

8. The model helps provide an explanation for sluggish relative prices, not sluggish absolute prices. However, under certain circumstances, sluggish reactions of relative prices may help explain the persistence of inflation.

9. If negative inventories are impossible, and stockouts occur instead, prices will be more sensitive to positive demand shocks than to negative demand shocks.

10. Investment in inventory carrying capacity can be viewed as one of several ways for a firm to acquire greater flexibility in reacting to unanticipated events.

LIST OF REFERENCES

Y. Amihud and H. Mendelson, "Monopoly under Uncertainty: The Enigma of Price Rigidity," mimeo, March 1980.

A.S. Blinder and S. Fischer, "Inventories, Rational Expectations, and the Business Cycle," Journal of Monetary Economics, 1981 forthcoming.

M.S. Feldstein and A. Auerbach, "Inventory Behavior in Durable Goods Manufacturing: The Target-Adjustment Model," Brookings Papers on Economic Activity, 2, 1976, pp. 351-396.

L.J. Maccini, "An Aggregate Dynamic Model of Short-Run Price and Output Behavior," Quarterly Journal of Economics, 1976.

L. Phlips, "Intertemporal Price Discrimination and Sticky Prices," Quarterly Journal of Economics, 92 (May 1980), 525-542.

P.B. Reagan, "Inventories and Asymmetries in Price Adjustment," mimeo, MIT, April 1980.

P.B. Reagan and M.L. Weitzman, "Asymmetries in Price and Quantity Adjustments by the Competitive Industry," mimeo, MIT, April 1980.

E. Zabel, "Multiperiod Monopoly under Uncertaintly," Journal of Economic Theory, 5, (1972), pp. 524-536.