1 Small Menu Costs and Large Business Cycles: A Macroeconomic Model of Monopoly

N. Gregory Mankiw

1 Introduction

The conflict between modern neoclassical and traditional Keynesian theories of the business cycle centers upon the pricing mechanism.[1] In neoclassical models, prices are fully flexible. They represent the continuous optimization of economic agents and the continuous intersection of supply and demand. In Keynesian models, prices are often assumed to be sticky. They do not necessarily equilibrate all markets at all times. One of the reasons for the resurgence of the equilibrium approach to macroeconomics has been the absence of a theoretical underpinning for this Keynesian price stickiness.

This note shows that sticky prices can be both privately efficient and socially inefficient. The business cycle results from the suboptimal adjustment of prices in response to a demand shock. To the extent that policy can stabilize aggregate demand, it can mitigate the social loss due to this suboptimal adjustment.

In some Keynesian models, prices are simply exogenously fixed.[2] In others, agents must set their prices in advance of the transaction date.[3] The act of altering a posted price is certainly costly. These costs include such items as printing new catalogs and informing salesmen of the new price. Yet these "menu" costs are small and, therefore, generally perceived as providing only a weak foundation for these fixed-price models. However, this inference is flawed. Small menu costs can cause large welfare losses. The claim that price adjustment costs are small does not rebut the claim that they are central to understanding economic fluctuations.

I present a simple static model of a monopoly firm's pricing decision. The firm sets its price in advance, and changes that price ex post only by incurring a small menu cost. I show the firm's price adjustment decisions are in no sense socially optimal.[4]

2 A Partial-Equilibrium Analysis

Consider a monopoly firm facing a constant cost function (1) and the inverse demand function (2):[5]

Quarterly Journal of Economics 100 (May 1985): 529–539. Copyright © 1985 by the President

$$C = kqN \tag{1}$$

$$P = f(q)N, \tag{2}$$

where C is the total nominal cost of producing quantity q, k is simply a constant, P is the nominal price the firm obtains if it places q on the market, and N is a nominal scale variable. The variable N denotes the exogenous level of aggregate demand. It can be regarded as the overall price level, the money stock, or the level of nominal GNP. Both the nominal cost to the firm C and the nominal price it receives P increase proportionally to the level of nominal demand N.

Figure 1 shows the standard profit-maximizing solution to the monopoly firm's problem.[6] An increase in the nominal scale variable shifts the cost and demand functions proportionally, thereby increasing P_m without affecting the firm's output q_m.

Let $c = C/N$ and $p = P/N$. The firm's problem is then viewed as independent of aggregate demand. That is, if faces (1′) and (2′), which do not shift when N changes:

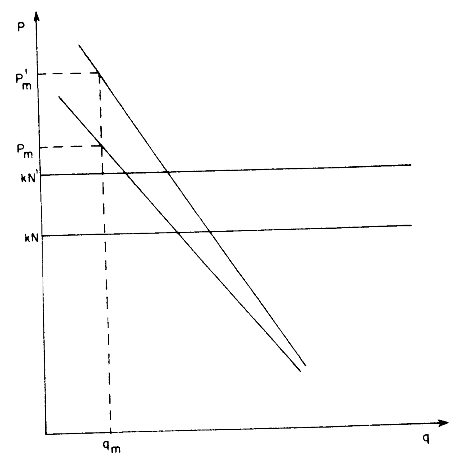


Figure 1

$$c = kq \tag{1′}$$

$$p = f(q) \tag{2′}$$

(see figure 2). The firm chooses p and q to maximize profits. The nominal price it sets is $p_m N$. (I hereafter refer to p as the "price," even though it is the nominal price normalized by the nominal scale variable.) The firm earns profits (producer surplus) equal to the rectangular area between k and p_m. Consumer surplus—the excess of value over price—equals the triangle above profits. Total surplus, which is the measure of welfare used in this paper, is the sum of profits and consumer surplus.

Now suppose that the firm needs to set its nominal price one period ahead based upon its expection of aggregate demand N^e. It sets the nominal price equal to $p_m N^e$. If its expectation turns out correct ex post, then the observed price p_0 is p_m. If it turns out incorrect, then the observed price is $p_0 = p_m(N^e/N)$.[7]

If N is lower than expected, p_0 is higher than p_m (see figure 3). In this case, the firm's profits are lower by the area $B - A$, which is positive, since p_m is

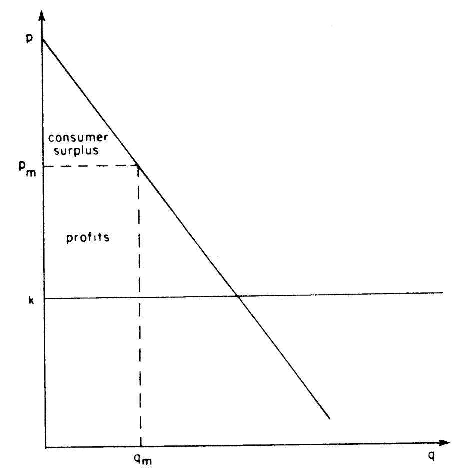


Figure 2

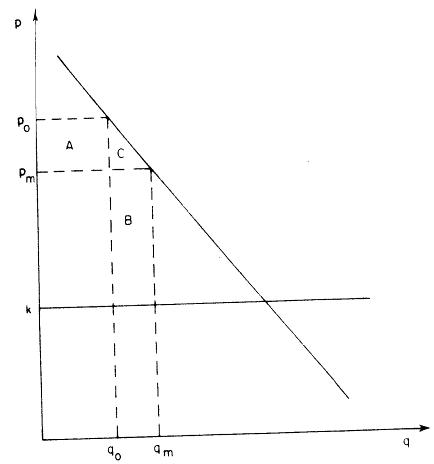


Figure 3

by definition the profit-maximizing price. Total surplus is reduced by $B + C$. Hence, the reduction in welfare due to a contraction in aggregate demand is greater than the reduction in the firm's profits.

Suppose that the firm is able to change its nominal price ex post, but only at a menu cost of z. If it chooses to do so, it reduces the price from p_0 to p_m and obtains the additional profits $B - A$. Hence, it lowers its price if and only if $B - A > z$. Yet, from the standpoint of a social planner, the firm should lower its price to p_m if and only if $B + C > z$. Thus, we obtain these results:

PROPOSITION 1 Following a contraction in demand, if the firm cuts its price, then doing so is socially optimal.

Proof of proposition 1 If the firm cuts its price, then $B - A > z$. Hence, $B + C > z + A + C > z$. Thus, doing so is socially optimal.

PROPOSITION 2 Following a contraction in demand, if $B + C > z > B - A$, then the firm does not cut its price to p_m, even though doing so would be socially optimal.

Proposition 2 suggests the downward price rigidity often mentioned in macroeconomic debate. The inefficiency results because there is an external benefit of $C + A$ in printing new menus. How large is this externality? A natural measure is the ratio of the social benefit from a price adjustment $B + C$ to the private benefit $B - A$. This ratio, of course, depends upon the size of the demand shock. Since the firm would adjust to the profit-maximizing price p_m, rather than the first-best price k, the increment to profits $B - A$ is of second order, while the increment to welfare $B + C$ is of first order. Therefore, this ratio approaches infinity as the size of the shock approaches zero. The ratio is more meaningful if evaluated for a shock of typical size. Hence, I compute it for a 1 percent contraction.[8] If the demand function has a constant price elasticity of ten, the ratio is twenty-three. If the elasticity is two, the ratio is over two hundred. For any plausible demand function, the social gains from price adjustment far exceed the private gains.

PROPOSITION 3 A contraction in aggregate demand unambiguously reduces welfare as measured by the sum of producer and consumer surplus. If the firm cuts its price in response, then the contraction only has the menu cost z. If not, then the contraction has the possibly much larger cost $B + C$.

Now let us examine an expansion in aggregate demand. Since $N > N^e$, we know that $p_0 < p_m$. At first, let $p_0 > k$ ($N/N^e < p_m/k$,) as in figure 4. The firm's profits are reduced by $D - F$, which is positive, since p_m is profit-maximizing. Total surplus is increased by $E + F$. The firm resets its price if the cost of doing so is justified by increased profit. That is, the firm raises its price back to p_m if and only if $D - F > z$.

PROPOSITION 4 Following an expansion in demand in which $N/N^e < p_m/k$, if the firm resets its price, total surplus is decreased by the menu cost. If the firm does not reset its price, surplus is increased by $E + F$.

If $N/N^e > p_m/k$, then after the demand expansion, p_0 is below k (see figure 5). Total surplus decreases by $I - J$, which could be negative or positive, making the welfare effect ambiguous. Firm profits (now negative) have been reduced by $G + H + I$. The firm resets its price to p_m if $G + H + I > z$. It is socially optimal to do so if $I - J > z$.

PROPOSITION 5 Following an expansion in which $N/N^e > p_m/k$, if the firm resets its price, total surplus is decreased by the menu cost. If the firm does

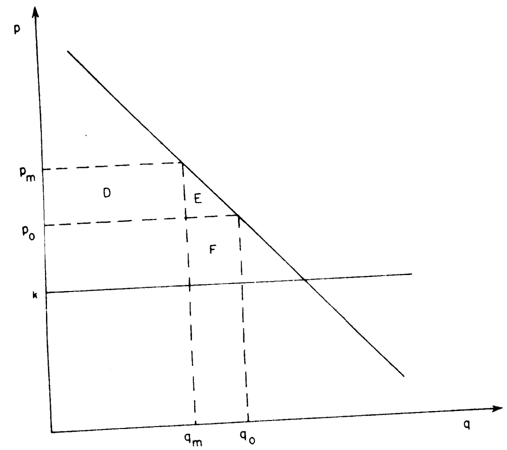
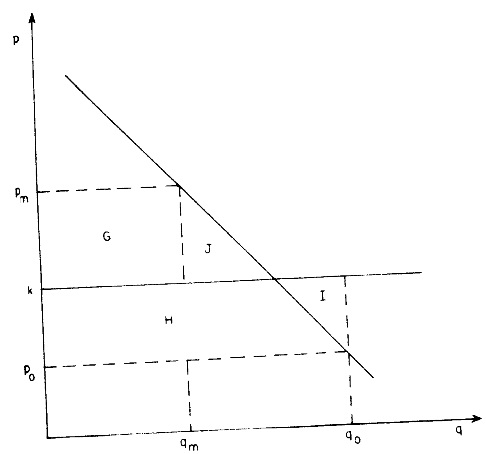


Figure 4



not reset its price, total surplus change is ambiguous, but total surplus does not decrease by more than the menu cost.

Proof of proposition 5 If the firm does not reset its price, then $G + H + I < z$. This implies that $I - J < z - J - G - H < z$. Hence, total surplus reduction, $I - J$, is less than the menu cost.

As a partial summary of the above results:

PROPOSITION 6 An expansion in aggregate demand reduces welfare by no more than the menu cost, and may even increase welfare. A contraction in aggregate demand unambiguously reduces welfare, possibly by much more than the menu cost.

3 Conclusion

The economy I describe is Keynesian, even though all agents are optimizing and all prices result from that optimization. The central postulate is that monopoly firm must incur a small menu cost if it alters its posted price after an aggregate demand shock. I show that the firm's price adjustment decisions are suboptimal. In addition, the welfare loss can far exceed the menu cost that is its cause.

The model also displays an asymmetry between contractions and expansions, since the natural rate of output is below the social optimum. Private incentives produce too much price adjustment following an expansion in aggregate demand and too little price adjustment following a contraction in aggregate demand. From the viewpoint of a social planner, the nominal price level may be stuck too high, but it is never stuck too low. In this sense, prices are downwardly rigid but not upwardly rigid.[9] Furthermore, the model's asymmetry parallels another observed phenomenon; namely, that while aggregate demand contractions are associated with grotesquely inefficient underproduction, aggregate demand expansions are not associated with similarly inefficient overproduction. There is no obverse to the Great Depression. Instead, periods of expansion, such as the late 1960s in the United States, are often considered periods of economic prosperity.[10]

The analysis presented here is all in the context of partial equilibrium. It is, however, possible to construct simple general equilibrium examples that encompass exactly these partial-equilibrium results.[11] I suspect that a more complete general equilibrium model would exhibit more pronounced price stickiness. In particular, the introduction of interfirm purchases would exacerbate price rigidity. In such a model the failure of one firm to reduce its price following a contraction in demand would prevent the costs of other firms from falling, thereby reducing those firms' incentive to cut prices. The primary qualitative conclusion—that trivial menu costs can have important efficiency effects—would certainly remain true in the context of general equilibrium.[12]

The theme of this paper appears robust: In almost all economic models, agents who have the power to affect prices, exert that power by restricting output. The economy's equilibrium, or natural rate, is thus below the social optimum.[13] Because of this, deviations below the natural rate impose greater costs on society than on the price-setting agents. These agents, therefore, have inadequate incentive to return the economy to its equilibrium.

An economy of this sort does not recommend passive monetary policy. As long as new information about exogenous demand factors (e.g., velocity) is made available to the monetary authority after private agents set their prices, systematic feedback rules can stabilize output.[14] These exogenous demand shocks cause substantial and inefficient fluctuations in output and employment if the monetary authority does not react. Although firms optimize, their prices are not socially optimal, and in particular, respond too little to adverse demand shocks. This inefficiency appears to be the target of policies that aim directly at the pricing mechanism, such as wage-price controls and tax-based incomes policy.

Appendix: A General-Equilibrium Example[15]

The real world is replete with intricate interrelations between product markets and labor markets. For the most part, individuals sell their labor at a determined wage and buy goods at a determined price. Firms buy labor at a determined wage and sell their goods in markets were they exercise at least some control over their prices. To illustrate how menu costs fit into a general equilibrium, I present a simple example that encompasses the partial equilibrium results in the paper.

This economy is composed of individuals and firms, both of which are distributed along the unit interval. All individuals are identical. Thus, their behavior is represented by a single individual, and I need not be concerned

hereafter with the fact that there is more than one. The firms produce differentiated products. They also differ in their menu costs. Thus, after a monetary shock, some firms decide to alter their posted price, and some firms do not.

The economy comprises a labor market, a money market, and a continuum of product markets. It differs from a standard market-clearing competitive economy in only one respect. On the supply side of each product market is a monopolist. The Walrasian auctioneer, who costlessly adjusts prices, is replaced with price-setting firms, each with its own menu cost.[16]

Our representative individual has the following utility function:

$$U = (1-\phi)^{-1}\int_0^1 q_i^{1-\phi}\,di + \theta\log(M^d/P) - L, \qquad \text{(A1)}$$

where U is utility, q_i is the quantity of product i he consumes, ϕ is the reciprocal of the elasticity of substitution between the different products $(0 < \phi < 1)$, M^d is his money demand, P is the general price level, L is his labor supply, and θ is a money demand parameter $(\theta > 0)$. The general price level P is the geometric average of all P_i, where P_i is the nominal price of the good produced by firm i. That is,

$$P = \exp\left(\int_0^1 \log P_i\,di\right). \qquad \text{(A2)}$$

Given P_i for all i and the nominal wage W, our individual maximizes U subject to his budget constraint (A3).

$$\int_0^1 P_i q_i\,di + M^d = WL + M + \text{Profits} \qquad \text{(A3)}$$

M is the money supply, which begins with the individual. Profits of the firms go to the individual to assure that Walras's Law is satisfied. Of course, the individual takes profits as fixed in his maximization problem.

With this particular utility function, the first-order conditions give product demand and money demand.

$$q_i = (P_i/W)^{-1/\phi} \quad \text{for all } i \qquad \text{(A4)}$$

$$M^d = \theta W \qquad \text{(A5)}$$

Labor supply comes from (A3), (A4), and (A5). I do not explicitly solve for equilibrium in the labor market, as this is guaranteed by Walras's Law coupled with equilibrium both in the product markets and in the money market.

Equilibrium in the money market is simply money supply equaling money demand.

$$M = M^d \qquad \text{(A6)}$$

From (A4) and (A5), with simple algebraic rearrangement, we obtain

$$P_i = \theta^{-1} M q_i^{-\phi}, \qquad \text{(A7)}$$

$$W = \theta^{-1} M. \qquad \text{(A8)}$$

Equation (A7) is the inverse demand function (1) faced by the firm. The price elasticity of demand is $1/\phi$.

Each firm takes (A7) as its demand function and (A9) as its production function:

$$q_i = L_i, \qquad \text{(A9)}$$

where L_i is labor input of firm i. From (A8) and (A9) we obtain the nominal cost function (A10).

$$C_i = W L_i = \theta^{-1} M q_i \qquad \text{(A10)}$$

The implied profit function is

$$\Pi_i = (q_i^{1-\phi} - q_i)\theta^{-1} M. \qquad \text{(A11)}$$

The profit maximizing output q_m for each firm is

$$q_m = (1-\phi)^{1/\phi}. \qquad \text{(A12)}$$

At this quantity, each price is

$$P_m = \theta^{-1} M (1-\phi)^{-1}. \qquad \text{(A13)}$$

Since $\theta^{-1} M$ is all that enters into the above expressions, increases in money supply M are exactly the same as reductions in money demand θ. For simplicity I hold θ constant hereafter and examine changes in M only.

Suppose each firm in the economy is at its "natural rate" of q_m. The money supply is M_0, and thus each price is $P_0 = \theta^{-1} M_0 (1-\phi)^{-1}$. Suddenly the money supply is changed to M_1. Suppose all the posted prices at first remain unchanged at P_0 even though the profit-maximizing price is now $P_m = \theta^{-1} M_1 (1-\phi)^{-1}$. Through the money market equilibrium (A6), the nominal wage changes from $\theta^{-1} M_0$ to $\theta^{-1} M_1$. Through product demand (A7), output changes from q_m to $q_0 = (M_1/M_0)^{1/\phi} q_m$ for all firms.

Now suppose each firm can change its menu ex post, but to do so would require a small labor input g. In particular, let the menu cost of firm i be

$$z_i = g(i) W = g(i)\theta^{-1} M \quad g' > 0 \qquad \text{(A14)}$$

Each firm, when deciding whether to post a new price, compares the increment in profit $\Delta\Pi$, as calculated from (A11), to its menu cost z_i. Thus, the marginal firm I is

$$I = g^{-1}(\Delta\Pi/W) = g^{-1}[(q_m^{1-\phi} - q_0^{1-\phi}) - (q_m - q_0)] \qquad \text{(A15)}$$

If $i < I$, then the firm finds it profitable to incur the menu cost, post the price P_m, and produce output q_m. If $i > I$, then the firm leaves its price at P_0, and produces q_0. Total output is therefore

$$Q = \int_0^1 q_i \, di = I q_m + (1-I) q_0. \qquad \text{(A16)}$$

The general price level is

$$P = \exp\left(\int_0^1 \log P_i \, di\right) = \exp[I \log P_m + (1-I)\log P_0]. \qquad \text{(A17)}$$

Employment equals output through the production functions (A9) plus the labor input necessary for changing prices.

$$L = Q + \int_0^I g(i)\, di \qquad \text{(A18)}$$

Utility is

$$U = (1-\phi)^{-1}[I q_m^{1-\phi} + (1-I) q_0^{1-\phi}] + \theta \log M - \theta[I \log P_m + (1-I)\log P_0] - \left[I q_m + (1-I) q_0 + \int_0^I g(i)\, di\right]. \qquad \text{(A19)}$$

In addition, straightforward algebra demonstrates that the real wage is procyclical.

The partial equilibrium analysis of the paper shows that the ex post price-adjustment decisions by profit-maximizing firms are not socially efficient. The same result can be shown in this general equilibrium example. Suppose a social planner could dictate which firms adjust prices ex post. In other words, instead of I being determined endogenously as in (A15), the planner chooses the optimal I^* to maximize U in (A19). The first-order condition[17] is

$$\frac{\mathrm{d}U}{\mathrm{d}I} = 0. \qquad \text{(A20)}$$

This implies

$$I^* = g^{-1}[(1-\phi)^{-1}(q_m^{1-\phi} - q_0^{1-\phi}) - \theta \log(P_m/P_0) - (q_m - q_0)]. \qquad \text{(A21)}$$

For a monetary expansion, $P_m > P_0$ and $q_m < q_0$. Comparing (A15) and (A21) shows that $I^* < I$. The social planner would have fewer firms adjust prices. Similarly, for a monetary contraction, $P_m < P_0$, $q_m > q_0$, and thus $I^* > I$. The social planner would have more firms adjust prices.[18] Private incentives produce too much price adjustment following a monetary expansion and too little price adjustment following a monetary contraction.[19]

Acknowledgements

I am grateful to Olivier Blanchard, Alan Blinder, Avery Katz, Eric Rasmusen, James Rauch, Deborah Roloff, David Romer, Julio Rotemberg, Robert Solow, Lawrence Summers, and two anonymous referees for helpful comments and to the National Science Foundation for financial support.

Notes

1. See Gordon (1981) for a general discussion of the role of price adjustment in macroeconomic debate.

2. See, for example, Malinvaud (1977). Gordon (1981) references other fixed-price models.

3. See, for example, Fischer (1977) and Blinder and Mankiw (1984). The economy I describe in this paper is a relative of the Type 3 (nominal price contracts) economy in Blinder and Mankiw.

4. Rotemberg (1982) examines a model of monopoly firms with quadratic price adjustment costs. He emphasizes the dynamic behavior of prices and aggregate output, rather than the welfare properties of the firm's decisions.

5. Allowing a more general cost function does not alter the results.

6. Although the demand functions in the figures are linear, the results do not depend upon this functional form.

7. I assume that the firm cannot ex ante announce a nominal price P indexed to N. The reasons ... might not arise are beyond the scope of this note. See Gray (1976)

8. Barro and Rush (1980) find that the standard deviation of the postwar U.S. monetary shocks is 1.4 percent with annual data and 0.5 percent with quarterly data.

9. The price adjustment rule followed by the firm is not itself asymmetric. Instead, the welfare properties of the adjustment process exhibit asymmetry. Studies that concentrate upon the positive aspects of the adjustment process (e.g., Barro 1972), rather than its normative aspects, thus report no asymmetry.

10. These periods are often considered times of excessive inflationary pressure. Yet there is little concern about the level of output per se.

11. An earlier version of this paper contained such an example. The simplest general equilibrium example contains n yeoman farmers, each choosing between leisure and production of his uniquely differentiated output. In equilibrium, each produces too little. The price adjustment rule each follows is suboptimal, as in the partial-equilibrium analysis presented in this note.

12. The dynamic nature of the price-setting process, which is undoubtedly important, is ignored in this paper. Whether a contraction is viewed as temporary or permanent probably affects the reaction of firms. Rotemberg (1982), in the context of a somewhat different model, considers the dynamics of price adjustment in more detail. Another important aspect of the problem ignored here is the effects of price desynchronization. Blanchard (1982) explicitly examines this issue.

13. Okun (1981, p. 267), in his already well-known *Prices and Quantities*, writes that "there are strong grounds for the presumption that in macro equilibrium the output of the price-tag economy is below a social optimum, and that the extra output generated by a strengthening of aggregate demand augments social welfare."

14. See Blinder and Mankiw (1984) for a more complete exposition of this issue.

15. This appendix was circulated as part of the 1982 working paper version of this article, but was not included in the version published in the *Quarterly Journal of Economics*.

16. An analogous model of monopoly unions in the labor market would imply similarly inefficient employment fluctuations and wage rigidity. Real economies are characterized by a variety of different types of wage-price stickiness, as is discussed in Blinder and Mankiw (1984).

17. This first-order condition holds with equality only if the optimum I^* is strictly inside the unit interval. Examination of the more general Kuhn-Tucker conditions does not alter the conclusions.

18. A larger monetary contraction would induce more firms to adjust their prices, but those that do not are even farther from the optimal output. The marginal effect upon aggregate output and welfare depends upon $g(\cdot)$ and is in general ambiguous.

19. Some numbers may be helpful. Let $\phi = .1$, $\theta = 1$, and $g(i) = .03i^8$. Simple calculation shows total menu costs for this economy are less than 1% of its natural rate output. A 2% monetary contraction causes a 6.9% reduction in output, since only 62% of the firms adjust their prices. The social planner would have every firm adjust its price, in which case there is no drop in output.

References

Barro, Robert J., "A Theory of Monopolistic Price Adjustment," *Review of Economic Studies*, 39 (1972), 17–26.

Barro, Robert J., and Mark Rush, "Unanticipated Money and Economic Activity," *Rational Expectations and Economic Policy*, Stanley Fischer, ed. (Chicago: University of Chicago Press