Macroeconomic Implications of Customer-Supplier and Worker-Employer Relationships

Ph.D.

Paul Sau Leung, Yip

London School of Economics and Political Science

1991

x-21-158895-3

F6963

Abstract

This thesis investigates a few examples of customer-supplier and worker-employer relationships which are thought to be important to macroeconomic analysis.

Among the literature on sticky price, we suggest that the theory of mark-up pricing and the theory of customer-supplier relationships advocated by Okun (1981), deserve more attention. We also suggest that there are at least three hypotheses implicit in the mark-up equation:

(1) a sticky pricing response to demand shocks;
(2) a relatively fast pricing response to cost shocks; and
(3) 1 % change in average cost will cause an equiproportionate rise in price.

In Chapter 2, a reputation cost of changing price is used to summarize Okun's discussion on suppliers' tendency to pledge the constancy of price for some reasonably long period (a type of customer-supplier relationship). A microfoundation model is then built to investigate the three hypothesis in details.

With regard to the first hypothesis, it is shown that (a) the reputation cost of changing price; (b) uncertainties about the persistence and generality of an observed demand shock; and (c) their interactions can jointly account for an extensive degree of price stickiness. We also explain that such a modelling of price stickiness could be more convincing than that by the Menu Cost Hypothesis. With regard to the second and third hypotheses, our conclusion is positive in the sense that it is a good approximation, but negative in the sense that it is at most an approximation. We then specify the conditions under which the mark-up equation can be used in macroeconomic analysis. In our discussion of hypothesis 2, we also touched upon the evolution of the practice of cost-oriented, as opposed to demand-oriented, pricing.

In Chapter 3, we start with the justification of an implicit, non-binding guarantee of employment for those within the firms (a type of worker-employer relationship). A dynamic programming model is then built to investigate the employment response of the representative employer to demand shocks. It is found that:

(a) In the case of mild negative demand shock, the producer will hoard the excessive amount of labour, and production effort will be the variable of adjustment;
(b) In the case of adverse negative demand shock, the producer will break the implicit guarantee of employment and make considerable amount of layoffs.

From the point of view of maintaining employment, it is always better to stimulate the economy before, rather than after, the layoffs. Mild stimulation policies after the layoffs will have no effect on employment.

Chapter 4 attempts to provide an estimate of the cost of changing price. It was found that the cost is much larger than can be explained by the Menu Cost Hypothesis. The estimation also provides some evidence against the Normal Cost Hypothesis. Finally, Chapter 5 is a simulation exercise to check whether Caplin and Spulber's neutrality result, arising from the disappearance of price stickiness on aggregation, can be applied to some more general specifications.

Acknowledgement

I would like to take this opportunity to thank my supervisor, Professor C. Bean, for his constructive suggestions during the development and progress of this thesis. Without these suggestions, the thesis would not have been completed with success. I am also very much indebted to Professor M. Morishima and Dr. T. G. Yung for their profound influence on my attitude to the subject. However, they are not in any way responsible for my way of expressions in this topic. Last, but not least, I would like to thank H.O. Chan, Y.M. Lai and Felicia Kok for their helpful comments on the presentation.

To my Father

Contents

Diagrams and Tables

Diagrams

Tables

Chapter 1

1.1 Importance of Microfoundations to Macroeconomics

Much of the debate between the Keynesians and their opponents originates from, or is related to, differences in perceptions about the flexibility (or rigidity) of prices and wages. In the models of Keynes, Keynesians and post-Keynesians, prices and wages are often assumed to be rigid (or sluggish) downward so that a reduction in nominal aggregate demand will cause a reduction in output and employment. As quantity instead of price is assumed to play the (major) role of adjustment, excessive levels of unemployment and excessive fluctuations in real activity can result. This school of thought therefore proposes a role for the Government in fine-tuning the economy. On the other hand, the Classical and Neo-classical models usually assume flexible prices and wages so that markets usually clear with full employment. Involuntary unemployment can at best be short-lived and transitory. The New Classical models, with the same assumptions of flexible prices and wages and instantaneous market clearing, holds that fully recognized shifts in aggregate demand cannot cause fluctuations in real activities. The Government can only influence real activities by "fooling" individuals about relative prices. However, this kind of "fooling" cannot be long lasting because rational agents will soon learn about this and the Government policy will become fully recognized. Accordingly, there is a natural rate of unemployment and there is no room for systematic counter-cyclical monetary policy.

As we can see, the crucial point lies in the flexibility of prices and wages. If prices and wages are fully flexible, the Classical view of the world should be preferred. If any one of the prices and wages is rigid downward, the Keynesian viewpoint becomes relevant. Thus, an understanding of the microfoundations for prices, output, wages and employment decisions is the key to the macroeconomic debate. The major aim of this thesis is to review and extend previous work on this topic.

It is interesting to note why the traditional Keynesians and the three Classical schools insist on their own assumptions and reject the others on methodological grounds. The traditional Keynesians think that wages (and sometimes prices) are structurally rigid because this is what they observe in this world. The Classical schools, who believe in the maximizing principle and the power of the invisible hand, think that prices and wages will always be adjusted to achieve full-employment equilibrium. As a result, traditional Keynesians are often criticized by their opponents for the lack of sufficient theoretical foundation for the assumption of rigid prices and wages, while the Classical schools are criticized by the Keynesians as unrealistic in neglecting the observed stickiness of prices and wages. Subsequent work on the microfoundations of sticky prices attempts to reconcile this by suggesting that the short run Keynesian stickiness of prices (or more correctly, stickiness of prices within bounds) can originate from long run Neo-Classical maximizing behaviour when there is a (menu and reputation) cost of changing price. The theory of implicit contracts also attempts to explain real wage rigidity by suggesting that an implicit guarantee of stable real wages will be beneficial to both the more risk-averse employees and the less risk-averse employers. The unsatisfactory fact is that there does not seem to be any apparent correspondence of methodology between the two separate sets of theories in the product and labour markets. Neither does the real wage rigidity of the implicit contract theory look comparable to the nominal stickiness of price in the product market. However, in the later part of the review, we will first argue that the implicit wage guarantee will be quoted in nominal terms with the understanding that the nominal wage will be occasionally adjusted to keep the variations of the real wage within some narrow bounds. We then argue that the breaking (or underfulfillment) of an implicit guarantee of employment and wage will imply a cost of layoff and a cost of cutting the wage (or a cost of having the wage below the agreed norm) which correspond to the cost of changing price in the product market. This enables us to build a general model that is applicable to the price, output, wage and employment decisions. Besides, effort is made to combine the "good" elements and discard (or modify) the "bad" elements from the previous work on microfoundations to give a consistent framework which suit the observed facts.

Thus, excess capacity, cost-oriented pricing, sticky price with respect to demand shocks, long run Neoclassical maximization, customer-supplier relationships, employer-worker relationships, wage rigidity, involuntary unemployment, procyclical productivity, labour hoarding, cyclical variations of output and employment, implicit guarantee of employment and wages can all be made consistent with each other after refinement. Of course, the existence of the "bad" elements implies that some previous works have to be refined. For example, in the theory of implicit contracts, the prediction of real wage rigidity is discarded, and replaced by nominal wage rigidity with some limited fluctuations in the real wage. However, the spirit of implicit contract theory is conserved and used to back up the model in Chapter 3.

The aim of the thesis is not just to provide a justification of the Keynesian approach, it also aims to clear some of the misconceptions within the Keynesian framework. For example, in the macroeconomic debate, the Keynesian model is often interpreted as a theory of quantity adjustment and the Neoclassical school as a theory of price adjustment. Accordingly, if one variable does not adjust (or does not make the full adjustment), the other variable must, by definition, carry the full (or remaining) burden of adjustment:

> *"changes in the nominal aggregate demand for goods and services have been accompanied by only a partial response of aggregate price level. Because prices do not carry the full burden of adjustment in the short run, quantities must by definition carry part of the load."[Gordon(1981)]*

This type of dichotomy runs the danger of rejecting the possibility of a third variable of adjustment. A good example is related to the labour market. As explained in the later review and Chapter 3, the employer's implicit guarantee of wage and employment implies some stickiness of wages and employment. Thus, neither the wage (the price variable) nor employment (the quantity variable) will adjust to moderate demand shocks. Instead, we propose a third variable -- production effort -- as the main variable of adjustment with respect to moderate or transitory demand shocks. Thus, neither the Keynesian, nor the three Classical schools, give the right picture about wage and employment response to moderate demand shocks. The trouble with the three Classical schools is that they assume both wages and employment are free to adjust. Traditional Keynesian models, while attempting to change the macroeconomic properties by introducing a wage floor, fall into the same trap as the three Classical schools in assuming employment is a free variable. Indeed, this is why the traditional Keynesian models give the embarrassing prediction of counter-cyclical productivity. In Chapter 3, we will develop a model in which employment, due to a cost of layoff, will be sticky with respect to a moderate reduction of demand. Such a result suggests that productivity per head will be procyclical, a phenomenon that is in accordance with observation and empirical findings.

Microfoundations are important because different definitions of stickiness may have different macroeconomic implications. For example, McCallum (1977) claims to have developed a model which preserves the Lucas-Sargent proposition of the infeasibility of counter-cyclical monetary policy even if prices are sticky. If McCallum is right, this will relieve the New Classical school from the criticism of paying insufficient care to the observed stickiness of price. However, a careful look reveals that the definition of "sticky price" in McCallum's paper is quite different from the usual definition. In the model of McCallum, stickiness of price is related to the gap between the market clearing level ($P_t^{\#}$) and the anticipated general price level (P_t^*):

$$P_t = \begin{cases} P_t^* & \text{if } P_t^* - \delta_1 < P_t^{\#} < P_t^* + \delta_2 \\ P_t^{\#} & \text{otherwise} \end{cases} \qquad (1.1)$$

where δ_1, δ_2 are both greater than zero.

If the market clearing price $P_t^{\#}$ is not too far away from the anticipated price P_t^* (ie $P_t^{\#}$ lies within $P_t^*-\delta_1$ and $P_t^*+\delta_2$) it is not worth the cost to have the actual price different from the anticipated level P_t^*. On the other hand, if the gap $P_t^{\#}-P_t^*$ is sufficiently large in magnitude (ie lies outside the range of $[\delta_1,\delta_2]$), it is worth the cost of adjusting the price P_t to the market clearing level $P_t^{\#}$. Nevertheless, there does not seem to be any sound justification for McCallum's specification. McCallum did (informally) attempt to use Barro (1972) as the basis for the specification. However, as admitted by McCallum himself, the result of Barro (1972) would imply P_{t-1} instead of P_t^* will appear in the above specification. As will be explained in the thesis, we hold the same view as Barro that P_{t-1} instead of P_t^* should be the reference point of price stickiness. If P_t^* of equation (1.1) is replaced by P_{t-1}. McCallum will not be able to derive his result and price stickiness will be inconsistent with the Lucas-Sargent proposition [1].

---

[1]To illustrate the inconsistency in McCallum's model, suppose the economy is at full employment at t-1. Let there be an anticipated reduction of aggregate demand at t which is small relative to the cost of changing price so that prices remain sticky. With Barro's definition of stickiness, we have $P_t = P_{t-1}$. Adding the assumption of rational expectation, we have $P_t^* = P_t = P_{t-1}$. Substituting into equation (1) of McCallum's paper implies that the notional aggregate supply is still at the full employment level

ie $\quad y_t = k_t + a_1 (P_t - P_t^*) + u_{1t} \quad$ and $P_t^* = P_t = P_{t-1}$

$\Rightarrow \quad y_t = k_t + a_1 (P_t - P_{t-1}) + u_{1t}$

However, as price is now fixed at P_{t-1} and no longer fully flexible to achieve market clearing, the reduction of aggregate demand implies that the effective aggregate supply will be below the full employment level. There is thus a role for stabilization policy and the Lucas-Sargent proposition is no longer consistent with the proper definition of price stickiness.

In addition to the stickiness of prices and wages, we would like to emphasize the stickiness of relationships between the customers and suppliers; and between employers and employees. The details of such relationships have already been analyzed by Okun (1981). What we want to emphasize is the macroeconomic implications of these relationships. One of these is related to the recent emphasis of hysteresis effect in international trade theory [see Baldwin and Krugman (1986), and Bean (1987)]. The theory, in sharp contrast to the conventional economic wisdom, holds that large fluctuations of the exchange rate will lead to entry or exit decisions, thus leading to the breaking or making of customer-supplier relationships that are not reversed when the currency returns to its previous level. Thus, the extreme strength of the US dollar in the early 1980s has caused a permanent loss of market position by US-based firms. This is so because once foreign firms have invested in marketing, research and development, reputation, distribution networks, etc., they will find it profitable to remain in the market even at a lower exchange rate. Once US firms have abandoned markets, a mere return of the exchange rate to former levels will not be enough to make the expensive recapture of these markets worthwhile.

Another example is related to the "stickiness" of the employer-employee relationship which will be analyzed in Chapter 3. In the model, we hold that employers will keep the same amount of employees within a certain range of demand. If demand falls outside this range, employers will make considerable amount of layoffs. Thus, in case of recession it is always better, from the point of view of aggregate employment, to stimulate the economy before rather than after employers make the layoffs. This is so because once these employers, who start with labour hoarding, have cleared the excessive labour through layoffs, they will not re-hire this excessive labour even if demand returns to its original level. On the other hand, if the Government succeeds in stimulating demand before employers making the layoffs, labour that would otherwise have been laid off will remain employed.

1.2 Review of the Literature

1.2.1 Price and Output Decisions of Firms

Traditional supply and demand analysis predicts that fluctuations of demand (and supply) will cause fluctuations in prices. However, as Okun(1981) has suggested, such kinds of fluctuations are limited to selected financial assets, agricultural products and primary metals. Observation tells us that the prices of most manufacturing and service outputs are rather sticky (or sluggish) to demand shocks. Given that manufactured and service output forms the major category of production value in our economy, the question of whether prices are sticky with respect to demand shocks is important. If prices are sticky with respect to demand shocks, quantity rather than price will bear the burden of adjustment. Cyclical fluctuations of output will be a usual phenomenon and there can be a role for stabilization policy even if expectations are rational. For this reason, we will first review some of the literature on sticky prices.

Four strands of thought are of particular relevance to our discussion of the price and output decisions of the firm. The first strand is the theory of mark-up pricing which provides a group of important hypotheses that we would like to either confirm or challenge. The second strand consists of theoretical models such as Barro(1972), Mussa(1981) and Rotemberg(1982), which illustrate that sticky prices can indeed result from the rational behaviour of the firm. While there are quite a lot of limitations associated with these models, they are contain some basic features, such as the methodology of Neoclassical maximization subjected to a cost of changing price, in common with our model. The third strand is due to the very persuasive discussion by Okun(1981) which provides the conceptual foundation of the customer-supplier relationships and the formulation of the cost of changing price in our model. Most of the features discussed in Okun's book, such as the shopping process, are taken to be complementary to the model we build in Chapter 2. The final strand is the current research work on near-rationality by Akerlof and Yellen (1985a,b), Mankiw (1985) and Blanchard and Kiyotaki (1987). In this part of the review, we hope to point out some of the inappropriately overemphasized features of this literature.

(A) Mark-up Pricing

In 1939, Hall and Hitch published an empirical paper casting doubt on the Neoclassical short run marginal cost pricing behaviour. Instead, they suggest that businessmen might be following a rule of thumb, which they called the "full cost principle", in their pricing decision process. They also suggest that once prices are fixed,

> *"[they] will be changed if there is a significant change in wage or raw material costs, but not in response to moderate or temporary shifts in demand."*

Such an assertion represents an important departure form the Classical Supply-Demand Analysis and the Neoclassical framework which assume that prices will be fully flexible to achieve equilibrium. As another contribution, Hall and Hitch suggest that the pricing response with respect to cost shocks is qualitatively different from that with respect to demand shocks. Unfortunately, Hall and Hitch do not appear to recognize that the full cost principle can be derived from long run profit maximizing behaviour.

The work by Hall and Hitch was followed by a series of papers which brought both conceptual innovations and a much better understanding of the microfoundations of the Keynesian approach. Instead of going through the details of the debate, we will merely summarize the discussions by Koutsoyiannis (1979) and Morishima (1984) as follows :

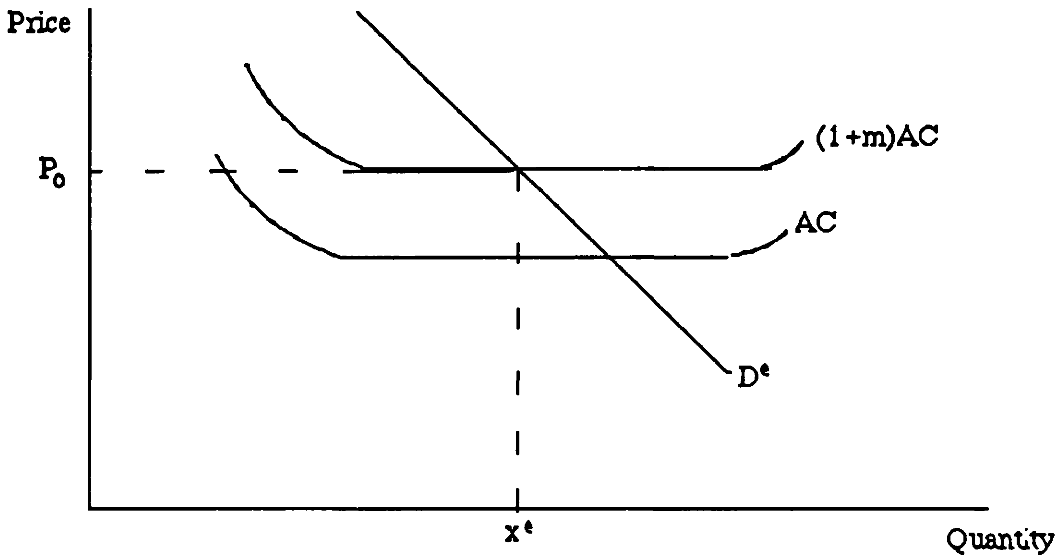
(a) When a representative producer starts a business, he will try to have a rough estimate of the demand (D^e in diagram (1.1));

(b) He then builds the plant (or service lines) with a flat-bottomed average cost curve and some kind of planned excess capacity. The reasons for the flat-bottomed AC curve and planned excess capacity, according to Koutsoyiannis are (i) to meet seasonal fluctuations of demand; (ii) to allow a smooth flow of production when break-down of some equipment occurs; (iii) to meet a growing demand until further expansion of scale is realized; and (iv) to allow for some flexibility for minor alterations of style of product in view of the changing tastes of customers. Excess capacity, according to Koutsoyiannis, also exists on the organizational and administrative level.

(c) The producer will charge the price (P) as a mark-up (m) of the average variable cost (AC) [2] so that

$$P = (1+m)\,AC \qquad (1.2)$$

In terms of Morishima's diagram,



where D^e is the expected demand curve;
AC is the average cost curve; and
m is the producer's ex ante profit mark-up plus a certain percentage associated with the ex ante average fixed cost.

diagram (1.1)

[2]Within the theory of mark up pricing, there is a debate between proponents of total mark up (full cost principle) and proponents of variable mark up [eg. Kalecki(1939)]. The debate is centered around whether AC should be interpreted as the average total cost or average variable cost. To keep the fluency of presentation, we will only go into the details of such debate in Chapter 4.

the producer will charge the price P_0 where the expected demand D^e cuts the curve (1+m)AC.

(d) Once the price is set, it will remain sticky even though actual demand turns out to be different from expected sales. The producer is reluctant to change price to the "short run" maximizing level because of the worry of losing goodwill which may seriously damage his sales in the future. Instead, the representative producer will adjust his production with respect to the demand shock. If the demand shock is persistent and larger than that allowed by the "planned excess capacity", the producer may still prefer expanding the capacity to raising price.

(e) Prices are more likely to change with cost shocks and the percentage change in price will be approximately equal to the percentage change in average cost.

Unlike Hall and Hitch, Koutsoyiannis (1979) and Morishima (1984) have recognized that the stickiness of prices with respect to demand shocks under the full cost principle may indeed be a result of long-run profit maximizing behaviour. Unfortunately, no explicit formulation about the cost and benefit of changing/keeping the price under these circumstances exists. Barro(1972), Mussa(1981) and Rotemberg(1982) [abbreviated as the B-M-R models] attempt instead to explain the stickiness by introducing an explicit cost of changing price.

(B) Price Stickiness with an explicit cost of changing price (the B-M-R models)

(1) Barro (1972)

The analytical tool used here is due to Miller and Orr(1966). Barro assumes that

(i) there is a fixed cost (A) for every change in price; and

(ii) the producer seeks to maximize profit

$$\pi = PY - C(Y) \qquad C'(Y) > 0$$

subject to the demand constraint

$$Y = Y^d = Q(P) + u \qquad Q'(P) < 0$$

where π is the profit, Y is the output, P is the price, Y^d is the quantity demand, C(Y) is the cost function, Q(P) is the normal demand function, and u is a stochastic component of demand which is assumed to follow a random walk.

After arguing that the "two-bins" policy will be the optimal solution of this problem,

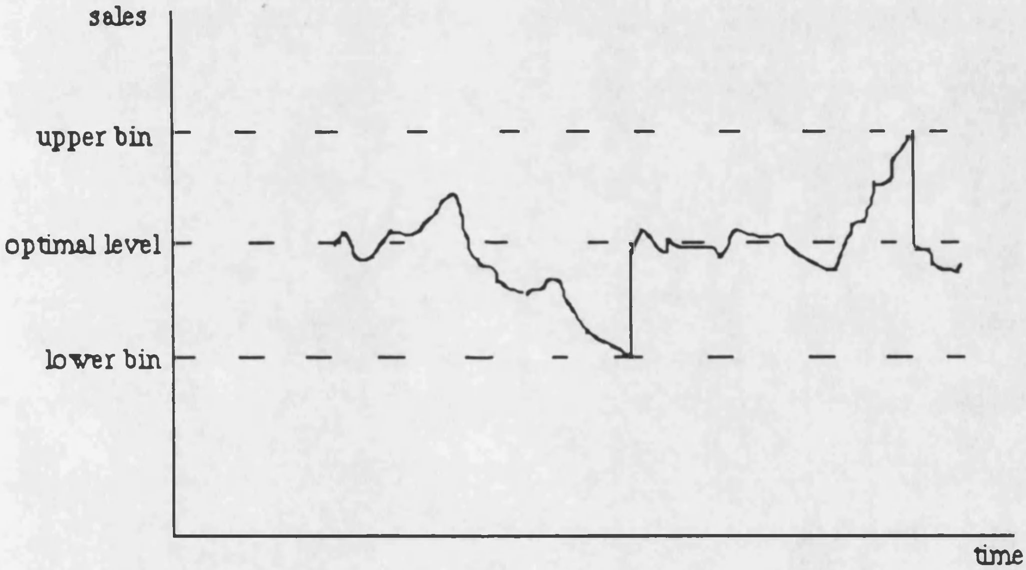


diagram (1.2)

Barro then proceeds to find the optimal floor and ceiling within which the producer should not change the price and beyond which the producer should adjust the price so as to make the sales return to the "optimal level". The contribution of this exercise is that it explains price stickiness with respect to stochastic fluctuation of demand when there is a cost of changing price. The trouble with this stochastic analysis is that it tells us nothing about the optimal pricing response in case of a permanent demand shock. If fluctuations of sales is a mixing result of permanent shifts and random fluctuations of demand and the producer cannot perfectly distinguish the two disturbances, the two bins policy may even fail to be the optimal solution of the problem.

(2) Mussa(1981)

By assuming

(i) the equilibrium price (P_t*) is growing at a constant rate;

(ii) the loss from not adjusting the price (P_t) to P_t* is a quadratic function of (ln P_t - ln P_t*); and

(iii) a fixed cost incurred with every change in price;

Mussa was able to guess that the optimal solution would be one with a periodic stepwise change of price for every fixed period T (he then proceed to solve for the optimal value of T). Thus, the optimal price path (ln P_t) will be something as shown in diagram (1.3) :

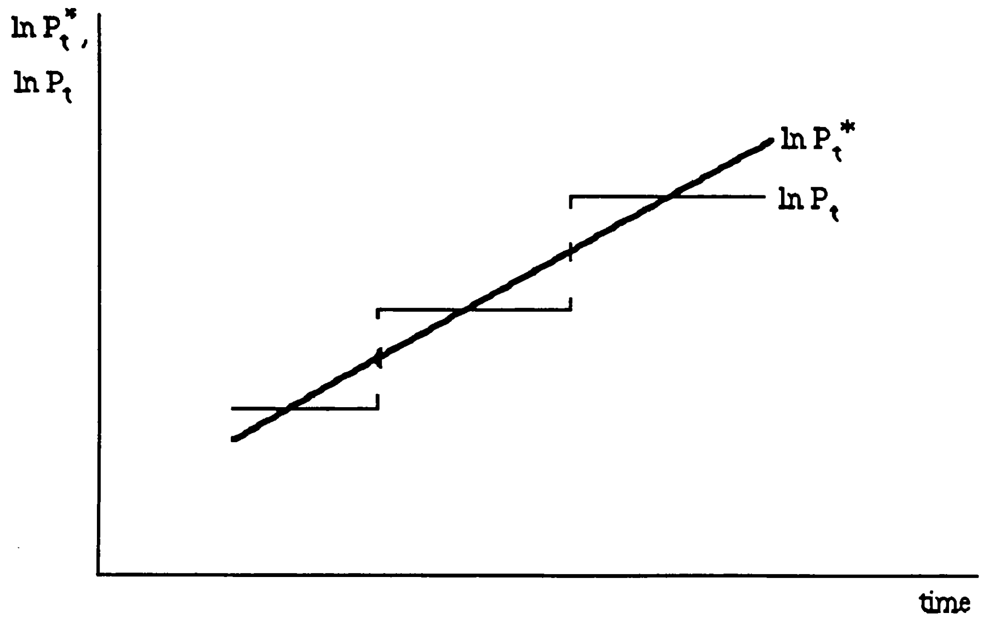


diagram (1.3)

where price will remain sticky in the horizontal section of ln P_t (and make a discrete jump in the vertical section). It should be noted that while Barro(1972) attempts to explain price stickiness in a world with purely random fluctuation of demand around a fixed normal level, Mussa(1981) attempts to explain price stickiness in a world with a constantly rising mean equilibrium price (eg. rising cost). Thus, the work by Barro and Mussa can be regarded as complementary.

While Mussa's model is most suitable in explaining price stickiness in a world with constantly rising cost, it does not explain price sluggishness with respect to a large permanent demand shock which will cause a once-and-for-all rise in P_t^*. If the once-and-for-all rise in P_t^* is large enough to overcome the cost of changing price, Mussa's model will predict an immediate rise of P_t to P_t^* and there will not be any price sluggishness.

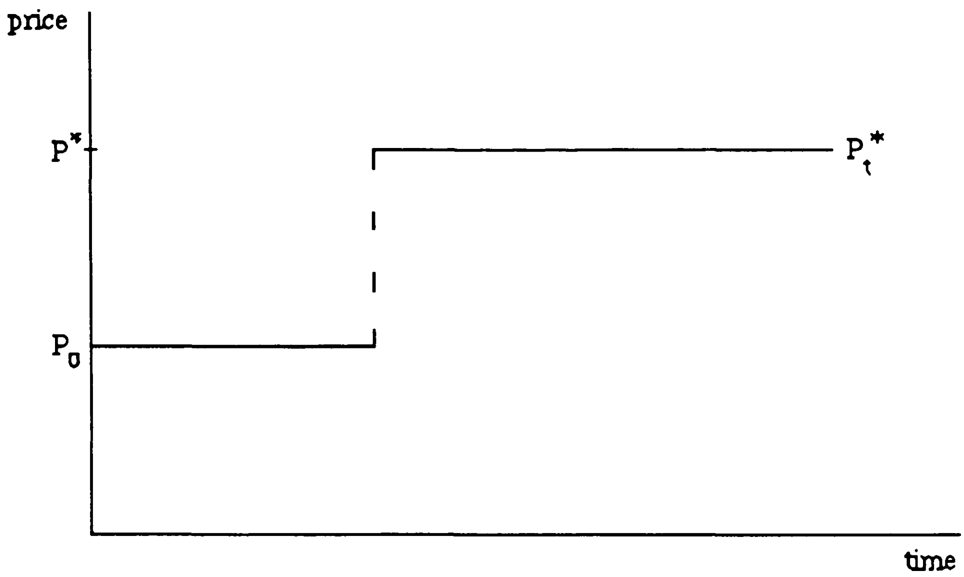


diagram (1.4)

However, as we will show in Chapter 2, price will still be sluggish in the face of such an event whenever there is a signal extraction problem between a persistent and a transitory shock. This greatly enhances the extent of price sluggishness that can be explained by the recent literature.

(3) Rotemberg (1982 a,b)

With the assumption of a quadratic cost of changing price and some other simplifying assumptions, Rotemberg shows that the optimal pricing behaviour of his representative producer will be

$$P_t = \alpha P_{t-1} + \frac{1}{\beta\rho c} \sum_{j=0}^{\infty} \left(\frac{1}{\beta}\right)^j P_{t/t+j}^* \qquad (1.3)$$

where α,β,ρ,c are some constants specific to the firm with
$0 < \alpha < 1$; $\beta > 1$; $(1-\alpha)(1-\beta) = -1/\rho c$; and
$P_{t/t+j}^*$ is the equilibrium price at t+j expected at t.

The presence of αP_{t-1} in (1.3), arise from the assumption of a quadratic cost of changing price, implies an intertemporal linkage in the pricing decision between different periods (i.e. the price set this period will affect the optimal level of the price chosen next period). The presence of αP_{t-1} in (1.3) also implies a sluggish pricing response with changes in the expected equilibrium price ($P_{t/t+j}^*$). For example, suppose there is a permanent shock in sales which changes the expected equilibrium price $P_{t/t+j}^*$ from P_o to P^*, (1.3) can be reduced to

$$P_t = \alpha P_{t-1} + (1-\alpha)P^* \qquad (1.4)$$

Approximating the discrete time model by a continuous one, we have diagram

(1.5):

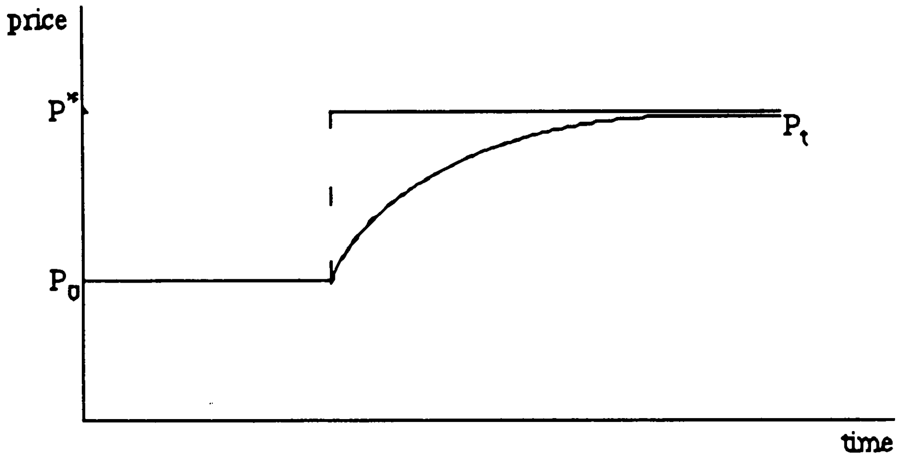


diagram (1.5)

in which P_t only adjusts gradually (sluggishly) towards the new equilibrium level. Thus, Rotemberg's paper can be regarded as a supplement of Barro(1972) and Mussa(1981) in suggesting a sluggish pricing response with respect to a permanent demand shock. It should be noted that the sole source of sluggishness here is the quadratic cost of changing price which naturally implies a gradual adjustment in all types of discrepancy between the actual and equilibrium price. Nevertheless, the overemphasis of a quadratic cost of changing price implies that the effects of all $P_{t/t+j}^*$ on P_t (in equation (1.3)) do lead to some unconvincing predictions. For example, suppose the producer perfectly anticipates a disturbance in $P_{t/t+j}^*$ as that shown in diagram (1.6):

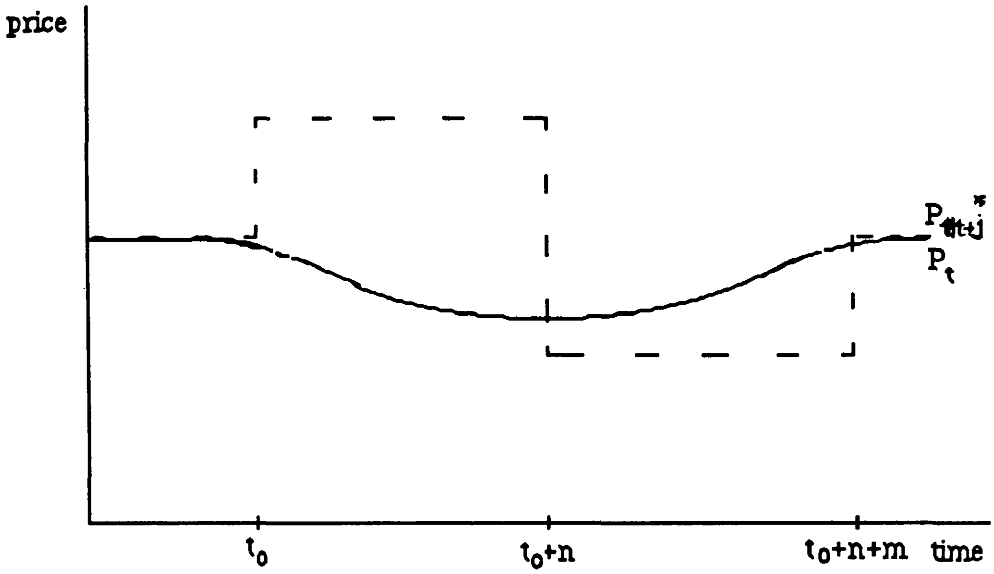


diagram (1.6)

The optimal price path P_t, according to equation (1.3), will be as shown in the above diagram. This would mean a reduction of price between t_0 and t_0+n during which demand is higher than usual. Such prediction does not appear to be what we observe in the real world. In Chapter 2, we will develop a model which will predict a more convincing pricing response than that of Rotemberg. It will also be clear that the effects of $P_{t/t+j}^*$ on P_t are not perfectly additive.

(C) Further comments on the B-M-R models and the theory of mark up pricing

(1) The B-M-R models versus the theory of mark-up pricing

The contribution of the B-M-R models is that they illustrate how sluggish price adjustment can result from profit maximizing behaviour. In particular, Barro(1972) can be regarded as an important work in the refinement of the full-cost principle. Instead of interpreting the planned profit mark-up (m) of equation (1.2) as a "fair" mark-up the producer wants to stick with, Barro's work implies that m can be interpreted as the "long-run profit maximizing" mark-up with which the producer would like to stick with even in face of random fluctuations of demand [3].

The crucial assumption in the B-M-R models is the presence of a cost of changing price. As the derivation does not require any assumption of "planned excess capacity" or a "flat-bottomed average cost curve", the result of a sticky price can also be consistent with many models other than the theory of mark-up pricing. This greatly enhances their relevance and applicability in the macroeconomic literature. However, seen from another point of view, one should recognize that the theory of mark-up pricing is a far more general and influential theory than the B-M-R models. Beside being a theory of price decision, the modern version of mark-up pricing is also, as we have seen, a theory involving capacity, production and quality decisions. The appearance of the B-M-R models only implies further development along this line should be done about the joint decisions of price, capacity, production and quality to explain the observed behaviour of firms over these

---

[3]The interpretation of m as the "long-run profit maximizing" mark up (or more correctly, the expected net discounted profit maximizing mark up) has the advantage in explaining why the profit mark ups of some growing industries are so high while the profit mark ups of some declining industries can be close to zero or negative. On the other hand, the interpretation of m as a "fair" mark up can at most explain the difference in mark up by the difference in risk premium. Such interpretation has difficulty in explaining why some profit mark ups can be close to zero (or negative) and why the difference in mark ups can be so high across industries.

areas of decisions. Even if one were only interested in the price decision, the theory of mark-up pricing does say quite a few things not recognized or properly tackled by the B-M-R models. Implicit in the formula of P=(1+m)AC, there are indeed three hypotheses

(i) the assumption of a fixed m which implies a sticky price with respect to demand shocks;

(ii) the separation of AC from m reveals the belief of an asymmetric pricing response with respect to cost shocks and demand shocks;

(iii) the unitary power index of AC implies that 1% rise of average variable cost will cause 1% rise of price (i.e. average cost pricing is superior than the short run Neoclassical marginal cost pricing).

The B-M-R models have at most given an imperfect account of the first hypothesis. In particular, by assuming a certain equilibrium price, Mussa(1981) and Rotemberg(1982) did not distinguish whether the change in equilibrium price is due to cost changes, demand changes or a mixture of the two. Such a distinction is important because (i) the cost of changing price with respect to demand shocks may be different from that with respect to cost shocks; and (ii) the representative producer's expectation of his competitor's pricing response with respect to a cost shock may be different from that of a demand shock. In Chapter 2, we will use these to explain why the pricing response with respect to cost shocks will be different from that with respect to demand shocks.

(2) <u>Signal Extraction Problem as an important source of price sluggishness</u>

Returning to the first hypothesis, we would like to emphasize that the degree of price stickiness explained by the B-M-R models is rather limited. In Chapter 2, we will emphasize the signal extraction problem faced by the firm in distinguishing between persistent and transitory demand shocks as an important source of price sluggishness. The logic is as follows :

Suppose there is an observed rise in sales at t-1. The producer, with imperfect information will not be sure whether the sales at t-1 will persist in the future or not. Instead of raising the price immediately to the "short-run profit maximizing level" recommended by the sales at t-1, the producer will

(i) in the presence of a fixed cost of changing price, prefer to wait for more observations of sales. Only if he is relatively sure that the demand shock is persistent, will he start raising the price; or

(ii) in the absence of any cost of changing price, guess the probability that the shock at t-1 is persistent and raise the price to the level recommended by this probability. If the shock at t-1 is permanent, the probability (and hence the price) will be revised upward with more and more observations of sales.

Unlike the B-M-R models, the signal extraction hypothesis suggests that the usual reputation and menu cost of changing price is not a necessary condition for (though it remains an important source of) price sluggishness. Without such cost of changing price, price will still be sluggish to permanent demand shocks as long as there is a signal extraction problem. [Another source of sluggishness that does not require the assumption of a reputation and menu costs of changing price is the presence of information transmission lags within the economy. However, this is beyond the scope of this thesis.]

Of even greater importance, the signal extraction hypothesis, when combined with reputation and menu costs of changing price, can help to extend the degree of price stickiness that can be explained by the recent literature. The argument can be separated into two stages. The first stage arises from our dissatisfaction with the formulation of the cost of changing price in the B-M-R model. In Barro(1972) and Mussa(1981), the cost of changing price is assumed to be a fixed component, A. Rotemberg(1982), on the other hand, assumed a quadratic cost of changing price (i.e. $c(\Delta P)^2$ where c is a constant). The trouble with these two types of assumptions is that they imply reducing price is as costly as raising price. If one believes that a permanent reduction of price is favourable to the reputation of firm, and the reputation cost is more important than the menu cost, the above specifications will be wrong. In Chapter 2, we will argue that a more convincing formulation should be the sum of

(i) a fixed component A [4] for any change of price. [This is intended to capture the menu cost and part of the reputation cost. It has the effect of penalizing too frequent changes in price.];

(ii) a significant linear component $\xi(\Delta P)$ [4] which is intended to capture most of the reputation cost/gain of changing price. [The advantage of having a significant linear component is that it implies a reduction of price will represent a reputation gain (at least no reputation cost) to the firm.]; and possibly

(iii) a non-linear component, such as that in the Rotemberg model,which will penalize too sharp a change in price.

The presence of a significant linear component implies that the fixed cost or quadratic cost in the B-M-R models should be much smaller than we are used to expect. This, in turn, reduces the degree of price stickiness that can explained by the B-M-R models (see Chapter 2 for the details).

The second stage of our argument concerns how the signal extraction problem can raise the degree of price stickiness. As we will show with the model of Chapter 2, a large enough permanent demand shock, in the absence of signal extraction problem, will cause an immediate jump in price form P_0 to P^* in diagram (1.7). With the presence of a signal extraction problem, but no cost of changing price, the price will be adjusted gradually towards P^* along the a path CD. The addition of a fixed cost of changing price will raise the degree of price stickiness and the optimal pricing response will be that shown by the

---

[4]It is possible that the values A and ξ of a rise in price will be different from that of a reduction in price.

stepwise path AB.

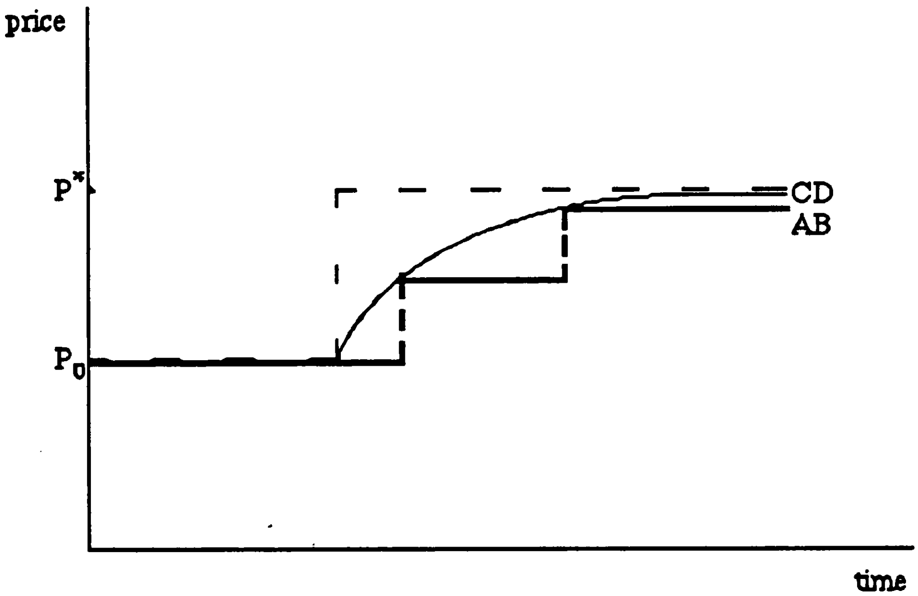


diagram (1.7)

The interesting thing is that, without the signal extraction problem, a fixed cost of changing price alone will not cause any price stickiness with respect to the demand shock. The presence of a signal extraction problem, on the other hand, not only causes sluggishness as shown by the gap between P* and CD, but also allows the fixed cost of changing price to cause an additional stickiness as shown by the gap between AB and CD. [In Chapter 2, it will also become apparent how the signal extraction problem can raise the price stickiness with respect to large enough transitory demand shock.] That is why we claim that the signal extraction problem helps to extend the degree of price stickiness that can be explained by the recent literature.

(D) Customer-supplier relationships

While Barro, Mussa and Rotemberg were, in some degree, able to explain price sluggishness with the assumption of a cost of changing price, their discussion about the nature of the cost remained shallow. Barro(1972) and Mussa(1981) simply assumed the cost is fixed and made no explanation for the assumption. Rotemberg(1982a,b) only said that the cost includes the physical cost of changing the posted price (ie. the menu cost) and the reputation cost. Okun(1981), however, provides a much more extensive discussion of the cost by emphasizing the importance of repeated purchases in his category of customer market in which products are sold with price tags set by the seller. The idea itself originates in the search theory and implicit contract literatures.

(1) Okun's explanation of sticky price and welfare implication of the relationship

Okun first assumes that there is a shopping cost and limited information about the location of the lowest price in the market place so that buyers do not find it worthwhile to incur all the costs required to find the seller offering the lowest price. Instead, customers will adopt the strategy of setting an acceptance price, being ready to settle for any price at which the additional cost of more shopping outweighs its benefit. Okun then explained that customers are valuable to suppliers because of their potential for repeat business. Thus, suppliers have the incentive to discourage customers from shopping elsewhere by pledging continuity of firms's policy of price, services and the like. In other words, suppliers want to promote and condition customer's reliance on intertemporal comparison shopping. On the other hand, customers are attracted by continuity because it helps to minimize shopping costs. Such action in turn encourages and justifies the seller to maintain a stable pricing policy and willingness to accept greater variations in quantity. Such an account of customer relationship not only explains the stickiness of price, but also predicts the stability of quality, service and the like.

Okun also distinguishes his customer relationship from the theory of administered prices [Berle and Means(1932), Means(1935,1939) and Blair(1974)] as an alternative explanation of sticky prices [5]. In sharp contrast to the theory of administered prices, the customer relationship model implies that inflexibility of price may have some socially desirable aspects. Although there is a social welfare loss associated with the persistent excess of price over marginal cost, the relationship does significantly reduce the transaction cost :

> *"The customer-market attachments save a huge volume of resources that customers would otherwise devote to shopping (and trying out) products with every transaction. To firms, an established clientele increases predictabilities of sales, permitting important savings in inventory costs and production scheduling [Okun(1981),P.155]".*

No matter whether the benefit falls short or outweighs the loss, the relationship suggests "diagnosis does not point to antitrust measures [as implied by the theory of administered price] as a likely remedy for chronic inflation or macroeconomic stability."

(2) Scope of the relationship

(a) "Big-ticket" items

In the explanation quoted in section (1), the possibility of repeated business is the crucial reason for suppliers to maintain the relationship. Nevertheless, Okun holds that such relationship also exists in markets of "big-ticket" items that are bought infrequently by consumers :

> *"In case of big-ticket items that are bought infrequently by consumers, like automobiles and household appliances, repair services are a means of maintaining relationships. More generally, firms seek to establish brand-name reliability in a way that counts on reputation (a flow of information from one consumer to the next) to substitute for repetition of experience by*

---

[5] Okun also suggests, in the absence of customer relations, the theory of administered price itself is insufficient in explaining why price remains sticky with a rise in demand. Interested readers should refer to Okun's book.

the same consumer." [Okun(1981),P151]

Despite the continuing high cost in providing repair services and making advertisements, most of the suppliers in these markets still find it worthwhile to take the cost in promoting the brand-name reputation which will in turn raise demand significantly. As the cost of promotion is huge in absolute size, the cost of dis-investment in reputation (such as having an erratic policy in pricing and the like) will also be high. This explains why suppliers of these markets dislike changing price frequently to capture small and short term fluctuations of demand.

(b) Within the production and distribution hierarchy

While Okun's attention was focused on the shopping process (ie. the relationship between the producers and the final consumers), a customer relationship also exists between the various participants of the production and distribution hierarchy [6] (i.e. between the secondary producer(s) and the tertiary producer(s); between wholesaler(s) and retailer(s) etc.). To borrow the terminology of Schultz(1985) concerning the employer-worker relationship, there exists large returns to maintaining the continuity of association between the customer and supplier. Customers prefer the relationship because, after a reasonably long period of trading,

(i) their specific demand about the product in various quality dimension become well understood by their supplier(s) and less mistakes (which can be costly) are expected;

(ii) they know that the suppliers have indicated by their previous trading that they prefer long-term business to short-term cheating;

(iii) they know about the speed and reliability of delivery of the suppliers;

(iv) credit and discount can be obtained; and

(v) bargaining costs over price and the like are significantly reduced;

[6] As the number of participants are much smaller here, the shopping process described by Okun may not apply in some cases. Nevertheless, customer relationship still exists between the buyer(s) and seller(s).

All these may not be available, or costly to acquire, if they were to change to a new supplier.

Suppliers (or subcontractors) also prefer the relationship because, after a reasonably long period of trading,

(i) they know more about the customer's creditability and financial ability in honouring the payment;

(ii) the possibility of repeated business gives them greater guarantee of future demand and hence important savings in inventory costs and in production scheduling;

(iii) as in the case of customers, they acquire non-transferable knowledge about the customer's specific demand in various quality dimensions and hence avoid some very costly mistakes; and

(iv) bargaining costs over the price and the like is significantly reduced.

As the initiating costs and risk premiums of trying out new suppliers (or customers) here are much higher than that in the shopping process, the relationship here is necessarily more sticky. This provides an explanation for the existence of "hysteresis effect" in the international trade theory mentioned in section (1.1).

(3) Cost-oriented pricing

In the previous sections, we have seen that suppliers have incentives to maintain and pledge constancy of the pricing policy. However, because suppliers are subjected to cost increases that they cannot control, it will be impossible (and not worthwhile) to maintain the same price over an infinite horizon. They can nonetheless establish some practices designed to build the confidence of their customers in their dependability and reliability. One important practice [7], according to Okun, is cost-oriented pricing :

[7]The other cost reducing practices, according to Okun, are the promise to meet competition, fixed time scheduling and prenotification of price revision. The use of these practices in some industries, however, does not deny the applicability of cost-oriented pricing. For example, fixed time scheduling of price revision can be combined with cost-

"Firms not only behave that way, but also condition their customers to expect them to behave that way. It is easy for anyone to understand that cost increases can force the firm to break the continuity of its offer. Higher costs are an accepted rationale for raising prices .... Price increases that are based on cost increases are fair, while those based on demand increases often are viewed as unfair." [Okun(1981,P153]

In the terminology of the B-M-R models, Okun's explanation is equivalent to saying that the cost of raising price based on cost increases is much lower than that based on demand shocks. Such difference in the cost of changing price, according to Okun, explains why suppliers adjust price more promptly and more reliably in response to changes in cost. Unfortunately, no explicit model of this argument has been constructed. This creates some ambiguities. For example, the verbal analysis does not make it clear whether the cost refers to full cost or variable cost. Neither does it make it clear whether the price should be changed by the same absolute amount, or by the same percentage, of the change in cost [8]. Besides, the competitor's pricing response does not appear to enter the above explanation. If competitors were to maintain their prices in the face of cost shocks, it is doubtful whether a representative supplier could persuade his customer to perceive his cost-oriented pricing as natural and "fair". To explain the evolution of cost-oriented pricing as a common practice, one also has to explain why all, or at least the majority, of the suppliers

(i) "agree" to raise price with respect to moderate cost shocks; and

(ii) do not "agree" to raise price with respect to moderate demand shocks.

To tackle these problems, an explicit model incorporating the competitor's pricing response will be built in Chapter 2. Two additional reasons, along with the difference in cost of changing price, were found to explain the observed asymmetric pricing response to cost

---

oriented pricing to give a loose version of cost-oriented pricing such as that mentioned in the last paragraph of the section.

[8]Okun did mention that the truth lies somewhere in the middle. It should nevertheless, be noted that no theoretical explanation has been given to his belief.

shocks and demand shocks. That is, a cost shock is likely to be more general and persistent than that of a demand shock. The former makes all, or most, of the suppliers expect (and expect each other to expect) the other to be affected by the cost shock and raise the price sooner or later. This will not happen to a demand shock as there is great uncertainty (specifically a signal extraction problem) as to whether an observed demand change is due to a general or specific demand shock. The latter makes the gain in discounted stream of profit from a prompt revision of price to the "optimal" level much higher than that of transitory (demand) shock, and hence make it more likely that the gain from changing price (to the "optimal" level) to be greater than the cost.

Indeed, it is interesting to ask why cost-oriented pricing seems to be widespread while demand-oriented pricing seems to be relatively rare in reality. In Okun's description, price increases based on cost increases are considered (by the customers) to be "fair" while those based on demand increases are viewed as "unfair", and hence the cost of raising price with respect to cost shocks will be much lower than that with respect to demand shocks. But why? If demand-oriented pricing instead of cost-oriented pricing had come to be the usual practice of this world, price increases based on demand increases would be considered as "usual" (if not "fair"), and the cost of raising price with respect to cost shocks will be much lower than that with respect to demand shocks. Hence, we propose that the difference in cost of changing price is unlikely to be the original reason for the evolution of cost-oriented pricing [9]. Nonetheless, once the practice of cost-oriented pricing is established for some reasons, it does imply that the cost of raising price with respect to cost shocks will be lower. This in turn becomes an additional justification (though not the original driving force) for the cost-oriented pricing.

[9] Instead, the worry of potential entry with price increases w.r.t. demand shocks (not a problem in case of cost shocks); the generality and persistence of cost shocks; and (if one starts to consider the continuing inflation in our world) the necessity to raise price sooner or later with the gradually rising cost are some of the potential explanations for the evolution of cost-oriented pricing.

Last but not least, it should be noted that cost-oriented pricing is only an approximation to the observed behaviour of most suppliers. There are at least two reasons for this. Firstly, a straight pursuit of cost-oriented pricing by suppliers requires the price to be adjusted with every change (no matter how small and frequent) in costs. As mentioned earlier, suppliers need to maintain and pledge constancy of price for a reasonably long period so as to encourage customers to rely on inter-temporal comparison and return to shop. Such need to maintain the price for some period implies that it is not worth the supplier adjusting price in response to very small change in costs. Instead, a change of price (by the same percentage change in costs) will only occur when the changes in costs have accumulated to some level or when time has come to the revision period [10]. Secondly, as mentioned by Okun, customers have little information about the actual change in costs:

> *"cost-oriented pricing is a deficient price standard because its operation is not readily observable by buyers. They cannot monitor the firm's diligence in carrying out an implicit contract that links price to costs"[Okun(1981), P154]*

Given such asymmetric information, suppliers have incentives to change the price, in addition to the same percentage rise in costs, by another amount that reflects the perceived long term change in demand. It appears to the writer that many suppliers do behave like that. Thus cost-oriented pricing is only, on the average, a good approximation to observed behaviour.

(4) Implications for the specification of the cost of changing price: distinction between different types of price changes

One important contribution of Okun's discussion of customer-supplier relationships is that it emphasizes the additional loss of present and future business that

[10] If the supplier has set up the practice of fixed time scheduling of price revision.

arises from customer's adverse response as the main source of price adjustment cost[11]. According to Okun, frequent (and irregular) changes of price will make it difficult for customers to guess the present price from that of last purchase, and hence destroy their incentive to return to shop. Such an explanation provides a deeper foundation of the cost of changing price than that of the B-M-R models. In Chapter 2, we will use Okun's idea as a guideline for a refined specification of the cost. Here, we will use Okun's idea to distinguish and compare the cost of some different types of price changes [12].

(a) Regular Price Changes Versus Irregular Price Changes

Regular price changes should be distinguished from irregular price changes because the latter is usually more costly than the former. Examples of regular and predictable price changes include

(i) Cinemas, car parks, restaurants have a higher charge during the peak hours; and
(ii) Air, coach and ship companies have seasonal price changes between the high and low seasons.

According to Okun, frequent changes of price are costly mainly because it raises customer's uncertainty about the price that the supplier will be offering. However, if the supplier were to print both the peak and off-peak price in the same price list and stick with what they have promised in the price list, the customer will be clear about the price pattern and there will be little cost of changing price in the Okun sense. Besides, as both the peak and off-peak prices are included in the same price list, there will be no additional menu cost for such regular price changes. These explain

(i) why some suppliers can have as frequent (regular) price changes such as having one price during the daytime and having another in the evening; and

[11]Menu cost is another, smaller but important cost.

[12]The distinction between the price changes with respect to cost shocks and that with respect to demand shocks have been discussed in the previous section.

(ii) these suppliers will usually print a price list that includes both the peak and off peak prices.

The evolution of regular (seasonal) price changes and printing all seasonal or off-peak prices in the same list provide further supporting examples to the hypothesis that some practices, institutional arrangements, and rules of thumb are indeed evolved from maximizing behaviour.

Not all industries have seasonal or off-peak pricing even though the costs of such regular price changes are negligible. This happens whenever the possible benefit from such regular price changes are negligible or negative. For example, supermarkets seldom raise their prices during peak hours. If they do, customers can buy the product during the slack hours and hoard the product until they use it. Thus, the possible benefit is significantly reduced. More importantly, if some competitors decide to keep a low price during the peak hours, those supermarkets who charge a higher price during the peak hours may suffer a loss. In general, seasonal pricing only occurs in industries where

(i) customers have difficulties in hoarding the product; and

(ii) the threat of competitors maintaining a low price during the peak season is small.

(b) Sales versus its alternative

Suppose a supplier finds it worthwhile to reduce the price temporarily. Instead of announcing a price reduction with another announcement (at a later date) of a rise in price, the supplier will usually choose to announce a sale which can be considered as a package of announcements in which customers are informed at the very beginning that the price will be reduced temporarily and eventually raised back to the original level. The advantage of the announcement of sales over two temporally-separated announcements is that customers are well informed about the price pattern at the very beginning. The two temporally-separated announcements, on the other hand, may give the customer an impression that the supplier has an erratic pricing policy which is harmful to the customer's attachment for recurrent purchases. Thus, "sales" can be considered as a new practice by the supplier to minimize the cost of changing price in the Okun sense.

(E) Near-Rationality

With a somewhat different emphasis, Akerlof and Yellen (1985a,b), Mankiw(1985) and Blanchard and Kiyotaki(1987) have drawn attention to cases in which inertial price-wage behaviour with respect to an aggregate demand shock will cause only a second-order welfare loss to near-rational agents, but first order effects on real variables such as output, employment and welfare. The inertial (or "near-rational") behaviour is rational because of the presence of a small menu cost of changing price and wage. The contribution of these papers is that it highlights the possibility that a second order menu cost is sufficient to account for large business cycles. However, the mere existence of this possibility does not imply this is what happens in the world. Our position is that the producer's cost of changing price with respect to a demand shock, due to the customer-supplier relationships suggested by Okun, are much larger than the menu cost. We believe that a successful microfoundation for business cycles does not lie in a second order menu cost causing a first order loss in real variables, but in the signal extraction problem coupled with a large cost of changing prices and wages that prevents prompt adjustments to reasonably large demand shocks.

Moreover, whether the loss of a "non-maximizer" with respect to a demand shock is small or not will depend on the size of the demand shock, the shape of the profit function, how the profit function shifts with the demand shock and the starting position of the non-maximizer. In particular, the greater the demand shock and/or the steeper the profit function, the less likely is the loss to be small. This can be illustrated with the following diagrams :

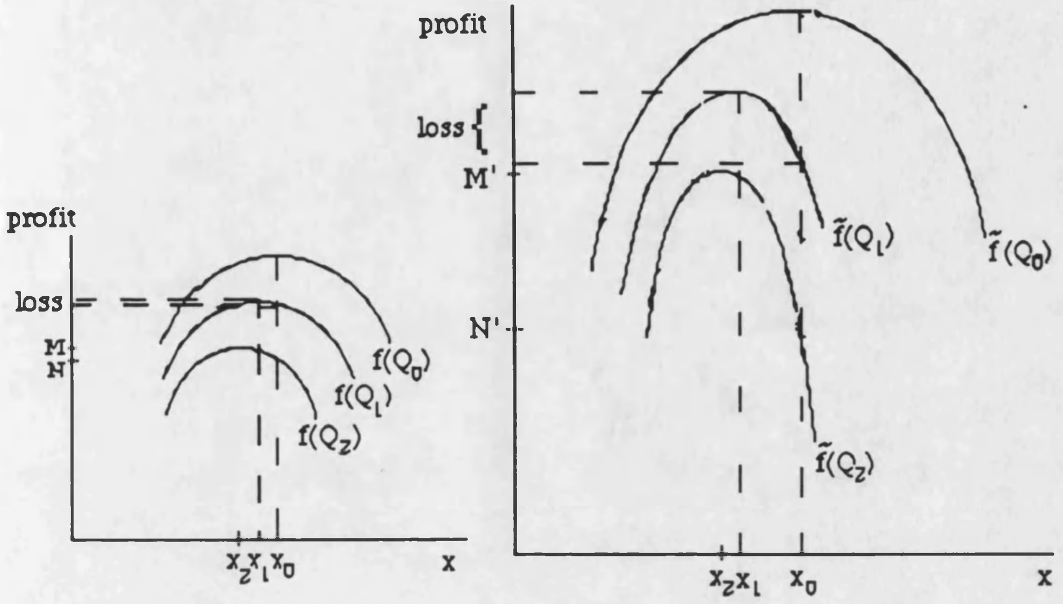


diagram (1.8a) diagram (1.8b)

where Q_0,Q_1,Q_2 are the demand, and Q_0>Q_1>Q_2
f(•) or f̃(•) is the profit function of firm

Suppose demand falls from Q_0 to Q_1. If the profit function is as shown in diagram (1.8a), the loss from not adjusting the price form x_0 to x_1 is quite small. However, if the profit function is as steep as that shown in diagram (1.8b), the loss is no longer of small. For a large demand shock such as that from Q_0 to Q_2, the loss will be MN and M'N' respectively. Even in case of diagram (1.8a), one will hesitate to declare the loss MN to be of second order. If the loss is not small, a small menu cost of changing price will not be sufficient to justify inertial behaviour [13]. This highlights the

[13]If the producers do change their price, real balance effect in the four papers of near-rationality implies that the social loss in output will be less. This creates the following embarrassing result : a just large enough aggregate demand shrink has little or no social loss but a just not large enough aggregate demand shrink will have a first order social loss in output. It is hard to imagine a smaller exogenous demand shrink to be more costly than a larger shrink. In the models we develop later, we will attempt to correct this.

limitations of near-rationality as an explanation of large business cycles.

Next, we would like to emphasize two theoretical problems inherent in this near-rationality literature. Both of these are associated with the attempt to avoid an explicit incorporation of the cost of changing price into the models. The first problem is that all the models are single period analyses. While the cost of changing price may be once-and-for-all for every change, the loss in profit may endure for more than one period. Suppose the reduction of demand from Q_0 to Q_1 is permanent, the loss of profit over the future will be $[f(Q_0)-f(Q_1)]/[1-\gamma]$ instead of $f(Q_0)-f(Q_1)$, where γ is the discount rate. Suppose $\gamma=0.9$ (ie interest rate = 11.1%), the slope of the sum of discounted profit function will be 10 times as steep as that shown in diagram (1.8a) and (1.8b). This further weakens the likeliness of near-rationality as a good explanation of large persistent business cycles. Seen from another point of view, one can state that the existing papers on near rationality fail to distinguish between the effect of persistent and transitory demand shocks. The second problem is that there is nothing to guarantee the "near-rational" agent starts at the optimum. If the agent is not choosing the optimum this period, how can we assume that the agent chose the optimum last period? If the agent does not start at the optimum of the last period, loss from inertial behaviour may or may not be of second order.

In Chapters 2 and 3, we will develop models that include inertia through near-rational behaviour as a special case, but do not necessarily require this to explain large business cycles.

1.2.2 Wage and Employment Decisions

(A) Implicit Contract

The early version of Implicit Contract theory [Bailey(1974), D.F.Gordon(1974) and Azariadis(1975)] attempts to explain observed real wage rigidity as a result of (implicit) contractual arrangements between risk averse employees and less risk averse employers. According to the theory, it pays both parties if the employers "insure" their employees with small variations in wage rate over the various possible states of nature; and in return the employers are compensated by risk premia in the form of lower average wages which workers are implicitly willing to pay for such wage insurance. The result of such insured (sticky) wages were then used by some macro-economists [such as Gray(1976,1978), Poole(1976) and Fischer(1977)] to account for non-neutrality of monetary policies on aggregate output and employment. Akerlof and Miyazaki(1980), however, challenge such an explanation of Keynesian unemployment by extending the Azariadis-Bailey result of insured wage to the implicit contract of insured employment. The logic, which they called the Wage Bill Argument, is as follows :

> *"Suppose there is an implicit contract whereby in a state of the world s a firm employs n_1 workers at a wage w but lays off n_2 workers, each worker randomly being laid off with the same probability $n_2/(n_1 + n_2)$. First observe that a risk averse employee would prefer, ex ante, a contract which guaranteed him employment in state s at the wage $wn_1/(n_1 + n_2)$ to the lottery of receiving w with probability $n_1/(n_1 + n_2)$ and 0 with probability $n_2/(n_1 + n_2)$. Secondly, note that the firm would be indifferent between this employment-guaranteeing contract and the layoff contract since its wage bill is unchanged at wn_1. These two observations (together with the assumed continuity of the worker's preferences) imply the existence of a wage rate w^* such that (i) $w^* < wn_1/(n_1 + n_2)$ and (ii) the*

worker strictly prefers guaranteed employment at w. Likewise a profit-maximizing firm would be willing to offer a full employment guarantee at w* because it can reduce its wage bill form wn_1 to $w^*(n_1 + n_2)$ without sacrificing its output. Because both firms and workers prefer full-employment to layoffs in any given state, it follows that unemployment cannot occur in an equilibrium with rationally negotiated contracts."*

While Akerlof and Miyazaki are right in pointing out the limitation of using the real wage rigidity arising from implicit contracts as an explanation of Keynesian unemployment, there are at least two important lines of criticism that help to explain involuntary unemployment despite the existence of insured employment. The first is due to the possibility of involuntary unemployment in efficiency wage models of the shirking and turnover cost variety that will be mentioned in the next section. Thus, even if employers find it beneficial to provide income insurance to those they employ, they might not reduce the wage to the full employment level (i.e. the second logic of the above quotation does not follow) as such a reduction in the wage may reduce production effort, raise shirking and increase turnover cost. In such a case, an guarantee of employment to those employed may co-exist with a pool of involuntary unemployment.

The second is related to the relative degree of risk aversion between employees and employers at various states of economic conditions. Charles Schultz(1985) has argued that employers may indeed be more risk averse in the case of a large and climatic reduction in demand. While it open to question whether such risk reversals do actually occur, the difference in risk aversion will in no doubt be small for the state of very adverse demand shocks. This makes the potential gain from insured employment against adverse demand shocks rather small. Because of their preference for flexibility, employers may not – at the very beginning – find the small reduction of wage worthwhile for the promise of fully insured employment including the case of very adverse demand shocks. Even if they do, they might find it worthwhile to break the promise in the case of a general adverse demand shock because he (rightly) expects many other employers will do the same thing as he does. Workers, seeing this kind of possibility, will not accept much wage cuts even if some employers "claim" to provide the insured employment against very adverse cyclical demand shocks. In other words, it is not the risk reversal between workers and employers, but (i) the worker's distrust that employers will honour the guarantee of employment, (ii) the employer's preference on flexibility, and (iii) the small potential gain from insured employment against adverse demand shocks that make the scheme of insured employment only partial [13] instead of full. Thus, in the case of a moderate reduction in demand, employment of individual firm will remain unchanged. On the other hand, if the reduction of demand is large and climatic, the employer will consider laying off some workers [14], introducing a part-time work-schedule, or overall dismissal subjected to recall [15]. In Chapter 3, we will build a model of employment containing such features. Some other interesting results are also obtained.

(B) Efficiency wages: traditional, shirking and labour turnover models

Efficiency wage models were first developed in the context of Less Developed Countries by Stiglitz(1976) and then applied to developed economies by Solow(1979). In the simplest form presented in Yellen(1984), each employer is assumed to have a production function of the form $Q = F[e(w)N]$, where N is the number of employees, e is the effort per worker and w is the real wage. The representative employer is assumed

[13]That is, employers may only have incentive to guarantee employment up to a certain adversity of demand. It will be interesting to include such possibility into the model of Akerlof & Miyazaki and show that (usually) there exists one such scheme that is superior than a fully insured implicit contract.

[14]If workers' ability differ, employers might prefer retrenching the least productive worker.

[15]See Okun(1981).

to choose w and N to maximize the profit (π) :

$$\max_{w,N} \quad \pi = F[e(w)N] - wN$$

Writing w* and N* as the optimal value, the first order conditions imply

$$\frac{w^*}{e(w^*)}\ e'(w^*) = 1 \qquad (1.5)$$

$$w^* = e(w^*)\ F'[e(w^*)N^*] \qquad (1.6)$$

Equation (1.5) implies that employers should offer a wage at which elasticity of effort is unity. Given the w* determined in equation (1.5) , equation (1.6) states that employers should hire labour until the marginal product of labour, $e(w^*)F'[e(w^*)N]$, equal to the real wage, w*. Thus, aggregation of (1.6) over all firms will give the aggregate demand for labour. As Yellen(1984) has stated, *"as long as the aggregate demand for labour falls short of the aggregate supply and w* exceeds the labour's reservation wage, the firm will be unconstrained by the labour market condition in pursuing its optimal policy so that equilibrium will be characterized by involuntary unemployment."*

It must be noted, however, this model only demonstrates the possibility of involuntary unemployment, not its necessity. If the aggregate supply of labour turns out to be smaller than the aggregate demand for labour, the model will predict full employment instead of involuntary unemployment.

The shirking model in Shapiro and Stiglitz(1984), on the other hand, predicts that involuntary unemployment would be <u>a certain event</u> if

(i) shirking by employees implies a cost (such as lower productivity) to the employer;

(ii) employers can only imperfectly detect whether a worker is shirking or not; and

(iii) employers have difficulties in using employment fees, performance bonds or seniority wage schemes as a penalty for shirking [see Yellen(1984)].

The intuition of their result is as follows :

"To induce its workers not to shirk, the firm attempts to pay more than the

going wage, then if a worker is caught shirking and is fired, he will pay a penalty. If it pays one firm to raise the wage, however, it will pay all firms to raise their wages, the incentive not to shirk again disappears. But as all firms raise their wages, their demand for labour decreases, and unemployment results. With unemployment, even if all firms pay the same wage, a worker has an incentive not to shirk. For, if he is fired, an individual will not immediately obtain another job. The equilibrium unemployment rate must be sufficiently large that it pays workers to work rather than to take the risk of being caught shirking." [Shapiro and Stiglitz(1984)]

The turnover cost models, such as Stiglitz(1974), Schlicht(1978) and Salop(1979), give another possible explanation of involuntary unemployment. In these models, the representative employer has an incentive to economize on turnover costs by offering higher relative wages. As every employer attempts to do so, the aggregate demand for labour decreases. According to Salop(1979), if:

(i) search while unemployed is more efficient so that not all workers prefer on-the-job searching [16]; and

(ii) employers have difficulties in charging employment fees and having a sufficiently low wage for the new hires,

we <u>might</u> have an equilibrium with involuntary unemployment which will in turn justify (at least in some degree) the employers' attempt to reduce turnover costs [17]. Just as in the simple efficiency wage model, involuntary unemployment in this model is, again, merely a possibility instead of necessity.

---

[16]Even if all workers prefer on-the-job searching, the result will still hold if employers believe higher wage will reduce on-the-job searching.

[17]In case of full employment equilibrium, only these employers with higher relative wage will succeed in reducing turnover cost, and the success is at the cost of higher turnover for those employers with lower relative wage.

In all of the above models, the existence of involuntary unemployment is due to the high wage policy of every employer who initially attempts to offer a higher relative wage to encourage high production effort, discourage shirking and reduce turnover which are in fact achieved by the pool of involuntary unemployment associated with the high wage policy of all employers. Although schemes such as employment fees, performance bonds or seniority wages can theoretically eliminate the unemployment, Salop(1979), Shapiro and Stiglitz(1984) and Yellen(1984) have convincingly reviewed the difficulties or limitations in such use of the schemes in practice [18]. Nevertheless, the explanation of involuntary unemployment by the above types of models will be weakened or destroyed if there exists sufficient amount of jobs in which:

(i) employers have little cost in monitoring the production effort of workers ;

(ii) employers have little turnover cost ; and

(iii) the required production effort is fixed.

Examples of these include those in the category of causal labour market described in Okun(1981). As employers in this market have little or no problem in reducing the wage towards the reservation wage of the representative worker, involuntary unemployment can be greatly or totally eliminated if there exists a sufficient amount of these jobs. Of course, involuntary unemployment still exists if the relative amount of these jobs is small in comparison with the types of jobs described in the efficiency wage models, shirking models and labour turnover models. [Besides, the former case may still imply a large pool of search (instead of involuntary) unemployment if the wage gap between the two types of jobs is large.] Whatever the case is, the above discussion implies a dual labour market economy in which the career labour market [19] is characterized by high wage, little turnover and long queue of applicants; and the causal labour market is characterized by

---

[18]See Okun(1981)

[19]See Okun(1981).

low wage and high turnover.

In addition, the three types of models suggest that workers with identical characteristics can receive different wages at different firms as the effort functions, shirking, monitoring, and turnover costs of firms may differ from each other[20].

One criticism of these models is that they do not explain the cyclical unemployment arise from aggregate demand shocks and nominal wage rigidity [21]. In these models, there is only real wage rigidity. No discussion has been provided about how the nominal wage behaves in face of aggregate demand shocks. Thus, the models can at best provide a static explanation of unemployment (such as the efficiency, shirking and turnover considerations), and fail to give a proper account of the cyclical variations in unemployment. For this reason, we will turn to the discussion by Okun(1981) and Schultz(1985) in the next section. The difference in the source of the unemployment is important. In the models described in this section, the pool of unemployed workers has some desirable aspects, such as discouraging shirking and turnover, and the Government is incapable of removing the unemployment by means of macroeconomic policy [22]. On the other hand, unemployment arising from the combination of nominal wage rigidity and negative aggregate demand shocks is likely to be undesirable, and it may pay the Government to use stabilization policies to eliminate such cyclical variations in unemployment.

---

[20]Yellen(1984) also uses the efficiency wage model to explain employer's preference or discrimination in the recruitment of observationally distinct groups.

[21]Even if the initial wage is low enough with full employment, (sufficient) reduction in aggregate demand combined with nominal wage rigidity will cause involuntary unemployment. We add the work "sufficient" because of the guarantee of employment explained in the last section. See the next section for further explanation.

[22]Nevertheless, the involuntary unemployment suggested by the models can co-exist with that arises in the usual theory of business cycle.

(C) Okun(1981) and Schultz(1985)

In the previous sections, we have seen some outstanding models which attempt to highlight and explain some of the very important features of the labour market, such as wage rigidity, employment insurance and the presence of unemployment. Unfortunately, reality appears to be even more complicated than that postulated by the models. Because of this, Okun(1981) and Schultz(1985) eschew formal modelling and content themselves with a verbal analysis. The attractiveness of this verbal approach is that they can attempt a much richer analysis than is permitted by formal modelling, and thus hopefully get closer to reality. Okun gives a balanced view of both the wage and employment decisions of employers. Schultz, on the other hand, concentrates on the wage decision in face of Knightian uncertainty. This Section summarizes their ideas.

In the early version of implicit contract theory, the association between employers and workers is due to the difference in their risk aversion, and realized in the form of the employer's guarantee of wages and employment to the workers. Okun(1981) and Schultz(1985), however, emphasize another important incentive for the association. According to them, employees acquire non-transferable, firm specific knowledge or skill through on-the-job training. As long as the employee stays with the same employer, the knowledge or skill will be useful and productive to the firm. The employer, in return for higher productivity, will be willing to pay higher wages to the experienced employee. On the other hand, if the employment relationship is ever broken, there will be significant search and initiation costs for both parties until the (new) worker acquires the "non-transferable firm specific" skill through sufficient experience in the new job. The employer will have to bear, in addition to any recruitment costs (advertising and interview fees), at least part of the toll cost which includes direct costs such as any formal orientation program, expenditure to foreman or "breaking in" new employees as well as lowered productivity during the adjustment process. The worker, in addition to the search cost while unemployed, may have to accept a temporarily lower wage if the required knowledge or skill in the old job is different from that of the new. Thus, both the difference in risk aversion and the non-transferable firm specific skill provide incentives for both parties to maintain the employment relationship as long as they can.

According to Okun, an explicit contract may ensure such association, but only at considerable expense of negotiation and legal work, and only through sacrifice of flexibility. The latter is particularly important to the employer because, over a sufficiently long period, some Knightian change of economic climate is bound to occur. Instead, employers opt for an implicit contract (ie non-binding statement) to maintain the relationship over the long run :

> *"In addition to , or instead of, affecting expectations by specific binding obligations, the firm may try to influence the expectations of willing applicants and of potential quitters by various types of statements about the future that are not binding. They can have some force and some credibility by putting the firms reputation on line. The firm providing such implicit contracts must decide how much of an investment it is prepared to make in its personnel policy. If it makes strong statements that paint a rosy future for its recruits, it can hope to increase the supply of labour in the short run, but it then faces greater risks of excessive payroll costs to fulfil its promise or of costly disappointments by its workers that trigger higher quits and lower productivity if it fails to fulfil these promises."[Okun(1981),P.89]*

Let us now come to the details of the implicit contract. In reality, the employer's guarantee over employment and wage is generally far more complicated (and with more dimensions) than suggested i the original implicit contract literature. For example, as mentioned in section 1.2.2(A), employers may only be willing to insure employment up to a certain point; a truly climatic fall in demand lead to layoffs. While employers will try their best to emphasize the improbability of such an event, it remains a potential worry to workers. To reduce such worry and raise the attachment of workers, they may also guarantee, or condition the thinking of workers by their previous action, that the utmost will be done to reduce the harm and experienced workers are to be the last affected. Thus,

> *"If the slack period is confidently expected by the firm to be very short in duration and if the goods produced by the firm are readily storable at low cost, these workers may be used to build up inventories in slack periods. But if there is considerable uncertainty about the possible duration of the slack and high costs to storage (including the impossibility of storing outputs that are services), output may be reduced. Firms may be able to assign the workers maintenance tasks like cleaning, repairing and painting, or it may really keep them in a state of on-the-job underemployment."[Okun(1981),P.57]*
>
> *"Finally, firms that normally subcontract part of their operations may be able to implement a no-cut strategy inexpensively by suspending that practice. Indeed, that is sometimes an important incentive for subcontracting during prosperity."[Okun(1981),P.107]*

If the recession is expected to be so adverse and prolonged that some cut in the labour force is necessary, some employers may choose to have, say, a four-day workweek instead of a twenty percent layoff within the labour pool. The employer will also promise that recall will be made as soon as the economy begins to recover. Besides, employers will try to convince the experienced and productive employee that they are to be the last affected by establishing the practice of :

(i) stopping new recruitment before any layoff or part-time working schedule is made;

(ii) making sure that only the least productive worker will be dismissed if layoff is unavoidable.

Thus, the guarantee of employment in reality is more complicated than that suggested in the early implicit contract literature. Moreover, the extent of guarantee may differ according to the difference in the nature of firms. For example, firms or organizations with more steady demand for output will try to emphasize greater insurance of employment to attract qualified employees. The extreme case is the government that virtually provides a full insurance of employment to the civil servant. On the other hand, firms with more erratic demand for output will opt for higher relative wages to attract and compensate qualified employees.

Similarly, the implicit guarantee of wage path in reality is somewhat different from that in the implicit contract literature. In the early implicit contract and efficiency wage models, the rigidity of wage is in real terms. Okun and Schultz, suggest that wage will be

(i) rigid in nominal terms for some length of period; and

(ii) periodically adjusted so that the fluctuations of real wage will be within a limited range and yet long term changes in wages reflect long term changes in market fundamentals.

Thus, we do not have the insurance of a fixed real wage in reality. Instead, employers only insure workers by guaranteeing small variations in real wages. This raises the question as to why employers do not find it optimal to offer the insurance of a fully fixed real wage. This writer, following the analysis of Schultz(1985), sees the explanation as follows :

> In the contracting approach, a certain known distribution of states is assumed. In such a case, the guarantee of a fixed real wage is perfect. In reality, however, there are Knightian changes (whose probability of occurrence cannot be determined in advance) in relative demand, technology and cost that necessitate some unknown changes of real wage

> over the very long period of association[23]. If so, the guarantee of a fixed real wage (or a fixed path of real wage) will not be optimal. The ideal wage scheme has to be capable to cope with two hard-to-reconcile facts. It has to be sticky enough to provide sufficient insurance to workers but yet allow flexibility to deal with Knightian uncertainty. This led Schultz to recommend the wage scheme mentioned earlier.

Beside the short term nominal rigidity of the wage, Schultz also suggests a very cautious and sluggish adjustment of the money wage to changes in conditions whose permanence is open to question. This bears a close resemblance to the signal extraction problem we explored in the context of the product market. Thus, employers prefer to wait for more observations before any change in the wage is made. Indeed, such a practice has become so "natural" that once the wage is raised, it will be extremely awkward to revise it downward. This is probably why traditional Keynesians tend to pre-suppose the Keynesian wage floor as an institutional feature, rather than as behavioural pattern arising from long run optimization.

(D) Role of Our Model

One of the troubles with most macroeconomic models is that they rely on wages, employment or both as the adjustment variable(s) to aggregate demand shocks. However, as seen in the previous review, employers have incentives to provide at least some implicit guarantee over both wages and employment. Thus, at least in the short run, wages and employment will remain sticky in the face of moderate reductions in demand. It also implies the labour market is more complicated than that supposed by the simple debate over price (wage) versus quantity (employment) adjustment. In Chapter 3, we will present a model which emphasizes the role of another variable – production

---

[23]Indexation is impossible because the list of variables will be too long and the functional form will be too complex. Moreover, employers and workers have incentives to disagree in the choice of variable, and to what extent it is used in the indexation.

effort – in the short run adjustment to moderate demand shocks. By incorporating some important elements of the efficiency wage models (an effort function) and the theory of implicit contracts (a cost of layoff), it is shown that

(a) In the case of moderate demand shocks, production effort (and hence output) instead of employment will be the main variable of adjustment. In other words, there will be some form of labour hoarding in mild recessions and a tighter working schedule in mild expansion.

(b) Because the reputation cost of layoff consists of a fixed and sizable component, employers will resort to layoffs only in the case of very adverse demand shocks. Once the layoff option is chosen, there will be a considerable amount of workers being laid off. Thus, in the case of general demand shocks, we expect a mild shock will be characterized by little change in unemployment and an adverse shock be characterized by massive and sharp fall in employment.

(c) While efficiency wage and traditional Keynesian models are criticized for their embarrassing prediction of counter cyclical productivity [see Akerlof and Yellen (1985) and Okun(1981)], result (a) demonstrates that these models - after allowing for the invariance of unemployment with respect to mild demand shock - can be consistent with the usual observation that productivity, measured in terms of output per head, is procyclical.

(d) In the face of an adverse demand shock, it is usually better, from the point of view of aggregate employment, to simulate the economy before rather than after employers make the layoffs. The rationale is as follows. If the Government succeeds in stimulating the economy before the recession, employers with an excessive amount of labour may, in view of the reputation cost of layoffs, hesitate to make the retrenchments. On the other hand, once the recession has started and employers have cleared the excessive amount of labour through layoffs, employment will not return to the original level even if aggregate demand does.

Last but not least, the model in Chapter 3 is consistent with many of the features of the labour market discussed in the previous review. It is consistent with the implicit guarantee of employment as there is a reputation cost of layoff. By assuming that cost is finite, albeit large, it is consistent with Okun's proposal of only partial guarantee of employment[24]. The same type of model is also applicable to the case in which the employer, in view of a recession, chooses between hoarding the excessive amount of labour and the introduction of part-time working. Again, once the option of part-time working is chosen, the model predicts a large and discrete reduction of working hours. (Whether the employer, in the case of adverse recession, chooses layoffs or part-time working will depend on the specific nature of the firm.) The model is also consistent with the toll cost explanation of employer-employee relationship by assuming a hiring cost which includes the cost of the on-the-job acquisition of the non-transferable firm specific skill. The model can also be extended to explain that the recruitment policy of employers in the career market is dominated by long-term considerations. In the face of a temporary rise in demand, employers may choose to have a tighter working schedule (ie higher production effort) despite its higher cost in the short run. This is so because (i) it takes time to train a worker; and (ii) new recruitment in case of temporary demand shock may mean excessive payroll or costly layoff in the future. On the other hand, if the rise of demand is permanent, employers may find it worthwhile to take the effort to recruit and train the new employees. Once the employers have made the recruitment, they dislike retrenchments because of the large reputation cost of layoff. Thus, in contrast to the Neoclassical theory, the model predicts labour in terms of heads (and hence the "usual" labour cost [25]), is fixed in the short run. While capable of explaining cyclical variations of unemployment, the model can also be consistent with a static pool

[24] So long as one assumes that wage rates are made before the employment decision, the model will also be consistent with the implicit guarantee of wage, short term rigidity on nominal wage and etc.

[25] This excludes the higher cost associated with the higher production effort.

of involuntary or search unemployment due to the high wage policy characteristic of efficiency wage models. Thus, in contrast to Akerlof and Miyazaki(1980), employment insurance can co-exist with unemployment.

(E) Cyclical variations of unemployment

In section 1.2.2(B), we discussed some models which suggest the existence of a static pool of unemployment. Unfortunately, the models do not explain the presence of nominal wage rigidity and fail to account for those cyclical variations in unemployment which should be the target of Government stabilization policies. Schultz uses Knightian uncertainty to explain "short-term nominal wage rigidity". While some degree of nominal wage rigidity, as supposed by many Keynesians such as Yellen(1984), is necessary to allow the aggregate demand shock to cause cyclical variations of unemployment, this is never a sufficient condition. For example, if employers are willing to offer a full guarantee of employment, one might not have any cyclical variation of unemployment even if the wage is nominally rigid. Nevertheless, not all employers are so risk neutral as to provide full guarantee of employment. Thus, in the case of very adverse demand shock, layoffs or part-time working will result. If we assume a large number of firms evenly distributed between the point of layoff and the point of hiring (ie some are growing while some are declining), moderate aggregate demand shocks will cause cyclical variations in unemployment. Besides, the guarantee of employment refers only to those employed by firms, but not to those who are unemployed. Suppose we start with a static equilibrium with some workers retiring and some school leavers entering the labour force. In the face of a reduction in aggregate demand, employers will stop or slow down new recruitment while allowing old workers to retire. When aggregate demand rises, they will speed up their long term plan of recruitment. This gives another explanation for cyclical variations of unemployment. It also explains why most of the burden of cyclical unemployment falls on school leavers or new applicants.

1.2.3 Empirical Works on the Theory of Mark-up Pricing

Since empirical work on the pricing decision has already been reviewed in Godley and Nordhaus (1972), and Laidler and Parkin (1975), our remarks here will be brief.

As reported by Laidler and Parkin, most of the empirical work in the United Kingdom, with only a few exceptions, show significant demand effects in the price equations. Among the few exceptions, the most appealing challenge is by Godley and Nordhaus (1972). By decycling factor price and productivity changes, they computed a time series for "normal costs" and then found empirical support for the "normal cost hypothesis", according to which prices respond to changes in normal costs and are independent of demand. On the other hand, empirical work in the United States, including the subsequent work by Gordon (1975), found that demand variables exert independent upward pressure on prices.

Laidler and Parkin, possibly influenced by the literature on the Phillips Curve at that time, attempt to reconcile the difference in results by claiming mis-specification in ninety out of the one hundred regressions conducted by Godley and Nordhaus (1972). They criticize Godley and Nordhaus(1972) for specifying the rate of price change as a function of <u>changes</u> in demand instead of <u>level</u> of demand. We disagree with this criticism. As we will show in Chapter 4, if $P_t = (1 + m_t) AC_t$ is the right equation and $\ln(1+m_t)$ can be approximated by a linear function of the logarithm of demand, the appropriate price equation should have the <u>level</u> of price depending on the <u>level</u> of demand (or the change in prices depending on the changes in demand.)

Although we disagree with Laidler and Parkin's previous claim of mis-specification in Godley and Nordhaus (1972), the result by Godley and Nordhaus may still be mistaken because of the presence of some other mis-specification(s). For example, as criticized by Laidler and Parkin, the coefficient on price changes predicted from normal cost changes is only 0.6 in Godley and Nordhaus's "preferred" equation. Such estimated value, being significantly less than the theoretical value of unity, is evidence of possible mis-specification in their equation.

Instead of just searching for mis-specifications in Godley and Nordhaus(1972), we would like to make some more general criticisms on all the empirical work referred to here. First of all, this work is all based on aggregate data. Second, it assumes the mark-up equation $P_t = (1 + m_t)AC_t$ holds for all observations. As we will show, by means of a more elaborated theoretical foundation in Chapters 2 and 4, $P_t = (1 + m_t)AC_t$ only holds for those observations in the "raise price" regime. For those observations in the sticky price regime, we will have $P_t = \overline{P}_t$, where $\overline{P}_t$ is the price in the last period. As discrete jumps in individual prices may be smoothed out on aggregation, our theory also suggests that empirical work using individual price data should be preferred. Last but not least, the demand variable used should be some kind of <u>expected</u> instead of <u>current</u> demand. Our empirical work in Chapter 4 attempts to overcome these criticisms. In addition to testing the "normal cost hypothesis", we would also like to provide some rough estimate(s) of the cost of changing price for the product chosen. Such estimation is important because the quantitative significance of the sluggishness/stickiness of prices will depend on the cost of changing price. It will also provide side evidence on whether the reputation cost or menu cost is a more important component in the cost of changing price.

Before moving to the next section, it is worthwhile repeating here one important comment by Laidler and Parkin (1975):

"Whether or not prices respond to excess demand independently of cost changes is not relevant to the overall existence of a short-run trade off between the rate of inflation and excess demand. This trade off will exist if either product prices or factor prices or both are responsive to excess demand since no one disputes that cost (factor price) changes affect product prices. It is an almost universal finding that prices respond to cost changes, a major element of which is wage change. Wages in turn, as we have seen, are usually found to be responsive to excess demand as well as to other variables, sometimes including current or expected price changes. By taking the wage and price equations together, it is possible to obtain quasi-reduced form relations with which both price and wage changes are functions of excess demand and expected inflation, as well as of other exogenous variables which might appear in other structural equation."

1.2.4 Disappearance of price stickiness on aggregation and the neutrality of money

In section 1.2.1(B), we have reviewed the various arguments for individual price stickiness/sluggishness. Under the assumption of a representative producer, they all implicitly or explicitly conclude there is aggregate price sluggishness and thus also the non-neutrality of money. The jump from individual price stickiness to the non-neutrality of money is however questioned by Caplin and Spulber (1987) who present a model in which individual price stickiness disappear on aggregation. The result is striking because it produces the neutrality result despite the presence of a fixed cost of changing price and individual price stickiness. However, because of the complexity in aggregation, Caplin and Spulber can only show their result for the very restrictive case in which (i) monetary growth is monotonic so that, in the absence of other shocks, one side (S,s) rule will be the optimal pricing policy; and (ii) the initial cross-sectional distribution of relative prices is uniform between the two thresholds (S,s) [26]. The robustness of their result to more general specifications is however questioned by Blanchard and Fischer (1989). To see this, first consider the intuition behind Caplin and Spulber (1987) :

> Suppose there is a 1% rise in money supply and the fixed cost of changing price is of the magnitude equivalent to, say, an 10% rise in the desired price. Because of the fixed cost of changing price and the assumed distribution of relative price, only one tenth of the firms will raise the price. However, these firms will raise the price by 10%, an amount large enough to make the rise of aggregated price index equal to 1%. Such percentage rise is exactly the same as that of the money supply. Thus, the discrete jump in price by a fraction of the firms implies individual price stickiness disappears on aggregation.

As we can see, the above argument depends very much on whether there exists a steady state cross distribution of price deviations and whether such distribution is uniform or not. For example, Blanchard and Fischer (1989) suggests that if money is not growing monotonically, but instead follow a symmetric random walk, the optimal pricing policy will be a two sided (S,s) rule and the steady state cross distribution of price deviations [27] is likely to have higher density at the return point (eg triangular) instead of being uniform. If so, a 1% rise in money supply will cause less than one tenth of the firms to make the 10% rise of price and the percentage rise in aggregate price index will be

[26] To justify the second assumption, Caplin & Spulber proceed to show that the uniform distribution will survive with the specific monetary shocks they considered.

[27] If, as will be explained in Chapter 5, we want to use idiosyncratic shocks to generate the dispersion of price endogenously, it would be more appropriate to analyze in terms of price deviations between the actual price and "optimal" price [as was done in Blanchard and Fischer(1989)] rather than the relative prices between individual price and aggregate price index [as was done in Caplin and Spulber (1987)]. We well return to this in Chapter 5.

smaller than the percentage rise in money supply [28]. Nevertheless, even in such a simple counter example, Blanchard and Fischer find that it is extremely difficult to derive the steady state cross distribution of price deviations[29]. Thus, the counter argument by Blanchard and Fischer here remains hypothetical :

> *"Unfortunately, aggregation is hard if not impossible in most models with state dependent rules, so we do not know the answer. From the few examples we have, however, it appears that the neutrality result is not robust ."[Blanchard and Fischer (1989), Chapter 8]*

Indeed, one reason that stops Caplin and Spulber from going towards more general specifications of the monetary generation process is due to the difficulty in analytic aggregation:

> *"A theoretical difficulty in modelling two-sided policies is that their properties under aggregation appear highly complex. Specifically, it is not possible to specify an initial cross-sectional distribution of prices which survive shocks ."[Caplin and Spulber (1987)]*

Because of this, we can only resort to numerical methods. The aim of Chapter 5 is to use numerical simulation to check the robustness of the Caplin and Spulber result to some more general specifications of monetary generation process. In particular, we would like to check whether the neutrality result will hold in Blanchard and Fischer's counter example (i.e. when the monetary generation process follows a symmetric random walk). Besides, we would also like to check whether the conclusion will again be changed with the addition of an underlying trend of monetary growth to Blanchard and Fischer's counter example.

---

[28]Indeed, as admitted by Caplin and Spulber themselves, *"if monetary growth is non-monotonic, the one-sided pricing policy has to be replaced by a two sided one and the neutrality propositions no longer holds""*. Again, aggregation problem here forbidden a rigorous analysis by Caplin and Spulber.

[29]Blanchard and Fischer emphasize that their statement refers to the steady state distribution of price deviations across price-setters, and not of price deviations for a given price-setter which will be triangular.

The following results are obtained :

(a) Money is neutral in the sense that one cannot keep raising (reducing) aggregate output by an indefinite rise (reduction) of the money supply; and

(b) Money is non-neutral in the sense that occasional reduction of money stock in an inflationary world will cause a reduction in aggregate output. If there were any other exogenous reduction in aggregate demand, the government may be able to use monetary policy to reduce the initial reduction of output. Monetary policy, however, cannot reduce the "long-run" [30] reduction in output arising from such an exogenous cut in aggregate demand.

1.3 Plan of the Thesis

In Chapter 2, we will first present a model on the price decision. The model is particularly helpful in showing how signal extraction problems and the cost of changing price can generate a significant degree of price stickiness not well explained by the recent literature. The remaining part of Chapter 2 is then devoted to examining the remaining two hypotheses in the theory of mark-up pricing. A brief discussion on the evolution of cost-oriented pricing and the applicability of the mark-up equation in macroeconomic analysis is also included. After that, the model is extended to the employment decision and is presented in Chapter 3. Here, emphasis is on (i) the role of production effort (instead of wage and employment) as the main adjusting variable with respect to moderate demand shocks; and (ii) the role of a fixed reputation cost of layoff (originating in the implicit, non-binding guarantee of employment) in creating a bang-bang employment decision. Chapter 4 is a piece of empirical work on the cost of changing

---

[30] "Long run" here is defined as the state where the actual price (P_{kt}) is adjusted to the desired price(P_{kt}*). See the discussion in Chapter 5.

price; we also check the "normal cost hypothesis" proposed by Godley and Nordhaus (1972). Chapter 5 is a simulation exercise to check the robustness of Caplin and Spulber's neutrality result to more general specifications of the monetary generation process. Finally, the conclusions and suggestions for further research are included in Chapter 6.

Chapter 2

In Chapter 1, we have briefly reviewed the limitations of the B-M-R models in explaining the degree of price stickiness in our world. Now we build a new model that is not only capable of explaining a greater degree of price stickiness, but also more realistic and elegant in tackling the signal extraction problem and the associated process of expectation revision. After this, we will go into the details of pricing response to various types of shocks. We will also check the following three hypotheses implicit in the theory of mark-up pricing :

(a) a sluggish pricing response to demand shocks;

(b) a relatively fast pricing response to cost shocks; and

(c) a unit elasticity of average cost (AC) in the formula P=(1+m)AC.

2.1 The Basic Model

2.1.1 Assumptions

Based on the argument by Koutsoyiannis (1979), we assume that a representative firm has some kind of planned excess capacity so that production will always be within the capacity [1]. The following assumptions for the formulations are made to achieve a more realistic condition :

(i) A representative producer does not have perfect information to distinguish between a permanent demand shock from a transitory one, and there is uncertainty about the persistence of any observed demand shock.

---

[1] This is a slightly restrictive assumption because it is possible that the demand shock is so large that the required production is beyond capacity. If the shock is also expected to be permanent, the producer may consider expanding the capacity instead of raising the price. In other words, a satisfactory formulation should, in addition to the decision of changing price, include the decision of changing capacity. We will return to the discussion later.

(ii) As a result, the producer does not know the exact demand until the end of each period. However, price has to be set before the goods can be sold (i.e. at the beginning of each period).

(iii) At the beginning of any period t, the producer forms an expectation for the probable sales and sets the price according to such an estimated demand.

(iv) At the beginning of the next period t+1, a new observation of sales at t arrives and the producer revises his expectations and pricing policy accordingly.

To facilitate a simple presentation of the model, we also make the following specific assumptions:

(1) We assume the level of demand α_t consists of a permanent component α_t^P and a transitory component α_t^T :

$$\alpha_t = \alpha_t^P + \alpha_t^T \qquad (2.1)$$

and the two components are generated from the following stochastic process :

$$\alpha_t^P = \alpha_{t-1}^P + \varepsilon_t^P \qquad (2.2)$$

$$\alpha_t^T = \varepsilon_t^T \qquad (2.3)$$

where ε_t^P is generated from a normal distribution with mean zero and variance σ_P^2 [i.e. $\varepsilon_t^P \sim N(0,\sigma_P^2)$]; and

σ_t^T is generated from a normal distribution with mean zero and variance σ_T^2 [i.e. $\varepsilon_t^T \sim N(0,\sigma_T^2)$].

With such kind of settings, it can be shown that

(i) given the information at the beginning of t, the optimal predictors for all future demand α_{t+j}, $j \geq 0$ are equal to each other. That is

$$\alpha_{t+j|t} = \alpha_{t+k|t} \equiv \hat{\alpha}_t \qquad \forall\, j,k \geq 0 \qquad (2.4)$$

where $\alpha_{t+j|t}$ is defined as α_{t+j} expected at the beginning of t; and

$\hat{\alpha}_t$ is a simplified notation for all these $\alpha_{t+j|t}$, $j \geq 0$.

(ii) $\hat{\alpha}_t$ the optimal predictor of future level of demand expected at the beginning of t, is a distributed lag of previous level of demand :

$$\hat{\alpha}_t = \lambda \sum_{i=0}^{\infty} (1-\lambda)^i \alpha_{t-i} \tag{2.5}$$

where $\lambda = \sigma_P^2 / \sigma_T^2$.

(2) Whenever a producer changes the price, there is an administrative cost of changing the posted price list; of notifying (or explaining to) the customers about the price change etc. There should also be a reputation cost (gain) for every rise (fall) in price. For presentation sake, we will first assume the following cost of changing price, $L(\Delta P_t)$:

$$L(\Delta P_t) = \begin{cases} A + \xi\,(\Delta P_t) & \text{if } \Delta P_t \neq 0 \\ 0 & \text{if } \Delta P_t = 0 \end{cases}$$

where (i) the fixed component, A, is used to capture the menu cost and part of the reputation cost [2] [It also has the effect of penalizing too frequent changes in prices]; and

(ii) the linear component, $\xi(\Delta P_t)$, is used to captured most of the reputation cost/gain of changing price.

In section 2.3.2, we will consider more general specifications of $L(\Delta P_t)$ such as the inclusion of a non-linear component which will penalize too sharp a change of price [Since this will create an interlinkage of price decision at different periods (c.f. the quadratic cost in Rotemberg (1982a,b)), the solution procedure will be somewhat more

---

[2]It is also possible to assume that the fixed component to be different for $\Delta P_t > 0$ and $\Delta P_t < 0$. Some may also prefer the fixed component to be zero for $\Delta P_t < 0$. These will not cause much complication and is left as an exercise to the reader.

complicated than that with fixed and linear cost].

(3) With the expectation in assumption (1) and cost of changing price in assumption (2) [3], the producer, at the beginning of any period t, is choosing between :

(a) Changing the price with an initial cost $A + \xi(\Delta P_t)$

If the demand shock turns out to be a permanent one, the profit at t will be higher than that of action (b) (not changing the price). This is so because the price has been changed to a more appropriate level.

On the other hand, if the demand shock turns out to be a transitory one, the profit at t will be lower than that of action (b) since the price has been changed to an inappropriate level. Besides, to avoid further loss of expected profit in later periods, the producer might have to reverse the initial price change which will involve a cost $A - \xi(\Delta P_t)$.

(b) Maintaining the price

If the demand shock turns out to be a permanent one, the profit at t will be lower than that of action (a) because the price is maintained at an inappropriate level. Besides, to avoid further loss of profit in later periods, the producer might have to change the price at some period t+j ($j \geq 0$) which will involve a cost $A + \xi(\Delta P_t)$.

On the other hand, if the demand shock turns out to be a transitory one, the profit at t will remain at its optimal level. Neither is it necessary to change the price in the future.

Table 2.1 summarizes the various costs and benefits of actions (a) and (b) under the

---

[3] As we will see the presence of a fixed cost per price change will lead to a very sophisticated formulation.

two possible cases of permanent and transitory demand shocks.

actions \ states	If the demand shock is permanent	If the demand shock is transitory
(a) changing the price by ΔP_t	(i) initial cost of raising price; (ii) maximum profit at t.	(i) initial cost of raising price; (ii) lower profit at t + need to reverse the price change later (or keep having the lower profit)
(b) Maintaining the price	(i) no initial cost of changing price; (ii) lower profit at t + need to change price later (or keep having the lower profit).	(i) no initial cost of changing price; (ii) optimum profit at t

Table (2.1)

(4) The producer is assumed to choose between action (a) and (b) so as to maximize the expected-sum-of-discounted profit. If the producer chooses action (a), he also needs to decide at which level the price should be raised. Hence, at the beginning of any period t, the producer's pricing decision involves two steps :

(i) Choosing the optimal ΔP_t to maximize the expected-sum-of discounted profit for action (a); and

(ii) Comparing the maximum profit of action (a) with that of action (b), and then decide to change or maintain the price accordingly.

(5) For simplicity, we assume a linear demand curve so that the producer's expected demand $\hat{Q}_t$ is :

$$\hat{Q}_t = \hat{\alpha} - \beta P_t ,$$

and the cost of production is :

$$C(\hat{Q}_t) = a + b\hat{Q}_t$$

Such simplifying assumptions are only for illustrative purpose. As we will see, formulations with more complicated demand and cost functions will still yield the same type of price sluggishness with respect to demand shocks.

2.1.2 The Formulation

Let γ be the discount rate. Suppose the price before any decision period t is $\overline{P}$ and the expected level with probability of demand is [4]. The maximum of the expected-sum-of-discounted-profit arising from actions (a) and (b) is defined as $f(\overline{P},\hat{\alpha})$ Hence, if the producer maintains the price at $\overline{P}$, the expected-sum-of-discounted profit will be

$$(\overline{P} - b)[\hat{\alpha} - \beta\ \overline{P}] - a + \gamma f(\overline{P},\hat{\alpha})$$

The first two terms of the above expression refer to the expected profit at period t; and the last term, by the Principle of Optimality in dynamic programming is the present discounted maximum of expected-sum-of-discounted profit for all periods after t.

If the producer changes the price to $\overline{P} + \Delta P$, the expected-sum-of-discounted profit will be

$$(\overline{P} + \Delta P - b)[\ \hat{\alpha} - \beta\ (\overline{P} + \Delta P)\ ] - a - [A + \xi\ (\Delta P)] + \gamma f(\overline{P} + \Delta P,\hat{\alpha})$$

Again, the first two terms of the above expression refer to the expected profit at period t. The third term is the cost of changing price by ΔP_t, and the last term is the present

---

[4]For simplicity sake, the time subscripts will be dropped in the remaining part of this chapter.

discounted maximum of expected-sum-of-discounted profit for all periods after t, with the new starting price $\overline{P} + \Delta P$.

Thus,

$$f(\overline{P},\hat{\alpha}) = \max \begin{cases} \sup\limits_{\Delta P \neq 0} (\overline{P}+\Delta P-b)[\hat{\alpha}-\beta(\overline{P}+\Delta P)]-a-[A+\xi(\Delta P)]+\gamma f(\overline{P}+\Delta P,\hat{\alpha}) \\ (\overline{P}-b)[\hat{\alpha}-\beta\overline{P}]-a+\gamma f(\overline{P},\hat{\alpha}) \end{cases} \tag{2.6}$$

where the first expression on the right hand side of equation (2.1) represent the supremum (maximum) return when the producer chooses to change price at t, and the second expression represents the maximum return when the producer choose to keep the price fixed at t. Whether the producer will choose to raise or maintain the price will depend on the relative size of the first and second expressions, which in turn depends on $\hat{\alpha}$. Putting $Q(P, \hat{\alpha})=(P-b)[\hat{\alpha}-\beta P]-a$, equation (2.1) can be written in the following compact form :

$$f(\overline{P}, \hat{\alpha}) = \max \begin{cases} \sup\limits_{\Delta P \neq 0} Q(\overline{P}+\Delta P, \hat{\alpha}) - [A+\xi(\Delta P)] + \gamma f(\overline{P}+\Delta P, \hat{\alpha}) \\ Q(\overline{P}, \hat{\alpha}) + \gamma f(\overline{P}, \hat{\alpha}) \end{cases} \tag{2.7}$$

Hence, the producer's pricing decision consists of two steps :

(i) Chooses the optimal $\Delta P^* \neq 0$ and calculates the supremum of the first expression; and

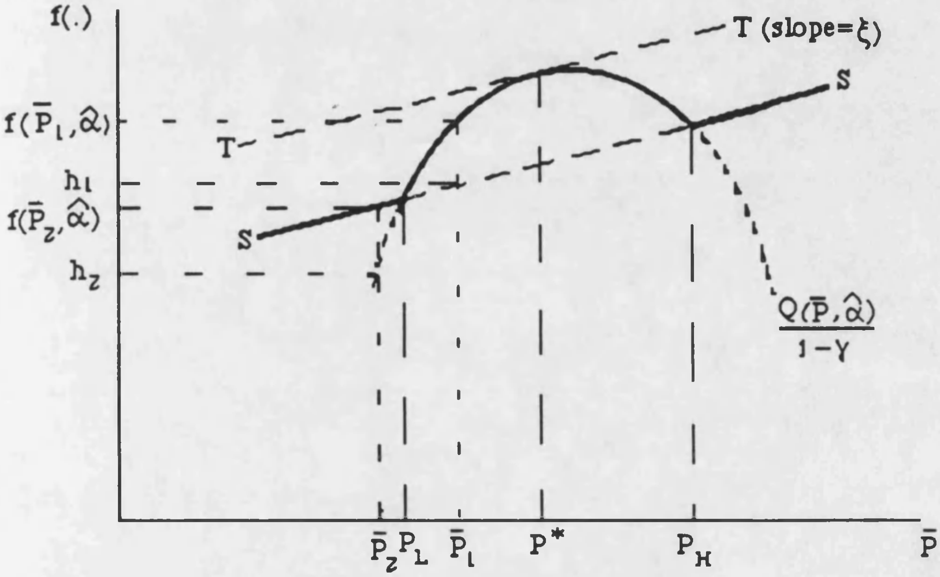
(ii) Compares the values of the first and second expressions. If the supremum of the first expression has a higher value, changes the price by ΔP^*. Otherwise, maintains the price at $\overline{P}$.

2.1.3 Outline of the Solution Result

As shown in Mathematical Appendix (I), equation (2.7) can be rewritten as

$$f(\bar{P}, \hat{\alpha}) = \max \begin{cases} \sup\limits_{\Delta P \neq 0} \dfrac{Q(\bar{P} + \Delta P, \hat{\alpha})}{1 - \gamma} - [A + \xi(\Delta P)] \\ \dfrac{Q(\bar{P}, \hat{\alpha})}{1 - \gamma} \end{cases} \tag{2.8}$$

In the same mathematical appendix, it is also shown that :



where $h_1 = f(P^*, \hat{\alpha}) - [A + \xi(P^* - \bar{P}_1)]$
$h_2 = Q(\bar{P}_2, \hat{\alpha})/(1-\gamma)$

diagram (2.1)

(i) the graphical shape of the second expression of equation (2.8) is represented by the parabola $[Q(\bar{P}, \hat{\alpha})/(1 - \gamma)]$ in diagram (2.1); while

(ii) the first expression of equation (2.8) is represented by the straight line SSS' with slope ξ in diagram (2.1).

From equation (2.8), we know that $f(\overline{P},\hat{\alpha})$ will be represented by the full-line curve in diagram (2.1).

The interpretation of the pricing decision is as follows :

(a) If the starting price lies within the curvature section of $f(\overline{P},\hat{\alpha})$ (i.e.between P_L and P_H), the producer will keep the price fixed[5]; and

(b) If the starting price is on the straight line section (i.e. below P_L or above P_H), the producer will change the price to P^*[6].

We have also shown that the shape and position of $f(\overline{P},\hat{\alpha})$ depends on :

(i) <u>the fixed cost per price change (A)</u>

The greater A is, the lower the SS line, the greater the gap between P_L and P_H, hence the less likely for the producer to change his price; and if the producer happens to change his price, the greater the size of each revision of price.

(ii) <u>the linear component (ξ)</u>

The lower ξ is, the flatter the straight line section of $f(\overline{P},\hat{\alpha})$ and the greater the values of P_L, P^* and P_H. However, as shown in Mathematical Appendix (II), the gap between P_L and P_H is independent of ξ.

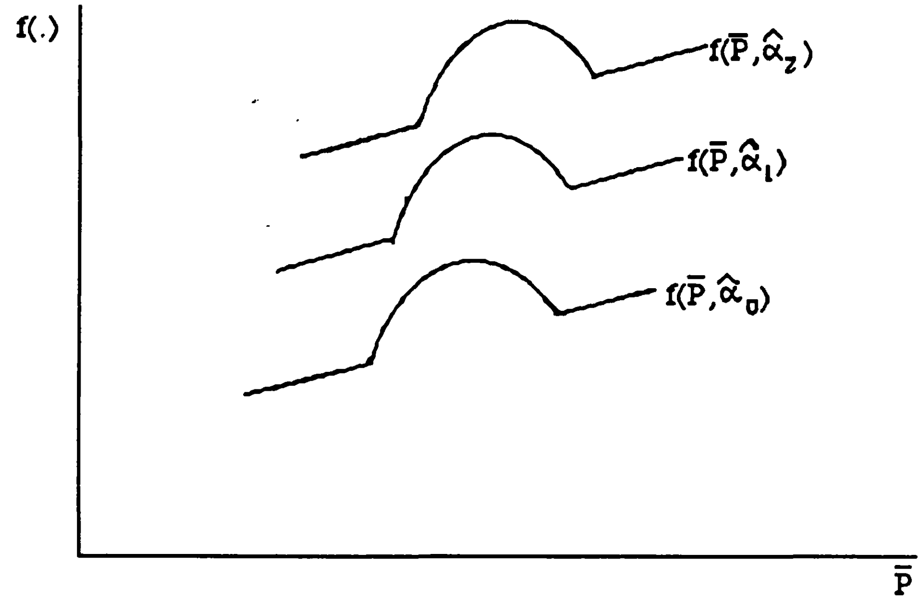
---

[5]The expected-sum-of-discounted profit for a starting price such as $\overline{P}_1$ will be represented by $f(\overline{P}_1,\hat{\alpha})$ in diagram (2.1). As shown in the same diagram, this is higher than that of raising price to P^* which will give $h_1 = \{f(P^*,\hat{\alpha}) - [A + \xi (P^* - \overline{P}_1)]\}$.

[6]The expected-sum-of-discounted profit for a starting price such as $\overline{P}_2$ will be represented by $f(\overline{P}_2,\hat{\alpha})$ which is equal to $\{f(P^*,\hat{\alpha}) - [A + \xi (P^* - \overline{P}_2)]\}$. As shown in diagram (2.1), this is higher than that of keeping the price at $\overline{P}_2$ which will give $h_2 = [Q(\overline{P}_2,\hat{\alpha}) / (1-\gamma)]$.

(iii) <u>the function $[Q(\overline{P}, \hat{\alpha})/(1-\gamma)]$</u>

This is in turn a function of β, a, b, γ and $\hat{\alpha}$. Of particular interest is that $Q(\overline{P}, \hat{\alpha})$, and hence $f(\overline{P},\hat{\alpha})$ is an increasing function of $\hat{\alpha}$. Thus, with an upward revision of expected demand from $\hat{\alpha}_0$ to $\hat{\alpha}_1$ and then to $\hat{\alpha}_2$, $f(\overline{P},\hat{\alpha})$ will be shifting to north east such as that shown in diagram (2.2) :



<u>diagram (2.2)</u>

In general, the greater the upward revision of $\hat{\alpha}$, the greater the shift in f(.).

Besides, we have shown that

$$P^* = \frac{\hat{\alpha} + b\beta - \xi(1-\gamma)}{2\beta} \qquad (2.4)$$

$$P_L = P^* - \sqrt{[A(1-\gamma)/2\beta]} \qquad (2.5)$$

$$P_H = P^* + \sqrt{[A(1-\gamma)/2\beta]} \qquad (2.6)$$

so that the gap between P_L and P_H is $2\sqrt{[A(1-\gamma)/2\beta]}$. Note,

(a) the greater the fixed cost per price change (A), the greater the gap between P_L and P_H, and the less likely for the producer to change his price;

(b) the greater is β, the smaller the gap between P_L and P_H, and hence the more likely for the producer to change his price. [The economic reasoning is as follows: the greater β is, the more sensitive are sales to a given change in price[7] and hence the more costly to maintain a suboptimal price (as against the fixed cost per price change).];

(c) if A=0, P_L and P_H collapse to P^*, and price will be adjusted immediately to any shock. This will be the case even if $\xi \neq 0$. Thus, the linear component alone cannot cause any price stickiness/sluggishness. As the gap between P_L and P_H is independent of ξ, neither will the linear component help to raise the price stickiness arising from the fixed component[8];

(d) the gap is independent of $\hat{\alpha}$, but P^* and the position of $f(\overline{P},\hat{\alpha})$ do.

(e) the "optimal" price change between any two periods of different $\hat{\alpha}$ is $\Delta P^* = (\Delta \hat{\alpha})/2\beta$.

---

[7]This is represented by a steeper turn in the parabola section of $f(\overline{P},\hat{\alpha})$.

[8]The only effect of ξ is on the value of P^*

2.2 Pricing Response to Various types of Specific Demand Shocks

Now, we come to use the basic model to explain the pricing response to various types of specific demand shocks.

2.2.1 Sluggishness with respect to a large permanent demand shock

Suppose the producer has experienced a long period of mixed occurrence of permanent and transitory demand shocks, then suddenly comes a permanent shock so that the mean level of demand rises from α_0 to α_0+u. Without other information, the producer cannot, at the beginning stage, tell from the sales whether the shock is permanent or not.

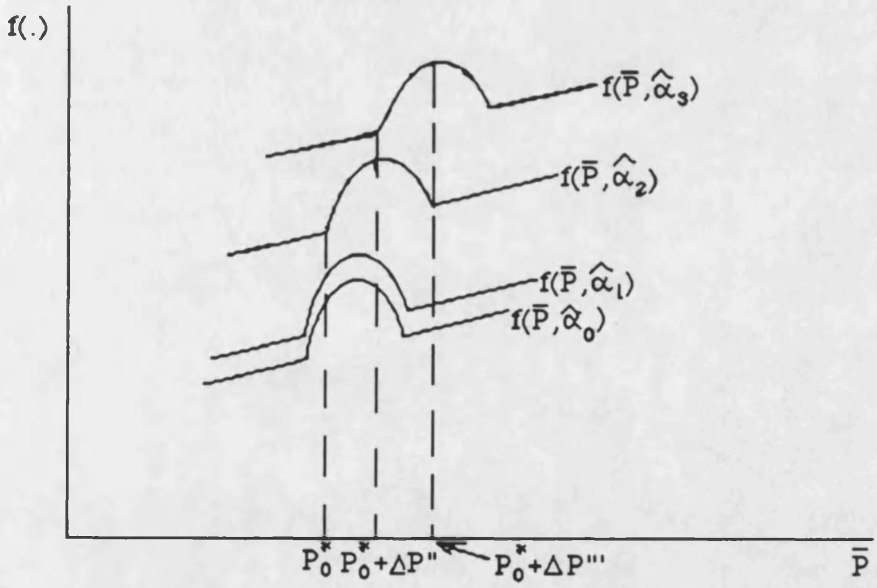


diagram (2.3)

According to equation (2.5), the producer will only make a small upward revision of $\hat{\alpha}$ which implies a very minor shift of f(.) from $f(\bar{P},\hat{\alpha}_0)$ to $f(\bar{P},\hat{\alpha}_1)$ in diagram (2.3). If the starting price is P_0^*, the producer will not raise the price. As time passes, the producer begins to get more observations on sales. If the shock is permanent, the sales will show favourable observations and the producer will start revising $\hat{\alpha}$ towards (α_0+u) which implies further shifts of f(.) to the north east. Sooner or later, f(.) will shift beyond $f(\overline{P},\hat{\alpha}_2)$. The producer will then raise the price by $[(\hat{\alpha}_2 - \hat{\alpha}_1)/2\beta]$ to $P_0^*+\Delta P''$. Thereafter, the producer will stick to the price until further observations suggests a continuous upward shift of f(.) to $f(\overline{P},\hat{\alpha}_3)$. This time, the producer will raise the price by $[(\hat{\alpha}_3 - \hat{\alpha}_2)/2\beta]$ to $P_0^*+\Delta P'''$. This process will continue until $\hat{\alpha}$ is revised to the permanent level (α_0+u). Hence, the producer's pricing response with respect to a permanent shock will be a stepwise function such as that shown in diagrams (2.4a)[9].

Note:

(i) The analysis can be extended to more general specifications of $\hat{\alpha}$. As long as $\hat{\alpha}$ increases with more observations of the permanent shock, the result will be qualitatively the same. One can also extend the analysis to the case in which the producer has extraneous information on the persistence of the shock (the extraneous information will cause an immediate jump in $\hat{\alpha}$).

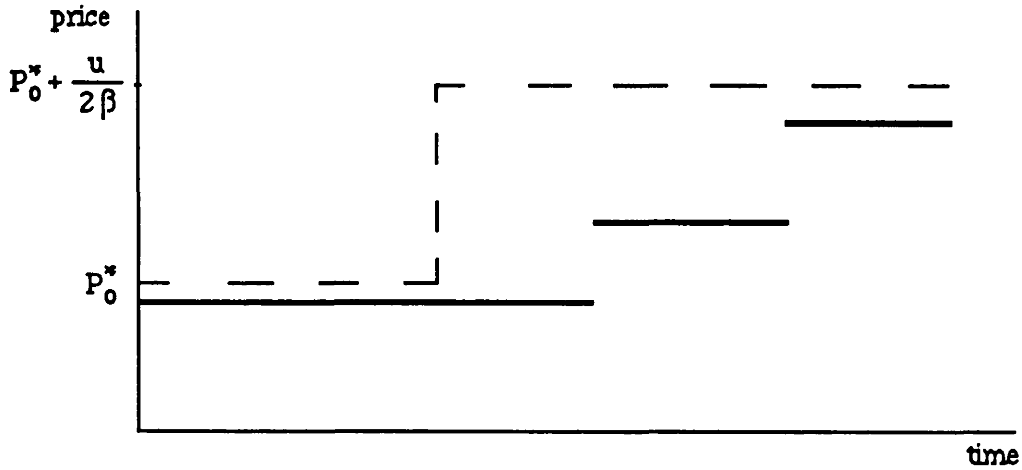
(ii) The size and length (and hence the pattern) of each step depend on how fast $\hat{\alpha}$ is adjusted and whether $\hat{\alpha}$ shows discrete jumps. (If $\hat{\alpha}$ is adjusted steadily, it implies the size of the steps will be the same).

(iii)In general, the more extraneous information a producer has, the better he can distinguish between a permanent shock from a transitory shock. In the extreme case, if, due to the extraneous information, he has complete certainty on the permanent

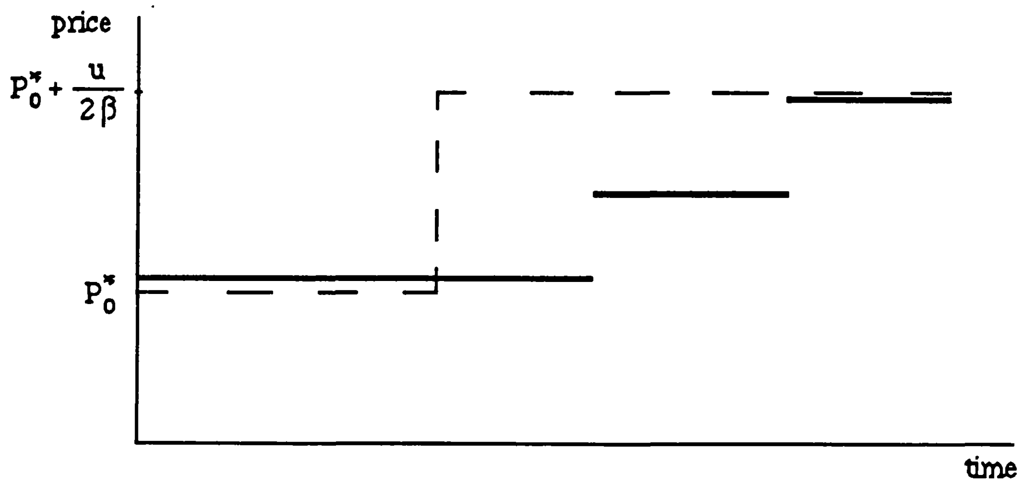
[9]Diagrams (2.4b) and (2.4c) are also drawn respectively for the cases where the starting price is lower than and greater than P_0^*.

shock, price will be adjusted immediately from P_0^* to $[P_0^*+u/2\beta]$.

(a) starting price at P_0^*



(b) starting price less than P_0^*



(c) starting price higher than P_0^*

<u>diagram (2.4)</u>

(iv) Unless the final price revision occurs at the time $\hat{\alpha}$ is revised to α_0+u, there is no necessity for the final price to reach $[P_0^*+u/2\beta]$. Indeed, it is likely that the final price revision occurs at somewhere $\hat{\alpha} < \alpha_0$+u which implies the final price will be somewhat below (above) $[P_0^*+u/2\beta]$ when u is positive (negative). Nevertheless, the gap will be limited by A, the cost of changing price. Otherwise, it will pay the producer to make another price revision.

(v) Diagram (2.4a) is drawn with the assumption that the starting price is P_0^*. Using the argument of (iv) above, the starting price can also be below or above P_0^* such as those shown in diagram (2.4b) and (2.4c).

(vi) If the producer is risk-averse (i.e. he also prefers <u>less variation</u> in return in addition to a higher <u>level</u> of expected return), he might not raise the price immediately at the time P_L of f(.) shift beyond the initial price $\overline{P}_0$. Instead, he might prefer to wait longer and raise the price by a larger step. In other words preference over less variation in return might intensify the stickiness of price. The formal proof of this is however beyond the scope of this thesis.

Finally, it is interesting to note that there are indeed two sources of sluggishness discussed in the model [10] :

(1) Sluggishness due to the pure signal extraction problem (without the cost of changing price); and

(2) Stickiness arises from the fixed cost of changing price.

The explanation is not difficult with the help of the model. Suppose there is no fixed cost per price change (ie A=0), from equations (2.10) and (2.11), P_L and P_H will collapse to

---

[10] There can also be other sources of price sluggishness such as the information transmission lags discussed in Chapter 1.

P_0^*. In addition, $\sup\limits_{\Delta P \neq 0} \dfrac{Q(\bar{P} + \Delta P,\ \hat{\alpha})}{1 - \gamma} - [A + \xi\,(\Delta P)]$ and hence $f(\bar{P},\hat{\alpha})$ will be represented by the tangent line TT of diagram (2.1). This is redrawn in diagram (2.5a). Suppose we have the permanent shock again. As $\hat{\alpha}$ is being revised upwards with more observations, the producer will raise his price continuously as shown by the dotted path CD in diagram (2.5b) [or C'D' in diagram (2.5a)]

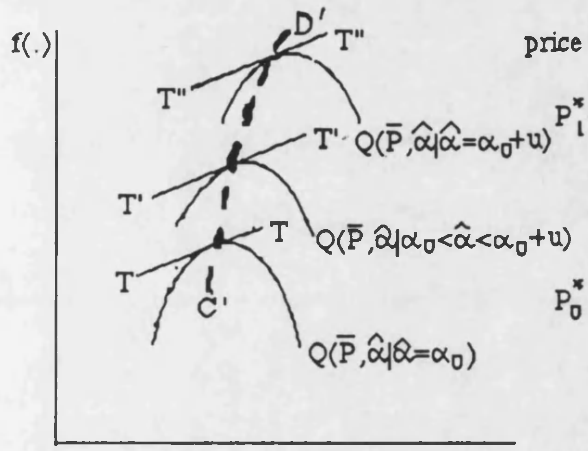


diagram (2.5a)

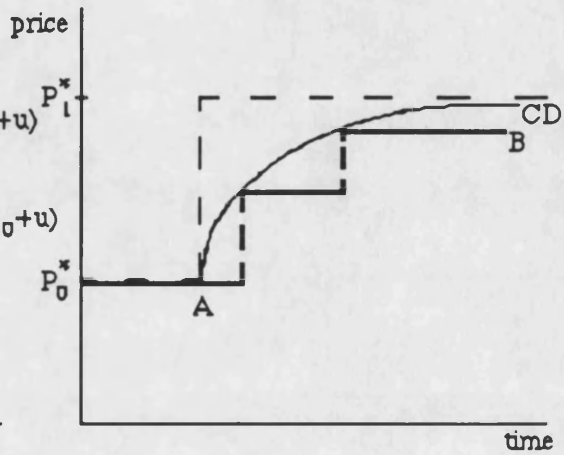


diagram (2.5b)

Thus, quite contradictory to that suggested by Mussa (1981), a cost of changing price is not a necessary condition for price sluggishness. There exists some other sources of price sluggishness, such as a confusion over the persistence of a stock.

For comparison purposes, the price path with the fixed cost of changing price is drawn in the same diagram (i.e. the stepwise path AB in diagram (2.5b)). There are two interesting interpretations with the diagram :

(i) With a signal extraction problem, the presence of the fixed cost of changing price will make the pricing response more sluggish (as shown by the gap between AB and CD);

(ii) Without a signal extraction problem, the representative producer, in the face of a large demand shock, will raise the price immediately to P_1^* even if there is a cost of changing price. The addition of a signal extraction problem, not only exerts its own effect on price sluggishness as shown by the gap between P_1^* and path CD, but also allows the fixed cost of changing price to cause further price sluggishness as shown by the gap between CD and AB.

Thus, the addition of a signal extraction problem produces a greater degree of price stickiness than that would be generated by the B-M-R models reviewed in Chapter 1.

2.2.2 Pricing Response to a Transitory Demand Shock

Let us suppose the demand shock discussed in diagram (2.3) is a transitory one. Again, without other information, the producer cannot, at the beginning tell from sales whether the sales shock is transitory or not. As usual, the producer will only make a small revision of $\hat{\alpha}$ in the first period. Thus, unless the shock (u) is really large, it is unlikely that $\hat{\alpha}$ in the first period is large enough to shift f(.) by such amount that the lower threshold of the new f(.) exceeds the starting $\overline{P}$ [11]. Hence, it is unlikely that the producer will raise the price in the first period. Given the shock is a transitory one, further evidence on sales will cause the producer to reduce the $\hat{\alpha}$ [12] which means that f(.) will be shifting back towards $f(\overline{P},\hat{\alpha}_0)$ and price will remain unchanged throughout the whole process. This explains why

[11] The only exceptions are that the transitory shock is so large or the starting price is near to the lower threshold of f(.) (i.e. the producer is at the margin of raising price) so that the (mistaken) upward revision of $\hat{\alpha}$ causes the lower threshold of f(.) to shift beyond the starting price.

[12] If there is another transitory shock at t,$\hat{\alpha}$ at the beginning of t+1 will be raised. However, by the definition of "transitory" the shock will vanish sooner or later. Hence, $\hat{\alpha}$ is expected to adjust back towards α_0.

a producer's pricing policy seldom reacts to a transitory shock.

Moreover, if we include the variance of return[13] (along with the expected return) into the producer's utility function, the producer might prefer to wait for more observations. Given the shock is a transitory one, more observations will result in a revision of $\hat{\alpha}$ back towards α_0. Thus, preference for less variation in return might imply a producer is less likely to raise price with respect to a transitory demand shock. Nevertheless, a formal proof of this is again outside the scope of this thesis.

[13]The inclusion of risk consideration into producer's pricing decision will mean a slightly more complicated formulation than the present model of expected return maximization. We will however, as mentioned earlier, have this as an interesting hypothesis for further research.

2.2.3 Pricing Response to a Small but Permanent Demand Shock

As before, f(.) will shift upwards with the rise in $\hat{\alpha}$. However, if u is too small, P_L of $f(\overline{P},\hat{\alpha}_0+u)$ might never exceed the initial price $\overline{P}_0$ [14] such as that shown in diagram (2.6) :

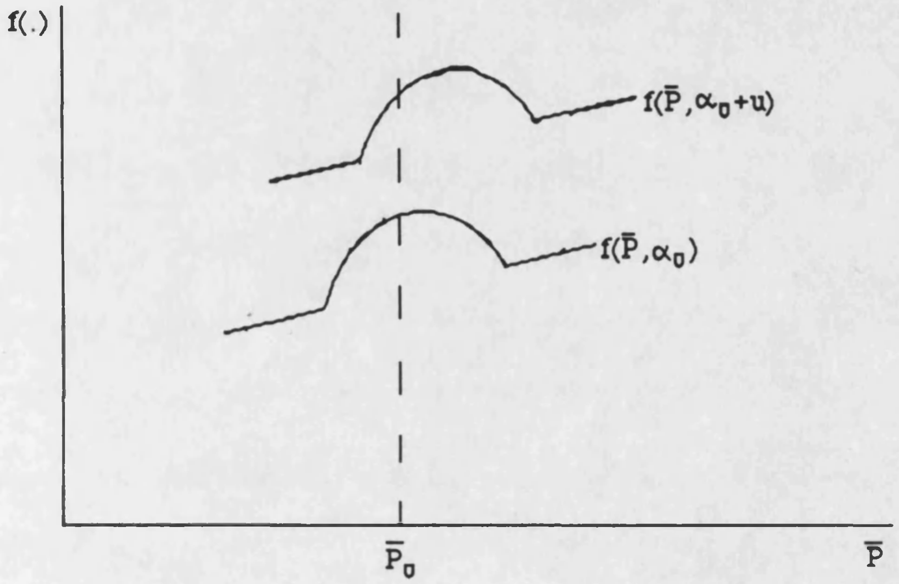


diagram (2.6)

In such case, even if we have a permanent shock so that $\hat{\alpha}$ will reach (α_0+u) sooner or later, the producer will not change his price with this permanent but minor shock[15].

[14]The only exception is when the producer is at the margin of raising price.

[15]It should, however, be noted that the small permanent shock has made the starting price $\overline{P}$ much closer to the margin of raising price. Its effect on price will be regained when some other shocks shift the lower threshold of f(.) beyond $\overline{P}$.

2.3 Solution for more general functional forms

In the previous section, we have derived our result by assuming the demand curve, the total cost curve and the cost of changing price are all linear. In this section, we will outline the solution procedure for more general functional forms.

2.3.1 Non-linear demand and total cost curves

If demand and total cost curves are non-linear, Q(.) of equation (2.7) may no longer be a parabola. Nevertheless, if one went through the solution procedure in Mathematical Appendix (I), one would find that the derivation from equation (2.7) to equation (2.8) will be applicable to all functional forms of Q(.), so long as there is a global maximum. Thus, one can derive a diagram similar to that of diagram (2.1) except that the new Q(.)/(1-γ) may no longer be a parabola. One such example is drawn in diagram (2.7) :

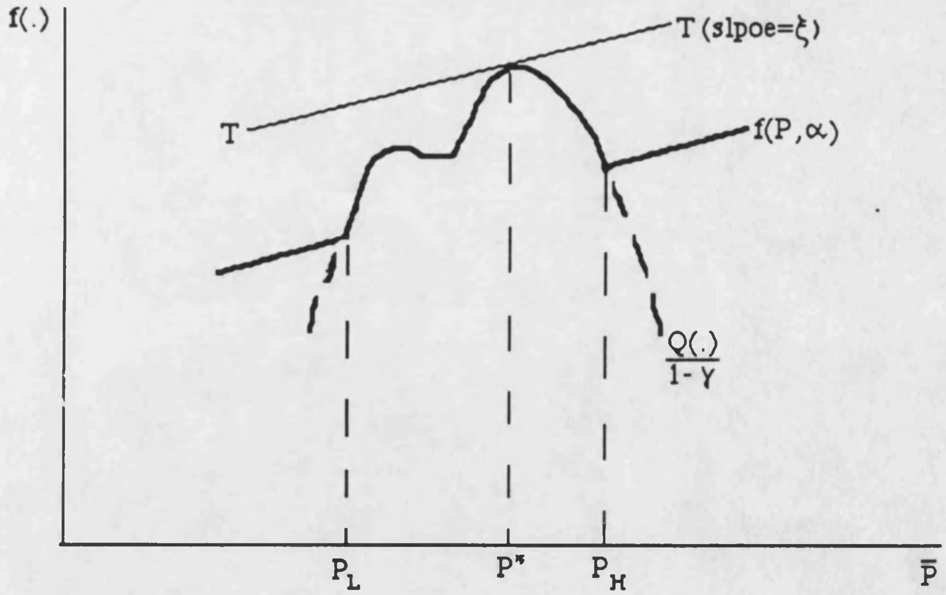


diagram (2.7)

Thus, the results will be basically similar to those discussed in section 2.1.3.

2.3.2 Non-linear cost of changing price

With a non-linear cost of changing price, the solution procedure will be more complicated. In addition, the degree of complication depends on whether $L(\Delta P_1+\Delta P_2)$ is greater than, equal to, or less than $L(\Delta P_1)+L(\Delta P_2)$. If we assume $L(\Delta P_1+\Delta P_2) \leq L(\Delta P_1)+L(\Delta P_2)$ for the relevant range of ΔP_1 and ΔP_2 [16], step (I) of Mathematical Appendix (I) still applies so that

$$f(\bar{P}, \hat{\alpha}) = \max \begin{cases} \sup_{\Delta P \neq 0} \dfrac{Q(\bar{P}+\Delta P, \hat{\alpha})}{1-\gamma} - L(\Delta P) \\ \dfrac{Q(\bar{P}, \hat{\alpha})}{1-\gamma} \end{cases} \tag{2.12}$$

where (i) $L(\Delta P)$ is any non-linear cost of changing price satisfying the condition $L(\Delta P_1+\Delta P_2) \leq L(\Delta P_1)+L(\Delta P_2)$. Examples of these are $\{\tilde{A}+\psi_1[\exp\{\psi_2(\Delta P)\}-1]\}$ and $[A + \xi(\Delta P) + \psi(\Delta P)^2]$ [17,18] with $\psi_1, \psi_2 << A$ and $\psi << A$ so that, for the relevant range of ΔP, $L(\Delta P) \leq L[\lambda\Delta P]+L[(1-\lambda)\Delta P]$ $\forall \lambda \leq 1$; and
(ii) $Q(P, \hat{\alpha}) = (P-b)[\hat{\alpha} - \beta P]$.

As $L(\Delta P)$ is no longer linear, the optimal (or desired) price for the first expression of

---

16Such condition states that, for the same total rise in price, it is always less costly to do it by one single large revision than by many small revisions. This implies, for the relevant range of ΔP_1 and ΔP_2, the fixed component will dominate the non-linear component. If this condition does not hold, the solution procedure will be even more complicated since we now have to compute an optimal price path, instead of an optimal price level, for all future periods. As we rarely observe changes of price in two consecutive periods, where the length can be as short as a month or a week, it seems that the assumption $L(\Delta P_1+\Delta P_2) \leq L(\Delta P_1)+L(\Delta P_2)$ is quite reasonable.

17The specification of exponential cost will be better than that of quadratic cost. This is so because the latter will give the unreasonable prediction that for some large enough negative ΔP_t, the reduction of price will imply a loss instead of a gain in reputation.

18It is also possible to assume the fixed component for $\Delta P_t<0$ is different from that for $\Delta P_t>0$.

equation (2.12) is no longer fixed but depends on the value of the starting price $\overline{P}$ [19] (i.e. there is an inter-temporal decision through the linkage between the optimal price and the starting price). This make a somewhat more complicated solution procedure and we outline it as follows :

(a) <u>First Step</u>

Our first step is to remove the fixed cost of changing from $Q(\overline{P}, \hat{\alpha})/(1-\gamma)$. This gives the curve $Q(\overline{P}, \hat{\alpha})/(1-\gamma) - A$ in diagram (2.8).

(b) <u>Second Step</u>

Above, we have seen that the desired price of the first expression of equation (2.12) depends on $\overline{P}$. Putting the argument the other way round, it implies that the price at any point such as E,F or G in diagram (2.8) will be the "desired" price for some corresponding $\overline{P}$. Pre-supposing the price at E, F or G as a potential optimal price, we draw the curves $Q(\overline{P},\hat{\alpha})/(1-\gamma)-L(\Delta P)$ for points E,F and G. For the case where $\{\bar{A}+\psi_1[\exp\{\psi_2(\Delta P_t)\}-1]\}$[20], these curves are shown by the dotted curves passing through E,F and G respectively. We then develop the envelope curve above this set of curves. Such an envelope curve (shown by the smooth full line curve AD of diagram (2.8)) indeed represents $\sup_{\Delta P \neq 0} Q(\overline{P},\hat{\alpha})/(1-\gamma) - L(\Delta P)$, the first expression of equation (2.12).

(c) <u>Third Step</u>

By taking the maximum of the envelope curve and $Q(\overline{P},\hat{\alpha})/[1-\gamma]$, $f(\overline{P},\hat{\alpha})$ is

---

[19]Differentiating $H(.) = [Q(\overline{P} + \Delta P, \hat{\alpha}) - a]/[1-\gamma] - L(\Delta P)$ with respect to ΔP and setting the derivatives to zero gives $[(P^*-b)(-\beta) + (\hat{\alpha} - \beta P^*)]/[1-\gamma] - L'(\Delta P^*) = 0$ where $\Delta P^* = P^* - \overline{P}$. Thus, as long as $L(\Delta P)$ is non-linear, $L'(\Delta P)$ will be non-zero and the optimal price P^* will depend on $\overline{P}$.

[20]The case for $L(\Delta P) = [A + \xi(P) + \psi(\Delta P)^2]$ will be similar and only differs in curvature to that shown in diagram 2.8.

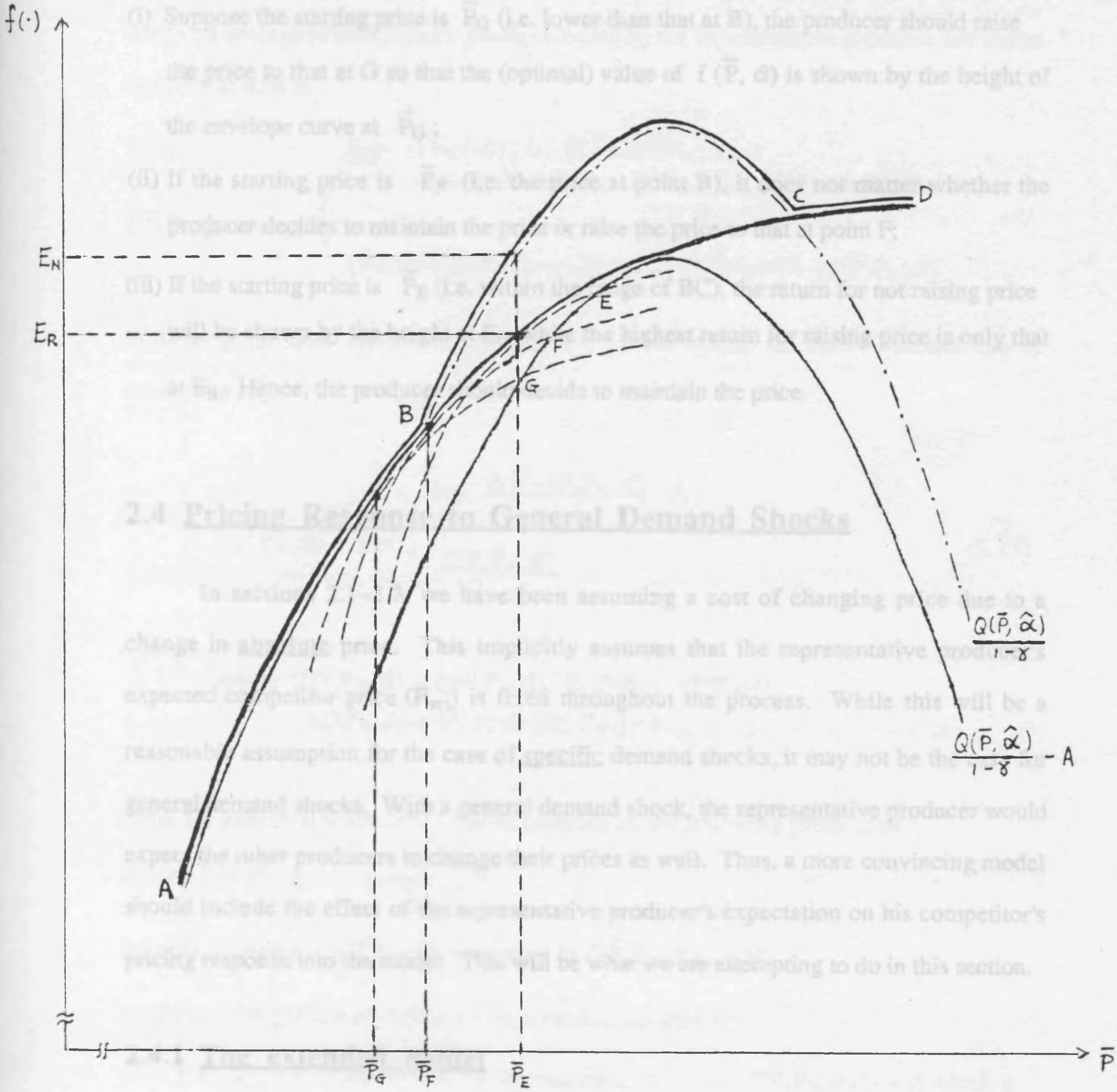


diagram (2.8)

indeed represented by the full line curve ABCD.

Thus, the pricing decision of the producer will be as follows :

(i) Suppose the starting price is $\overline{P}_G$ (i.e. lower than that at B), the producer should raise the price to that at G so that the (optimal) value of $f(\overline{P}, \alpha)$ is shown by the height of the envelope curve at $\overline{P}_G$;

(ii) If the starting price is $\overline{P}_F$ (i.e. the price at point B), it does not matter whether the producer decides to maintain the price or raise the price to that at point F;

(iii) If the starting price is $\overline{P}_E$ (i.e. within the range of BC), the return for not raising price will be shown by the height at E_N while the highest return for raising price is only that at E_R. Hence, the producer should decide to maintain the price.

2.4 Pricing Response to General Demand Shocks

In sections 2.1–2.3, we have been assuming a cost of changing price due to a change in absolute price. This implicitly assumes that the representative producer's expected competitor price (P_{wt}) is fixed throughout the process. While this will be a reasonable assumption for the case of specific demand shocks, it may not be the case for general demand shocks. With a general demand shock, the representative producer would expect the other producers to change their prices as well. Thus, a more convincing model should include the effect of the representative producer's expectation on his competitor's pricing response into the model. This will be what we are attempting to do in this section.

2.4.1 The extended model

Instead of assuming a linear cost of change in absolute price, we now assume a linear cost that depends on the deviation of relative price from a norm. Thus, we write the linear cost as $\xi[(P/P_w) - (P/P_w)^{\#}]$, where $(P/P_w)^{\#}$ is the relative price norm[21]. However, because of the menu cost and the producer's need to pledge some degree of constancy of nominal price, we assume a fixed cost per change in absolute price. If we let P_w be a weighted average of competitors' prices expected by the representative producer, the model can be rewritten as

$$f(\bar{P},P_w,\hat{\alpha}) = \max \begin{cases} \operatorname*{Sup}_{\Delta P \neq 0} (\bar{P}+\Delta P-b)[\hat{\alpha} - \beta(\bar{P}+\Delta P)/P_w] - a \\ \quad - \{A+\xi[(\bar{P}+\Delta P)/P_w - (P/P_w)^{\#}]\} + \gamma f(\bar{P}+\Delta P,P_w,\hat{\alpha}) \\ \\ (\bar{P}-b)[\hat{\alpha}-\beta(P/P_w)] - a - \xi[(P/P_w) - (P/P_w)^{\#}] + \gamma f(\bar{P},P_w,\hat{\alpha}) \end{cases}$$

Following the same solution procedure of the basic model, the above equation can be rewritten as :

$$f(\bar{P},P_w,\hat{\alpha}) = \max \begin{cases} \operatorname*{Sup}_{\Delta P \neq 0} \dfrac{Z(\bar{P}+\Delta P,P_w,\hat{\alpha})}{1-\gamma} - A \\ \dfrac{Z(\bar{P},P_w,\hat{\alpha})}{1-\gamma} \end{cases} \tag{2.13}$$

where $Z(P,P_w,\hat{\alpha}) = Q(P,P_w,\hat{\alpha}) - \xi[(P/P_w) - (P/P_w)^{\#}]$
$Q(P,P_w,\hat{\alpha}) = (P-b)[\hat{\alpha} - \beta(P/P_w)] - a$

Thus, the graphical shape of f(.) can be obtained by the following procedures :

(a) plot $Z(\bar{P},P_w,\hat{\alpha})/(1-\gamma)$ against $\bar{P}$;

(b) draw a tangent line TT at the maximum of $Z(\bar{P},P_w,\hat{\alpha})/(1-\gamma)$;

(c) draw a straight line SS below TT by a vertical distance A;

(d) $f(\bar{P},P_w,\hat{\alpha})$ can be represented by the maximum of SS and $Z(\bar{P},P_w,\hat{\alpha})/(1-\gamma)$ which is shown by the full line curve in diagram (2.9).

---

[21]For the case where producers are identical, $(P/P_w)^{\#}$ will be unity.

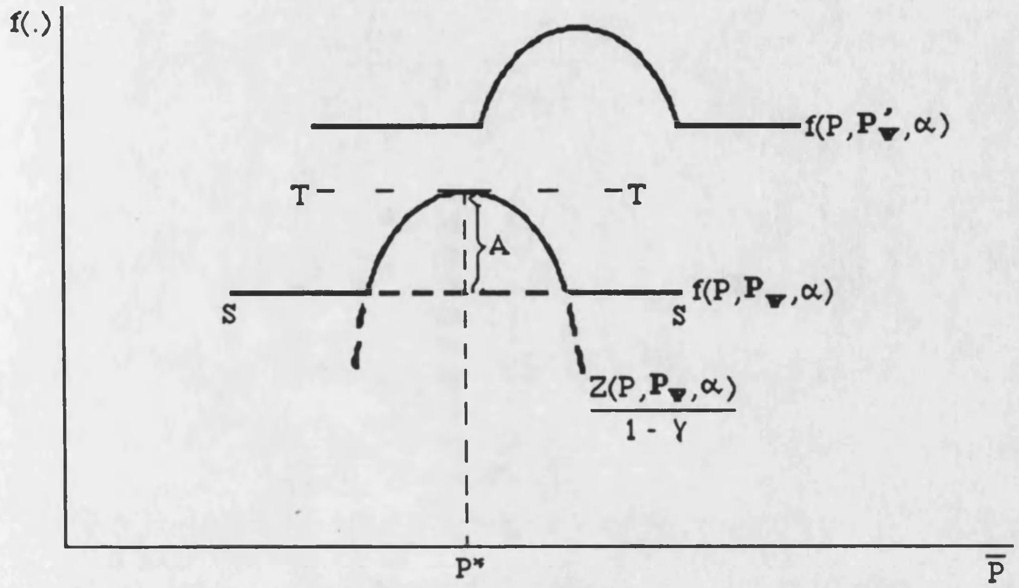


diagram (2.9)

It is interesting to note that in the above procedure of obtaining f(.) :

(i) $Z(\bar{P},\mathbf{P_w},\hat{\alpha})/(1-\gamma)$ instead of $Q(\bar{P},\mathbf{P_w},\hat{\alpha})/(1-\gamma)$ is plotted against $\bar{P}$;

(ii) the effect of the linear component $\xi[(P/\mathbf{P_w}) - (P/\mathbf{P_w})^{\#}]$ is included in $Z(.)/(1-\gamma)$ so that any rise in $Q(.)/(1-\gamma)$ or reduction in the linear component will cause a rise in $Z(.)/(1-\gamma)$; and

(iii) a rise in $\mathbf{P_w}$ to $\mathbf{P_w}'$ will shift $Z(.)/(1-\gamma)$ to the north east direction by raising $Q(.)/(1-\gamma)$ and reducing $\xi[(P/\mathbf{P_w}) - (P/\mathbf{P_w})^{\#}]$.

In other words, a rise in expected $\mathbf{P_w}$ with a general shock will make it more likely for producer to raise the price because

(a) the higher $\mathbf{P_w}$ implies that the desired price in $Q(.)/(1-\gamma)$ is higher; and

(b) the linear cost of having a higher price is reduced when $\mathbf{P_w}$ is higher.

We will further discuss this in section 2.5.

2.4.2 Signal Extraction Problem between General and Specific shocks as another source of Price Sluggishness

In the basic model, we have shown that the signal extraction problem between permanent and transitory shocks; and the cost of changing price are two important sources of price sluggishness/stickiness. Now we come to the third source of price sluggishness : signal extraction problem between general and specific shocks (referred as signal extraction problem (II) in later discussions). The three sources of price sluggishness will then be used to explain cost-oriented pricing (i.e. producer raise price fairly quickly with moderate cost shocks but react slowly to moderate demand shocks) in section 2.5.

To make the presentation simple, we will try to see how the third source alone [22] can cause a sluggish pricing response. We do so by assuming the following hypothetical case in which

(i) there is no cost of changing price;

(ii) there is a permanent general demand shock so that the mean level of demand rise from α_0 to α_0+u. Producers are completely certain that the shock is permanent, but are not sure whether the shock is general or specific; and

(iii) producers are identical [23].

---

[22]The interaction between the cost of changing price and signal extraction problem (II) will produce a more complicated pricing response. Nevertheless, it can be concluded the higher the fixed cost of changing price; and the less certain the producer about the persistence and generality of a shock, the more likely for a sluggish/sticky pricing response.

[23]Such assumption is only made for the sake of simple presentation. As the reader will be aware, the subsequent proof will go even for non-identical producers.

These assumptions imply that equation (2.13) can be simplified to

$$f(\bar{P},P_w,\alpha_0+u) = \operatorname*{Sup}_{P\neq\bar{P}} \frac{(P-b)[\alpha_0+u - \beta(P/P_w)] - a}{1-\gamma}$$

Solving for the optimal price (P^*) and adding time subscript t to P^* and $\mathbf{P_w}$, we have

$$P_t^* = b/2 + (\alpha_0+u)P_{wt}/2\beta \tag{2.14}$$

which implies that the optimal price P_t^* will depend on the expected competitor price index $\mathbf{P_{wt}}$.

(a) <u>If the producers recognize the shock is general</u>

If the producers know at the beginning , or getting to recognize with the passage of time, that the shock is general, then

$$\mathbf{P_{wf}} = P_f^*$$

where subscript f is defined as the "final period" when the producers get to recognize the shock is general.

Substituting this into equation (2.14), we have

$$\mathbf{P_{wf}} = \frac{b/2}{[1 - (\alpha_0+u)/2\beta]} \qquad \text{and} \qquad P_f^* = \mathbf{P_{wf}}$$

Thus, the permanent general shock, in the absence of signal extraction problems and cost of changing price, will cause an immediate rise of price from $P_0^* = \mathbf{P_{w0}} = [b/2]/[1-\alpha_0/2\beta]$ to $P_f^* = \mathbf{P_{wf}} = [b/2]/[1 - (\alpha_0+u)/2\beta]$.

(b) <u>With signal extraction problem about the generality of the shock</u>

However, with the co-existence of general and specific shocks and absence of extraneous information, the representative producer will, at the beginning, assign a low probability that the shock is general. This makes his expected competitor price index ($\mathbf{P_{wt}}$) less than $\mathbf{P_{wf}}$. With further observations from the general demand shock, the producers begin to revise $\mathbf{P_{wt}}$ upwards. Thus $\mathbf{P_{wt}} \geq \mathbf{P_{w,t-1}}$. Combining the two, the revision path of

$\mathbf{P_{wt}}$ must satisfy the following inequalities :

$$\mathbf{P_{w0}} \leq \mathbf{P_{w,t-1}} \leq \mathbf{P_{wt}} \leq \mathbf{P_{wf}} \qquad \forall\, t > 0 \qquad (2.15)$$

The adjustment speed of $\mathbf{P_{wt}}$ towards $\mathbf{P_{wf}}$ will depend on how fast the representative producer can recognize the shock as a general one. As long as it takes some time for $\mathbf{P_{wt}}$ to go from $\mathbf{P_{w0}}$ to $\mathbf{P_{wf}}$, equation (2.9) implies that P_t^* (the price chosen by the representative producer) will only adjust sluggishly from P_0^* to P_f^*. Thus, the signal extraction problem can also be a source of sluggish price movements.

2.5 The three hypotheses in the theory of Mark-up Pricing

In Chapter 1, we have argued that the theory of mark-up pricing is a more general theory than the B-M-R models. Beside being a theory of price decision, the modern version of mark-up pricing is also a theory involving capacity, production and quality decisions. It says quite a few things not recognized or properly tackled by the B-M-R models. In particular, implicit in the formula of $P_t = (1+m_t)AC_t$, there are indeed three hypotheses :

(a) the assumption about m_t (the planned profit margin) implies prices change sluggishly or remain unchanged with demand shocks;

(b) the separation of AC_t from m_t reviews the belief of an asymmetric pricing response with respect to cost shocks and demand shocks (prices rise relatively fast with cost shocks, but either change sluggishly or remain unchanged with demand shocks); and

(c) the unit elasticity of AC_t implies that 1% rise in average variable cost will cause 1% rise in price. (Note, the use of AC_t in the formula also reflect their belief that the average cost pricing is superior than the one period Neoclassical marginal cost pricing rule.)

Hypothesis (a) has indeed been justified/supported by the B-M-R models and our discussion in sections 2.2 – 2.4. The aim of this section is to check the robustness of the other two hypotheses with our model.

2.5.1 The Second Hypothesis

In sections 2.2 – 2.4, we have shown that

(i) signal extraction problem between persistent and transitory shocks;

(ii) signal extraction problem between general and specific shocks; and

(iii) the cost of changing price

lead to sluggish pricing response with respect to a persistent and general demand shocks.

The argument will, however, be weaker for the case of cost shocks :

> Unlike demand shocks, cost shocks are usually persistent and general – a rise in average variable cost is likely to affect the whole industry and persists in the near future. As a result, there is little uncertainty about the generality and persistence of the cost shocks, and this implies that there is little sluggishness arising from the two types of signal extraction problems. Besides, the generality of cost shocks implies producers will expect each other to raise price sooner or later. This raises the expected $\mathbf{P}_{wt}$ which implies a significant reduction in the reputation cost of raising price. These suggest that prices will rise fairly quickly with moderate cost shocks. Last but not least, if – because of the above reasons – price did rise fairly quickly with cost shocks in the past and had become a usual practice of the economy, everyone will find it acceptable [in addition to the "fairness" emphasized by Okun(1981)] to raise price with the rise in cost. In other words, the practice implies a further reduction in the cost of raising price with respect to cost shocks (as contrast to that with respect to demand shocks) which make it possible for price to rise with even smaller cost shocks.

The above arguments can be seen clearly with the help of our model and diagram. Rewriting our model for the case of a permanent and general cost shock and solving, we have

$$f(\bar{P},\mathbf{P}_w,b) = \max \begin{cases} \sup\limits_{\Delta P \neq 0} \dfrac{(\bar{P}+\Delta P-b)[\alpha-\beta(\bar{P}+\Delta P)/\mathbf{P}_w]-a-\xi[(\bar{P}+\Delta P)/\mathbf{P}_w-(P/\mathbf{P}_w)^{\#}]}{1-\gamma} - A_c \\ \dfrac{(\bar{P}-b)[\alpha-\beta(\bar{P}/\mathbf{P}_w)]-a-\xi[(\bar{P}/\mathbf{P}_w)-(P/\mathbf{P}_w)^{\#}]}{1-\gamma} \end{cases} \quad (2.16)$$

where : A_c is the fixed cost of raising price with respect to cost shocks which will be lower than that with respect to demand shocks (i.e. $A_c < A$) when cost-oriented pricing has become a practice in our economy.

To illustrate the point, we draw the cases of a reasonably large and a moderate cost shock in diagram (2.10a) and diagram (2.10b) respectively :

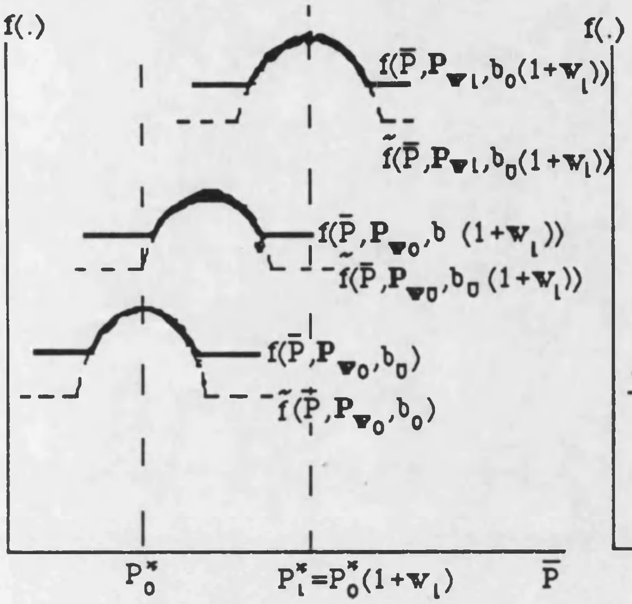


diagram (2.10a)

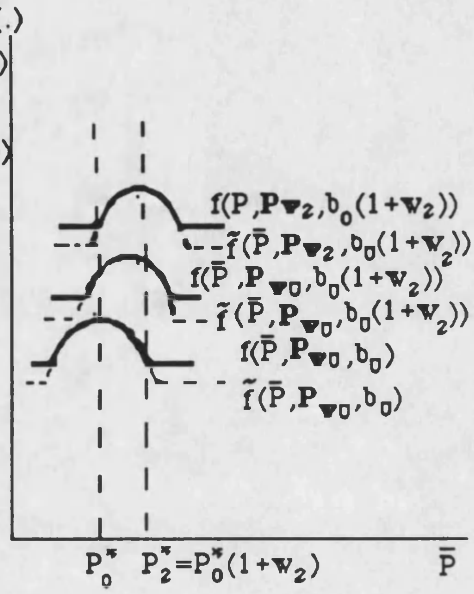


diagram (2.10b)

First consider the case of large cost shock in diagram (2.10a). Without any change in the expected competitor price index, a rise in cost from b_0 to $b_0(1+w_1)$ will shift $\tilde{f}(.)$ from $\tilde{f}(\overline{P},\mathbf{P}_{w0},b_0)$ to $\tilde{f}(\overline{P},\mathbf{P}_{w0},b_0(1+w_1))$. However, if the shock is known to be general, producers will expect each other to raise price. This cause a rise in $\mathbf{P}_w$ which implies further shift to $\tilde{f}(\overline{P},\mathbf{P}_{w1},b_0(1+w_1))$. Besides, if cost-oriented pricing has long been a practice of the economy, A_c will be much smaller than A, and f(.) instead of $\tilde{f}(.)$ in the above diagrams will be the relevant profit curves. If, as suggested in the verbal argument, there is no uncertainty about the persistence and generality of the cost shock, the shift from $f(\overline{P},\mathbf{P}_{w0},b_0)$ to $f(\overline{P},\mathbf{P}_{w1},b_0(1+w_1))$ will be immediate. As P_0^* lies outside the curvature section of $f(\overline{P},\mathbf{P}_{w1},b_0(1+w_1))$, the cost shock will cause an immediate rise of price to P_1^*.

Thus, the absence of signal extraction problems suggest that price will be adjusted immediately with a reasonably large cost shock.

Once the cost-oriented pricing has become a practice and producers usually find it right to expect a rise in $\mathbf{P}_{wt}$ with a cost shock, it will be more likely for price to rise with moderate cost shock. This is illustrated in diagram (2.10b) in which the higher $\mathbf{P}_w$ [24] and the lower A_c make it just possible for price to be raised with the moderate cost shock w_2. Thus, the second hypothesis in the theory of mark-up pricing (asymmetric pricing response with respect to cost shocks and demand shocks) is supported by our model. Nevertheless, it must be noted that cost-oriented pricing is only a good approximation of the actual pricing behaviour of producers. There are at least two reasons in saying so :

(a) A strict attachment of cost-oriented pricing by producers requires the price to be adjusted with every change (no matter how small and frequent) in cost. However, for any cost shock that is less than that shown in diagram (2.10b), the producer will choose to keep the price unchanged. The reasoning is simple : because of supplier's need to pledge constancy of price for a reasonably long period so as to encourage customers to rely on inter-temporal comparison and return to shop, it is better not to raise the price until the cumulated rise in cost has reached some reasonable amount. Thus, the need to pledge constancy of price for some time reflect that A_c will not be too small, albeit smaller than that of demand shock; and

(b) If there had been some long term change in demand before the cost shock so that the optimal price is somewhat below or above the starting price P_0^* shown in diagrams (2.10a) and (2.10b), the producer will change the price, in addition to that arising from the change in cost, by another amount that reflect the change in demand. This is

[24]As explained in section 2.4.1, the rise in $\mathbf{P}_w$ will make it more likely for producer to raise the price because

(i) the desired price in $Q(.)/[1-\gamma]$ is higher; and

(ii) the linear cost of having a higher price is reduced with a rise in $\mathbf{P}_w$.

possible because customers do not have perfect information about the actual change in cost (c.f. section 1.2.1(D)(3) in Chapter 1).

2.5.2 The Third Hypothesis

Suppose prices do respond fairly quickly to moderate cost shocks, our next question is to check whether price will rise equiproportionately with the cost shock. Solving for the optimal price in the first expression of equation (2.16), we have

$$P^* = b/2 + (\alpha P_w)/2\beta - \xi/2\beta \qquad (2.17)$$

where ξ is the linear cost coefficient which will "on the average " rise by x% when there is a x% rise in cost and general price level.

Suppose there is a general rise in cost by x% and every producer expects P_{wt} to rise by x% as well, the sum of the first two terms on the right hand side of equation (2.17) will rise by x% as well. As ξ also "on the average " rises by x%, P^* will rise by x% which in turn justify producers' initial expectation of a x% rise in P_{wt}. This leads us to conclude that the unitary power index of AC_t in the formula $P_t = (1+m_t)AC_t$ is a good approximation of optimal pricing behaviour in the case of a general cost shock. However, the approximation will be a bad one to the case of specific cost shocks. Suppose there is a specific cost shock (due to, say, a technological improvement specific to the firm) that changes b. Since the shock is a specific one, the producer will not expect any change in P_w. According to equation (2.17), the percentage change in price will be different from that of cost. Hence, the approximation will be a bad one in case of specific cost shocks. Nevertheless, this should not be too discouraging to the use of the mark-up pricing equation in macroeconomics because the size of inflation in our world implies general cost shocks (i.e. rising P_w, b_t and ξ_t with inflation) instead of specific cost shocks are dominating [25].

---

[25]The solution in case of inflation is discussed in section 2.6.

2.6 Remarks in the case of expected inflation

In the previous sections, we have seen the flexibility of our model and its power in explaining pricing behaviour with respect to various types of once-and-for-all demand/cost shocks. There is however one important limitation associated with the complexity of solution [26] in case of continuously rising prices. For example, in equation (2.13) or (2.16), the most general equation in the previous sections, we have assumed that the representative producer is expecting only a mean value of competitor price. While this may be a reasonable assumption for the case of once-and-for-all demand/cost shocks, we need to assume a path of $\mathbf{P_w}$ for the case of continuous rising cost and competitor price. In this case, equation (2.8) has to be replaced by the following equation :

$$f(\bar{P}_t,\{\mathbf{P}_{w,t+i|t}\},\{b_{t+i|t}\},\{a_{t+i|t}\},\alpha) = \max \begin{cases} \sup\limits_{\Delta P_t \neq 0} (\bar{P}_t+\Delta P_t-b_t)[\alpha-\beta(\bar{P}_t+\Delta P_t)/\mathbf{P}_{wt}] - a \\ \qquad - \{A+\xi[(\bar{P}_t+\Delta P_t)/\mathbf{P}_{wt} - (P/\mathbf{P}_w)_t^{\#}]\} \\ \qquad + \gamma f(\bar{P}_t+\Delta P_t,\{\mathbf{P}_{w,t+i|t}\},\{b_{t+i|t}\},\{a_{t+i|t}\},\alpha) \\ \\ (\bar{P}_t-b_t)[\alpha-\beta(\bar{P}_t/\mathbf{P}_{wt})] - a \\ \quad - \xi[(\bar{P}_t/\mathbf{P}_{wt}) - (P/\mathbf{P}_w)_t^{\#}] \\ \quad + \gamma f(\bar{P}_t,\{\mathbf{P}_{w,t+i|t}\},\{b_{t+i|t}\},\{a_{t+i|t}\},\alpha) \end{cases}$$

where $\{X_{t+i|t}\}$ is the sequence of X_{t+i} expected at the beginning of period t, and X_{t+i} can be $\mathbf{P}_{w,t+i}, b_{t+i}$ or a_{t+i}.

Unfortunately, solution of the above equation is still a mathematical problem. Nevertheless, approximate solutions for some special cases can be computed. For example, if we assume that (a) there is no demand shock; and (b) inflation is constant, the problem will be the same as that in Mussa (1981). The path of actual price ($\ln P_t$) will have pre-adjustments and under-adjustments around the path of "desired" price ($\ln P_t^*$) such as

[26] Note, the problem arising from the complexity of solution rather than the formulation of the model.

that shown in diagram (2.11) :

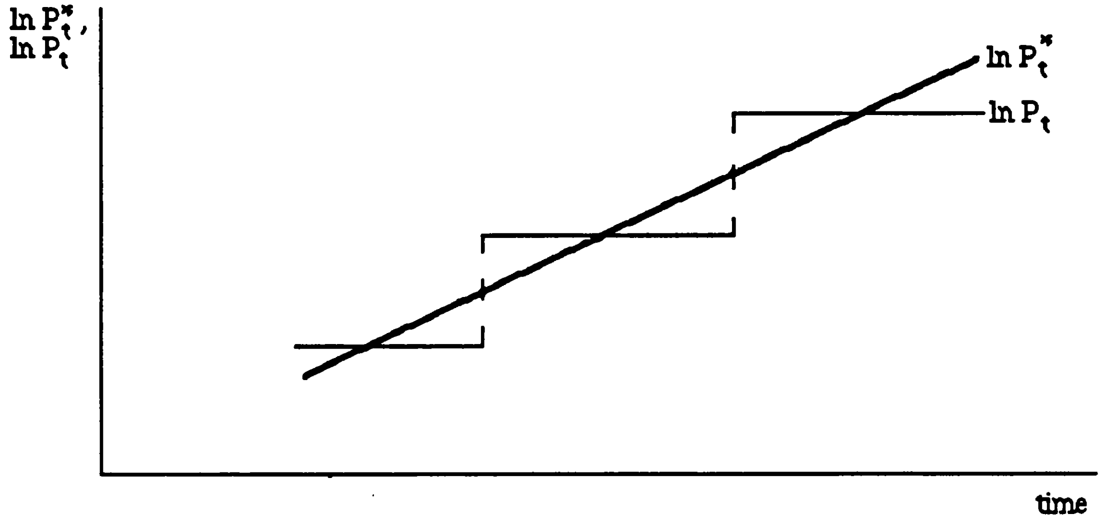


diagram (2.11)

In Chapter 4, we will try to approximate the solution for the case with (a) demand shocks; and (b) constant expected inflation.

2.7 Conclusions and Remarks

2.7.1 Mark-up Pricing as an approximation of Actual/Profit Maximizing Pricing Behaviour

In this Chapter, a dynamic programming model was built to check the following three hypotheses in the theory of mark-up pricing:

(a) a sluggish/sticky pricing response with respect to demand shocks;

(b) the pricing response with respect to cost shocks would be much faster than that with respect to demand shocks;

(c) a unit elasticity of average cost (AC) in the formula P=(1+m)AC.

With regard to the first hypothesis, we have analyzed three important sources of individual price stickiness/sluggishness with respect to demand shocks. These are the signal extraction problem between persistent and transitory shocks; the signal extraction problem between general and specific shocks; and the cost of changing price. Within the third source, the fixed component will cause a bang-bang solution : price will remain sticky for moderate fluctuations in demand; and the percentage change of price will show discrete jumps once the cumulated demand shocks exceed the threshold. Interestingly enough, the linear component was found to have no effect on the price stickiness. This is so because the cost of changing price arising from such a component will be the same as long as the total ΔP are the same, no matter whether the producer makes the ΔP by one large change or many small changes. Thus, a large cost of changing price does not necessarily imply a high degree of price stickiness. If most of the cost is due to the linear component, price will not be very sticky despite a large cost of changing price.

Perhaps a comparison with the existing literature will help to reveal the contribution/significance of our model in explaining the extensive degree of price stickiness/sluggishness in the world. For example, in the review of the menu cost hypothesis in Chapter 1, we have pointed out two important limitations in these models:

(i) there is nothing in these models to guarantee that the "near-rational" agent starts at the optimal price (ie a crucial assumption without which the envelope theorem will not be applicable); and

(ii) these models are usually for but a single period, and a replacement of the single period framework by a multi-period setting would imply that the individual loss from inaction may no longer be of second order (ie we need something more than the menu cost to explain the inaction of the producer).

The first limitation suggests that an explicit specification of both the cost of changing price and the nature of the trade off between action (changing the price) and inaction (not changing the price) are necessary. The model in this Chapter was built with this need in mind, and it demonstrates that the representative producer does not usually have the starting price at the optimal level. With regard to the second limitation, we emphasize the role of a significant reputation cost, which will be supported by the empirical work discussed in Chapter 4, in balancing off a possibly first order loss of inaction.Thus, although the menu cost hypothesis may create the interesting possibility that a small menu cost may cause a significant degree of price stickiness, we believe that the reputation cost is a more important source of price stickiness.

The model also helps to highlight a few limitations in the B-M-R models. For example, it suggests that (i) the overemphasis of a quadratic cost of changing price in Rotemberg (1982a,b) has led to an unrealistic solution of pricing response to a demand shock such as that shown in diagram (1.6) of Chapter 1; (ii) the solution in Mussa (1981) is not suitable for any once-and-for-all permanent demand shock; and (iii) the two bins policy in Barro (1972) may fail to be an optimal policy whenever sales are affected by both the permanent and transitory demand shocks. However, the most innovative contribution of our model over the B-M-R models is its capacity in dealing with the signal extraction problem. As highlighted in section 2.2, the signal extraction problem not only causes a certain degree of price sluggishness with respect to a permanent demand shock by itself [eg the gap between CD and P_1^* in diagram 2.5(b)], but also allows the fixed reputation cost of changing price to cause a further degree of price stickiness (eg the gap between AB and CD of the same diagram). Without the signal extraction problem, the fixed reputation cost of changing price will not be able to explain any price stickiness to a large enough permanent demand shock. Thus, our emphasis of a significant reputation cost of changing price, the signal extraction problems and their interactions have greatly extended the degree of price stickiness (with respect to demand shocks) that could be explained by existing models.

With regard to the second hypothesis, it was found that the three sources of price stickiness/sluggishness will be much weaker in the case of cost shocks:

> Unlike demand shocks, there is less uncertainty about the generality and persistence of cost shocks. Besides, the rise in expected competitive price associated with the generality of cost shocks; and the wide acceptance of cost-oriented pricing imply the cost of raising price with respect to cost shocks will be much lower than that with respect to demand shocks.

In section 2.5.1, the above argument were put in mathematical terms and it was shown that the size of cost shocks required to induce a rise of price would be, ceteris paribus, much smaller than that of demand shocks [27]. Thus the hypothesis that prices will be more responsive to cost shocks is justified.

The work in this Chapter also provides some hints on the evolution process of the "so called" cost-oriented pricing. Unlike Okun (1981) who attempts to use the concept of

[27]"Same size" here refers to sizes of various shocks that will give the same % change in the desired level of P*.

"fairness" to explain why price will be raised with respect to a cost shock but not a demand shock, we argue that this is unlikely to be the original reason for the evolution of cost-oriented pricing instead of a demand-oriented pricing [ie if demand-oriented pricing were ever evolved and widely accepted as a practice of the economy, consumers will find it "normal", if not "fair", even when prices are raised with demand]. Instead, we emphasize the persistence, the generality and the associated rise of expected competitor price with cost shocks as the major reasons for the evolution of cost-oriented pricing. We then proceed to integrate Okun's argument into our model by suggesting that once the practice of cost-oriented pricing instead of demand-oriented pricing is evolved, the argument of "fairness" will imply A_c is much smaller than A so that it is more likely for price to respond to cost shocks than demand shocks.

It must however be noted that the reasons we list here are only part of the story. For example, if we replace the assumption of profit maximization by some kind of satisficing/ inertial behaviour, inflation can be another reason for the evolution of cost-oriented (instead of demand-oriented) pricing [ie producers have to raise price (sooner or later) with the ever rising costs; but there is no corresponding force to push the producers to raise price with demand shocks]. While developing such an argument is beyond the scope of this thesis, it is certainly an interesting area of further research.

While it is true that the pricing response with respect to cost shocks would be much faster than that with respect to demand shocks, our model also suggests that:

(a) As long as $A_c>0$, prices will not be adjusted with very small change in cost; and

(b) If price is ever changed, the change will take into account the change in permanent demand as well as the change in cost.

These imply that the cost-oriented pricing proposed in the literature may not be a perfect description of profit maximizing or even actual pricing behaviour (ie it is at most an approximation). However, it appears to the writer that the approximation may still be a good one as long as there is a moderate inflation. This is because:

(i) with a moderate inflation rate, the small change in cost [mentioned in (a)] will be accumulated until there is a rise in price. Once the price is raised, the effect of the small changes in cost will be reflected in the change in price; and

(ii) with a moderate inflation rate, the effect of cost change on price will dominate that of demand shocks so that the problem mentioned in (b) is significantly weakened.

However, a formal proof of this cannot be done until a mathematical solution for the case with inflation is available (c.f. section 2.6).

In section 2.5.2, we also checked the third hypothesis by assuming that prices do respond fairly quickly to moderate cost shocks. It was found that the unit elasticity of AC in P=(1+m)AC will be a bad approximation in the case where specific cost shocks turned out to be dominating. Nevertheless, once again, the inclusion of a moderate inflation rate will tend to make general cost shocks dominating and hence help to justify the use of mark-up pricing as a good approximation in macroeconomic analysis.

The checking of the three hypothesis here further support the notions that average cost pricing can (i) originate in profit maximizing behaviour; and (ii) be superior than short run Neoclassical marginal pricing. From the above discussions, we also believe that the existence of a moderate inflation will ensure mark-up pricing as a good approximation to actual/profit maximizing pricing behaviour in the customer market. However, a formal proof of this has yet to be developed.

2.7.2 An Overview of Pricing Behaviour in Various Types of Markets

With the help of our model, let us make a general overview on the behaviour of price in various types of markets. First consider the case of customer markets where there is significant cost of changing price. As expected, we often find that prices of televisions, hair cuts, cinema tickets and etc are usually fixed for about half to one year period. Once the prices are raised, they are usually in the range of five to twenty percent. Indeed, our theory does predict that once the producer decides to change the price, the percentage change will be at least greater than some minimum size which is in turn a function of the elasticity of demand and the cost of changing price of the particular firm [28]. It is however possible to have the percentage price change anywhere [29] above the minimum size. For example, if the shock at that period is particularly large, such as that in the oil crisis, the percentage price change will be high. Besides, the price changes will be more dispersed if the cost of changing price do vary at

(a) different months simply because the producer has in the past developed a practice of changing price at some particular months; or

(b) different prices because of non-proportionate psychological feelings about numbers (eg. cost of changing price at £99 might be higher than that at £89 or £90).

Next consider the market of financial assets (such as shares, bonds, foreign exchanges and etc) where there is almost no cost of changing price. As predicted by our theory, price changes are frequent and it does not seem to be of any minimum size of price change except for the reason of visibility. The case of interest rates decision in Hong Kong is somewhere between the above two cases. As there is a cost for representatives to attend the meetings,

---

[28]If, for some particular period, the producer gets a very good reason to raise price (such as oil or material surcharge in cases of temporary rise in material and fuel cost), the percentage change of price might be lower than the minimum size because the cost of changing price at this period is particularly low.

[29]The upper limit will depend on the cumulated change of demand and cost between the present and previous price changes.

interest rates are usually[30] reviewed fortnightly (instead of hourly or daily). Before the next revision, quantity (amount of deposit or loan) will be the variable of adjustment. However, as there is no reputation cost and menu cost of changing interest rate, there is almost no minimum size of change in interest rate.

2.7.3 Extensions to other types of Decisions

Finally, we would like to emphasize that our model can be modified to analyze many decisions other than price and output. For example : a fixed component of changing capacity can be used to explain why changes in capacity are discrete; a relatively high cost of unsatisfied demand (compared with the cost of having inventory or excess capacity) can explain why producers will try their best to avoid unsatisfied demand; a fixed component in the cost of layoff explains why producers hesitate to retrench workers in the case of moderate reduction in demand. Also, a fixed component of the reputation cost of changing quality can be used to explain the stickiness of quality for most manufactured products[31]. For some special cases such as restaurant services or newspapers where the additional cost of changing quality [32] (such as giving more food for each portion; having better quality of papers) is negligible, we might find that qualities are raised at the beginning of every rise in price so that the effect of price change on demand is smoothened. In the subsequent Chapter, we will go into the details of the employment decision.

---

[30]Except for the emergency case where the cost of not adjusting interest rates is higher than the cost of holding an emergency meeting.

[31]When there is a significant technological innovation where the gain from improving the quality is enormous, the quality improvement will be carried out. Again, the jump in quality will be discrete.

[32]Note that according to our definition, there is a cost of having higher quality, but there is no additional cost of changing the quality in these industries.

Mathematical Appendix (I): Solution of equation (2.7)

In this appendix, we come to the solution of the basic model (equation (2.7)) whose properties are outlined in section 2.1.3. Repeating equation (2.7) here:

$$f(\overline{P}, \hat{\alpha}) = \max \begin{cases} \sup\limits_{\Delta P \neq 0} Q(\overline{P} + \Delta P, \hat{\alpha}) - L(\Delta P) + \gamma f(\overline{P} + \Delta P, \hat{\alpha}) \\ Q(\overline{P}, \hat{\alpha}) + \gamma f(\overline{P}, \hat{\alpha}) \end{cases} \qquad \text{(A1)}$$

where $Q(P, \hat{\alpha}) = (P-b)(\hat{\alpha} - \beta P) - a \quad \forall P$; and
$L(\Delta P) = A + \xi(\Delta P)$

We note that $Q(P, \hat{\alpha})$ is a quadratic function in P, with the coefficient of P^2 being $-\beta$, a negative number. Hence $Q(P, \hat{\alpha})$ is a function, concave downward, with a unique maximum occurring at some P_0:

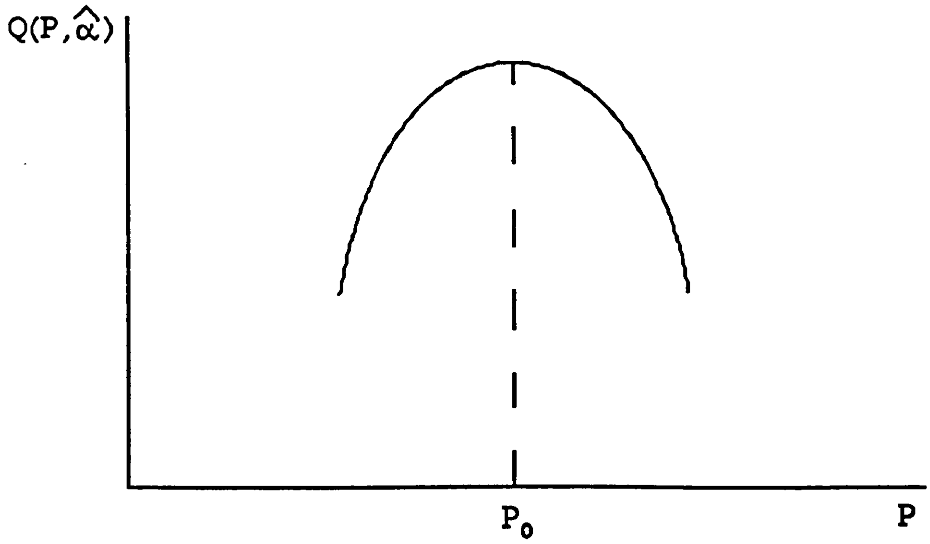


diagram (A1)

Since the producer's pricing decision involves the possibility of switching from the regime of maintaining price to that of changing price, or vice versa, $f(\overline{P}, \hat{\alpha})$ may sometimes take the value of the first expression and sometimes the second expression in (A1). Moreover, both expressions are in turn a function of $f(\overline{P}, \hat{\alpha})$ or $f(\overline{P}+\Delta P, \hat{\alpha})$. These make the interrelationship between $f(\overline{P}, \hat{\alpha})$ and the two expressions extremely complicated. Our solution procedure will involve the following two steps:

(1) We first prove that (A1) can be rewritten as:

$$f(\overline{P}, \hat{\alpha}) = \max \begin{cases} \sup_{\Delta P \neq 0} Q(\overline{P} + \Delta P, \hat{\alpha}) - L(\Delta P) + \gamma G(\overline{P} + \Delta P, \hat{\alpha}) \\ Q(\overline{P}, \hat{\alpha}) + \gamma G(\overline{P}, \hat{\alpha}) \end{cases} \quad \text{(A2)}$$

where $G(P, \hat{\alpha}) \equiv Q(P, \hat{\alpha})/(1-\gamma)$

[Note that (A2) is much simpler than (A1) in the sense that there is no more recursive relation between $f(\overline{P}, \hat{\alpha})$ and the two expressions in (A2).]

(2) We then solve (A2) in step (II).

In (A2), the complications remaining are:

(i) the possibility of regime switching between the two expressions; and

(ii) the need to find the optimal ΔP for the first expression, before the comparison of the two expressions (or regimes).

Although the problem is still somewhat complicated, step (II) of this appendix shows that they can be solved.

Step (I): Equivalence of (A1) and (A2)

To prove this, we need to consider the following two cases:

Case (1): $Q(\overline{P}, \hat{\alpha}) + \gamma f(\overline{P}, \hat{\alpha}) \geq \sup_{\Delta P \neq 0} Q(\overline{P} + \Delta P, \hat{\alpha}) - L(\Delta P) + \gamma f(\overline{P} + \Delta P, \hat{\alpha})$

If this happens, $f(\overline{P}, \hat{\alpha})$ will take the value of $Q(\overline{P}, \hat{\alpha}) + \gamma f(\overline{P}, \hat{\alpha})$

$$\Rightarrow f(\overline{P}, \hat{\alpha}) = Q(\overline{P}, \hat{\alpha}) + \gamma f(\overline{P}, \hat{\alpha})$$

$$\Rightarrow \mathbf{f(\overline{P}, \hat{\alpha}) = Q(\overline{P}, \hat{\alpha})/(1-\gamma) \equiv G(\overline{P}, \hat{\alpha})}$$

Thus, $f(\overline{P}, \hat{\alpha})$ in the second expression of (A1) can be rewritten as $G(\overline{P}, \hat{\alpha})$.

<u>Case (2): $Q(\overline{P}, \hat{\alpha}) + \gamma f(\overline{P}, \hat{\alpha}) < \sup_{\Delta P \neq 0} Q(\overline{P} + \Delta P, \hat{\alpha}) - L(\Delta P) + \gamma f(\overline{P} + \Delta P, \hat{\alpha})$</u>

Before starting the proof, we note that:

(i) For any fixed P_1, there exist ΔP_1 such that

$$Q(P_1 + \Delta P_1, \hat{\alpha}) - L(\Delta P_1) + \gamma f(P_1 + \Delta P_1, \hat{\alpha})$$

$$= \sup_{\Delta P \neq 0} Q(P_1 + \Delta P, \hat{\alpha}) - L(\Delta P) + \gamma f(P_1 + \Delta P, \hat{\alpha})$$

[We can justify this fact when the solution has been obtained.]

(ii) For the ΔP_1 satisfying (i) above, we have

$$Q(P_1 + \Delta P_1, \hat{\alpha}) + \gamma f(P_1 + \Delta P_1, \hat{\alpha})$$

$$\geq Q(P_1 + \Delta P_1 + \Delta P_2, \hat{\alpha}) - L(\Delta P_2) + \gamma f(P_1 + \Delta P_1 + \Delta P_2, \hat{\alpha}) \qquad \forall \Delta P_2$$

[ie if we start with $P_1 + \Delta P_1$ and $\hat{\alpha}$, which means the price is at its optimal level, the best policy is to keep the price unchanged.]

<u>Proof</u>:

If not, then there exists ΔP_2 such that

$$Q(P_1 + \Delta P_1 + \Delta P_2, \hat{\alpha}) - L(\Delta P_2) + \gamma f(P_1 + \Delta P_1 + \Delta P_2, \hat{\alpha})$$
$$> Q(P_1 + \Delta P_1, \hat{\alpha}) + \gamma f(P_1 + \Delta P_1, \hat{\alpha})$$

Thus, $Q(P_1 + \Delta P_1, \hat{\alpha}) - L(\Delta P_1) + \gamma f(P_1 + \Delta P_1, \hat{\alpha})$

$$< Q(P_1 + \Delta P_1 + \Delta P_2, \hat{\alpha}) - L(\Delta P_1) - L(\Delta P_2) + \gamma f(P_1 + \Delta P_1 + \Delta P_2, \hat{\alpha})$$
$$\leq Q(P_1 + \Delta P_1 + \Delta P_2, \hat{\alpha}) - L(\Delta P_1 + \Delta P_2) + \gamma f(P_1 + \Delta P_1 + \Delta P_2, \hat{\alpha})$$

[since $L(\Delta P_1 + \Delta P_2) \leq L(\Delta P_1) + L(\Delta P_2)$]

This contradicts the fact that

$$Q(P_1 + \Delta P_1, \hat{\alpha}) - L(\Delta P_1) + \gamma f(P_1 + \Delta P_1, \hat{\alpha})$$
$$= \sup_{\Delta P \neq 0} Q(P_1 + \Delta P, \hat{\alpha}) - L(\Delta P) + \gamma f(P_1 + \Delta P, \hat{\alpha})$$

With (i) and (ii) in mind, we now proceed to show that $f(\overline{P} + \Delta P, \hat{\alpha})$, the last term of the first expression of (A1), can be rewritten as $G(\overline{P} + \Delta P, \hat{\alpha})$:

Applying (A1) for $f(P_1 + \Delta P_1, \hat{\alpha})$,we know that it satisfies the following relationship:

$$f(P_1+\Delta P_1, \hat{\alpha}) = \max \begin{cases} \sup\limits_{\Delta P_2 \neq 0} Q(P_1+\Delta P_1+\Delta P_2, \hat{\alpha}) - L(\Delta P_2) + \gamma f(P_1+\Delta P_1+\Delta P_2, \hat{\alpha}) \\ Q(P_1 + \Delta P_1, \hat{\alpha}) + \gamma f(P_1 + \Delta P_1, \hat{\alpha}) \end{cases} \tag{A3}$$

From (ii) above, we note that the value of the second expression of (A3) is greater than or equal to that of the first expression. Therefore, $f(P_1+\Delta P_1, \hat{\alpha})$ will take the value of the second expression:

$$\Rightarrow f(P_1+\Delta P_1, \hat{\alpha}) = Q(P_1 + \Delta P_1, \hat{\alpha}) + \gamma f(P_1 + \Delta P_1, \hat{\alpha})$$

$$\Rightarrow \mathbf{f(P_1+\Delta P_1, \hat{\alpha}) = Q(P_1 + \Delta P_1, \hat{\alpha})/(1-\gamma) \equiv G(P_1+\Delta P_1, \hat{\alpha})}$$

Combining cases (1) and (2), it becomes obvious that (A1) can be rewritten as:

$$f(\overline{P}, \hat{\alpha}) = \max \begin{cases} \sup\limits_{\Delta P \neq 0} Q(\overline{P} + \Delta P, \hat{\alpha}) - L(\Delta P) + \gamma G(\overline{P} + \Delta P, \hat{\alpha}) \\ Q(\overline{P}, \hat{\alpha}) + \gamma G(\overline{P}, \hat{\alpha}) \end{cases} \tag{A2}$$

where $G(P, \hat{\alpha}) \equiv Q(P, \hat{\alpha})/(1-\gamma)$

Substituting the value of $G(P, \hat{\alpha})$ in, equation (A2) can be rewritten as:

$$f(\overline{P}, \hat{\alpha}) = \max \begin{cases} \sup\limits_{\Delta P \neq 0} \dfrac{Q(\overline{P} + \Delta P, \hat{\alpha})}{1-\gamma} - [A + \xi(\Delta P)] \\ \dfrac{Q(\overline{P}, \hat{\alpha})}{1-\gamma} \end{cases}$$

which is equation (2.8) of the main text.

Step (II): Solution of (A2)

Having proven the equivalence of (A1) and (A2), we now come to the solution of (A2). As the relationship in (A2) is less complicated (ie no recursive relationship between $f(\overline{P}, \hat{\alpha})$ and the two expressions), the two expressions can be analyzed separately. We start with graphical analysis of the two expressions:

(1) Graphical shape of the first expression

Lemma: $\operatorname{Sup}_{\Delta P \geq 0}\ Q(\overline{P} + \Delta P, \hat{\alpha}) - L(\Delta P) + \gamma G(\overline{P} + \Delta P, \hat{\alpha})$

can be represented by a straight line SS shown in diagram (A2)

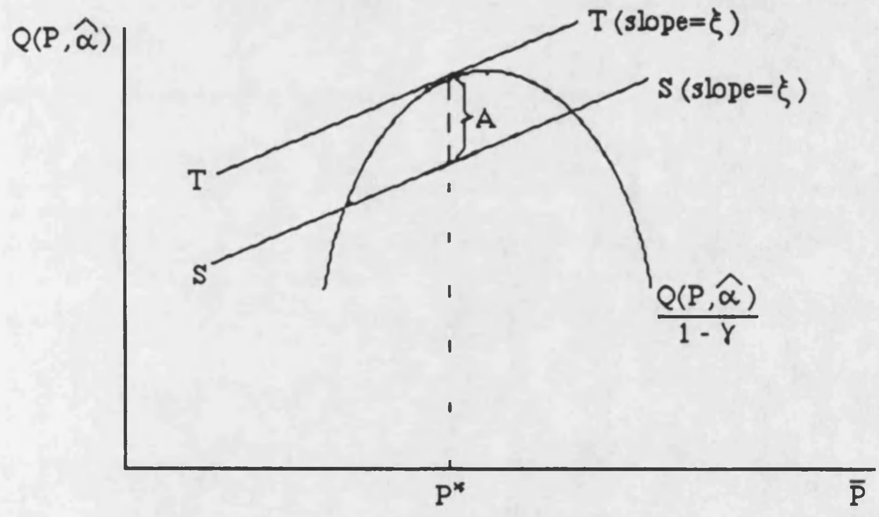


diagram (A2)

Hence, $\operatorname{Sup}_{\Delta P \geq 0}\ Q(\overline{P} + \Delta P, \hat{\alpha}) - L(\Delta P) + \gamma G(\overline{P} + \Delta P, \hat{\alpha})$

can be easily found out with the following steps:

(i) Plot $Q(\overline{P}, \hat{\alpha})/(1-\gamma)$ against $\overline{P}$;

(ii) Find a point on $Q(\overline{P}, \hat{\alpha})/(1-\gamma)$ so the slope of its tangent (TT) is ξ;

(iii) Draw a line SS which is vertically below TT by a distance A.

Proof:

(a) We note that

$$Q(\overline{P} + \Delta P, \hat{\alpha}) - L(\Delta P) + \gamma G(\overline{P} + \Delta P, \hat{\alpha})$$

$$= Q(\overline{P} + \Delta P, \hat{\alpha}) - L(\Delta P) + Q(\overline{P} + \Delta P, \hat{\alpha})/(1-\gamma)$$

[since $G(\overline{P}+\Delta P, \hat{\alpha}) = Q(\overline{P}+\Delta P, \hat{\alpha})/(1-\gamma)$]

$$= Q(\overline{P} + \Delta P, \hat{\alpha})/(1-\gamma) - L(\Delta P)$$

$$= Q(\overline{P} + \Delta P, \hat{\alpha})/(1-\gamma) - A - \xi(\Delta P)$$

$$= Q(P, \hat{\alpha})/(1-\gamma) - A - \xi(P - \overline{P}) \qquad \text{where } P = \overline{P} + \Delta P$$

(b) To find the value of $\sup_{\Delta P} Q(\overline{P} + \Delta P, \hat{\alpha}) - L(\Delta P) + \gamma G(\overline{P} + \Delta P, \hat{\alpha})$,

we differentiate the above expression with respect to P, noting that $\overline{P}$ is the starting value which can be considered as fixed during the differentiation. Setting the derivative to zero, we find that $Q(\overline{P}+\Delta P, \hat{\alpha}) - L(\Delta P) + \gamma G(\overline{P}+\Delta P, \hat{\alpha})$ has a maximum value at point P^*

$$\text{where} \qquad \left.\frac{\partial[Q(P, \hat{\alpha})/(1-\gamma)]}{\partial P}\right|_{P=P^*} = \xi \qquad \text{(A4)}$$

(c) Substituting the expression of $Q(P, \hat{\alpha})$ into (A4), we have:

$$\frac{(\hat{\alpha} - \beta P^*)(P^* - b)(-\beta)}{1-\gamma} = \xi$$

$$\Rightarrow P^* = \frac{\hat{\alpha} + b\beta - \xi(1-\gamma)}{2\beta}$$ which is equation (2.5) in the main text.

Note that P^* is independent of the starting value $\overline{P}$.

(d) We now proceed to our diagrammatic discussion by plotting $Q(\overline{P}, \hat{\alpha})/(1-\gamma)$ against $\overline{P}$ and assuming $\overline{P}_1$ as the starting value:

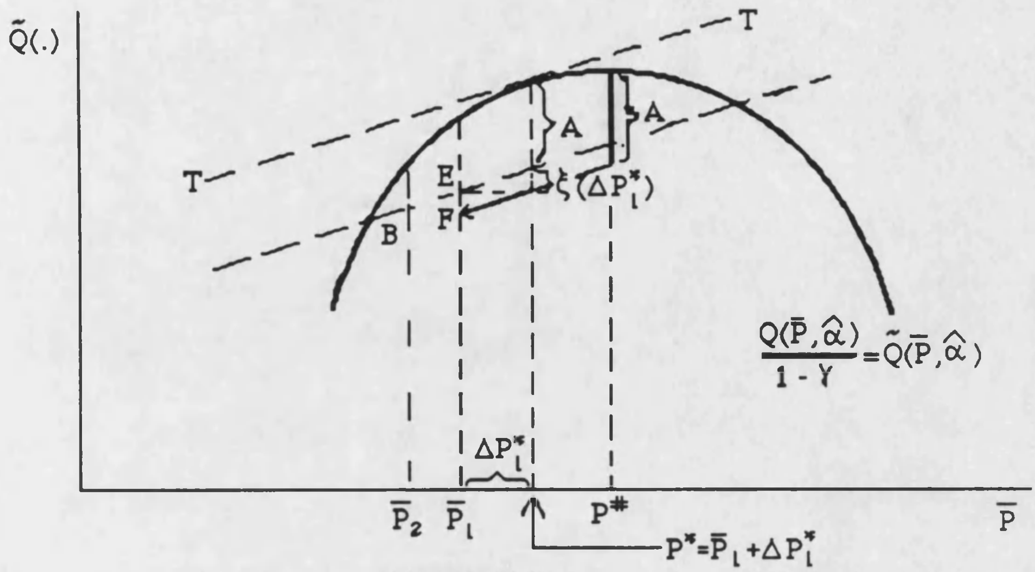


diagram (A3)

From (b), we note that $Q(\overline{P} + \Delta P, \hat{\alpha}) - L(\Delta P) + \gamma G(\overline{P} + \Delta P, \hat{\alpha})$ has a maximum value at P^* at which the slope of $Q(\overline{P}, \hat{\alpha})/(1-\gamma)$ is ξ. That is, for any starting value $\overline{P}_1$, the optimal change of price is ΔP_1^*. After the change in price, $Q(P^*, \hat{\alpha})$ is represented by the height of point D. With the deduction of the two costs of price changes from $Q(P^*, \hat{\alpha})$, $Q(P^*, \hat{\alpha}) - A - \xi(\Delta P_1^*)$ or

$\sup_{\Delta P \neq 0} Q(\overline{P}_1 + \Delta P, \hat{\alpha}) - L(\Delta P) + \gamma G(\overline{P}_1 + \Delta P, \hat{\alpha})$ is now represented by the height of point E.

For illustrative purpose, we also consider the case when the price is raised from $\overline{P}_1$ to $P^\#$. The value of $Q(P^\#, \hat{\alpha})/(1-\gamma) - A - \xi(P^\# - \overline{P}_1)$ is represented by the height of point F which is smaller than that of point E. Comparing with all other points, we can see the value of $Q(\overline{P}_1 + \Delta P, \hat{\alpha}) - L(\Delta P) + \gamma G(\overline{P}_1 + \Delta P, \hat{\alpha})$ takes a maximum when price is raised to P^* (not $P^\#$ or other points).

Now, for any other starting value $\overline{P}_2$,

$\sup_{\Delta P \neq 0} Q(\overline{P}_2 + \Delta P, \hat{\alpha}) - L(\Delta P) + \gamma G(\overline{P}_2 + \Delta P, \hat{\alpha})$ is represented by the height of point B. Hence, $\sup_{\Delta P} Q(\bar{P}+\Delta P, \hat{\alpha}) - L(\Delta P) + \gamma G(\bar{P}+\Delta P, \hat{\alpha})\ \forall\ \bar{P}$ is represented by the straight line passing through the points B and E (ie a straight line parallel to TT by a vertical distance of A).

(2) Graphical shape of the second expression

As $Q(\bar{P}, \hat{\alpha}) + \gamma G(\bar{P}, \hat{\alpha})$

$$= Q(\bar{P}, \hat{\alpha}) + \gamma [Q(\bar{P}, \hat{\alpha})/(1-\gamma)]$$

$$= Q(\bar{P}, \hat{\alpha})/(1-\gamma) \quad ;$$

the graphical shape of the second expression is parabolic:

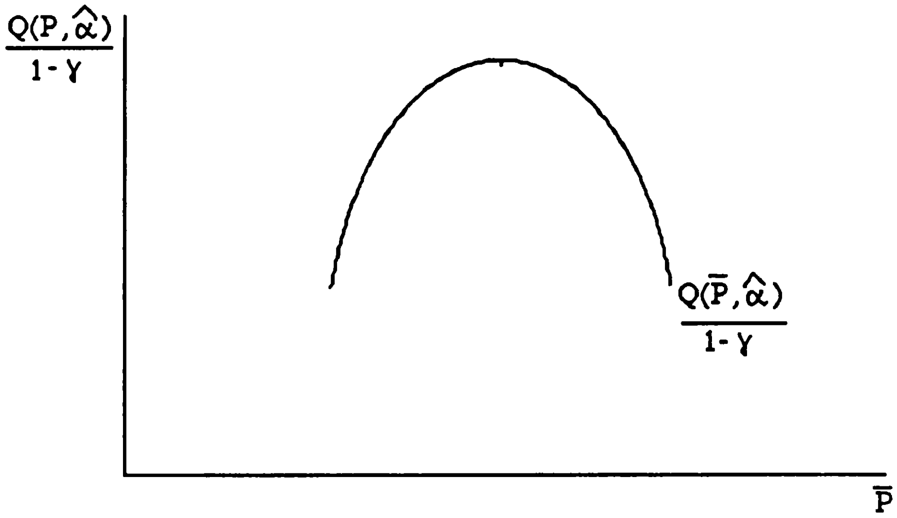


diagram (A4)

(3) Graphical shape of f (P̄, α̂)

We now put the graphs of the two expressions into the same diagrams. As shown in the two subsections above, the first expression is represented by the straight line SS and the second expression represented by the curve Q ($\bar{P}$, $\hat{\alpha}$)/(1-γ). From (A2), we know that f ($\bar{P}$, $\hat{\alpha}$) will take the maximum of the two. Hence, f ($\bar{P}$, $\hat{\alpha}$) will be represented by the thick dotted line curve in diagram (A5):

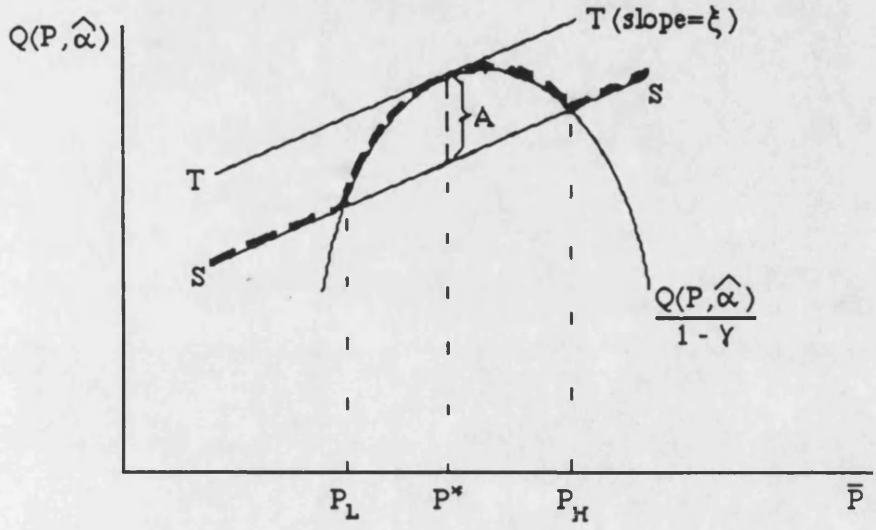


diagram (A5)

Mathematical Appendix (II): Solution of for P_L,P_H and P^*

We now come to solve for the value of P_L,P_H and P^* in diagram (2.1):

(i) Solution for P^*

As we have show in step (II)(1)(c) of Mathematical Appendix (I),

$$P^* = \frac{\hat{\alpha} + b\beta - \xi(1-\gamma)}{2\beta} \tag{A5}$$

Hence, the "optimal" change in price between any two periods of different $\hat{\alpha}$ is

$$\Delta P^* = \frac{\Delta \hat{\alpha}}{2\beta}$$

(ii) Solution for P_L and P_H

From equation (2.8), we have

$$f(\bar{P}, \hat{\alpha}) = \max \begin{cases} \sup\limits_{\Delta P \neq 0} \dfrac{Q(\bar{P} + \Delta P, \hat{\alpha})}{1-\gamma} - [A + \xi(\Delta P)] \\ \dfrac{Q(\bar{P}, \hat{\alpha})}{1-\gamma} \end{cases}$$

where $Q(P, \hat{\alpha}) = (P-b)[\hat{\alpha} - \beta P] - a$

Differentiating the first expression on the right hand side of the above equation with respect to ΔP and setting the derivative to zero, we know that the optimal ΔP^* satisfy

$$\frac{(\bar{P} + \Delta P^* - b)(-\beta) + [\hat{\alpha} - \beta(\bar{P} + \Delta P^*)]}{1 - \gamma} - \xi = 0 \tag{A6}$$

To solve for P_L and P_H (intersections of the curvature section and the straight line section SS in diagram (2.1)) is to look for the $\bar{P}$ at which the first and second expressions on the right hand side of (2.3) are equal to each other. Thus,

$$\frac{(\bar{P}+\Delta P^{*}-b)\,[\hat{\alpha}-\beta(\bar{P}+\Delta P^{*})]-a}{1-\gamma}-[A+\xi\,(\Delta P^{*})]=\frac{(\bar{P}-b)\,[\hat{\alpha}-\beta\,\bar{P}]-a}{1-\gamma}$$

$$\Rightarrow\ \frac{(\bar{P}+\Delta P^{*}-b)\,(-\beta\,\Delta P^{*})+\Delta P^{*}\,[\hat{\alpha}-\beta\,(\bar{P}+\Delta P^{*})]}{1-\gamma}-[A+\xi\,(\Delta P^{*})]=0 \qquad \text{(A7)}$$

Substituting (A6) into (A7) and rearranging, we have

$$\Delta P^{*}=\pm\sqrt{[A(1-\gamma)/2\beta]}$$

which implies

$$P_L = P^{*} - \sqrt{[A(1-\gamma)/2\beta]} \qquad ;\text{ and}$$

$$P_H = P^{*} + \sqrt{[A(1-\gamma)/2\beta]}$$

Thus the gap between P_L and P_H is $2\sqrt{[A(1-\gamma)/2\beta]}$.

Chapter 3

3.1 Introduction

In this Chapter, we will try to build a model on the employment decision of firms with respect to demand shocks. Before presenting the model, we will first discuss the idea of insured employment suggested by Akerlof and Miyazaki (1980). The idea was originated in the literature on implicit contracts and real wage rigidity which suggest that it pays both parties if the less risk averse employers "insure" their risk averse employees with small variations in wage rate over the various possible states of nature, and in return the employers are compensated by risk premia in the form of lower average wages which workers are implicitly willing to pay for such wage insurance. Akerlof and Miyazaki suggest that the difference in risk aversion will also imply implicit insurance of employment provided by employers :

Outline of their argument

By assuming that the worker's utility function (U[.]) is strictly concave with respect to wage, the paper argues that workers will prefer a lower expected wage with a guarantee of employment. That is,

$$U[w_j(s)n_j(s)/\lambda_j] > U[w_j'(s)] > \{n_j(s)/\lambda_j\}\ U[w_j(s)]$$

where (a) $U[w_j(s)n_j(s)/\lambda_j]$ is the worker's utility due to a guarantee of employment with wage $w_j(s)\ \{n_j(s)/\lambda_j\}$;

(b) $U[w_j'(s)]$ is the worker's utility due to a guarantee of employment with wage $w_j'(s)$, where $w_j'(s) < w_j(s)\{n_j(s)/\lambda_j\}$; and

(c) $\{n_j(s)/\lambda_j\}U[w_j(s)]$ is the worker's utility due to a non-guaranteed employment where there is a probability $\{n_j(s)/\lambda_j\}$ that his wage is $w_j(s)$ and a probability $(1 - \{n_j(s)/\lambda_j\})$ that his wage is zero.

If the employer is risk neutral, he will be indifferent between scheme (a) and scheme (c) because both schemes imply the same expected wage bill $n_j(s)w_j(s)$. Hence, compared with the case of no guarantee of employment in scheme (c), the employer can offer a lower wage $w_j'(s)$ with guarantee of employment (i.e. scheme (b)) so that expected wage bill is reduced from $n_j(s)w_j(s)$ to $n_j(s)w_j'(s)$ and yet workers are happier.

Once scheme (b) is concluded, there is little incentive for the employer to violate the guarantee of employment because this will ruin the reputation of the firm's personnel policy which would imply a higher wage to attract (or retain) sufficient workers in the future.

The contribution of the above analysis is that it provides a good explanation of the wide acceptance of some implicit guarantee of employment in the economy. However, there are quite a few difficulties with the above analysis. First, it does not allow for the possibility that

(i) employers might find it advantageous to break the promise in case of very adverse demand shocks; or indeed

(ii) employers <u>will not</u>, at the very beginning, promise to provide a full (or perfect) guarantee of employment covering any size of adverse demand shocks.

That is, for some large and persistent enough reduction of demand, the cost of layoff may be lower than the cost of maintaining the excessive wage bill. What will the employment decision be in such case? Will the employment response be different for different size of demand shocks? Second, for the case of mild demand shocks, the implicit guarantee of employment would imply stickiness of employment with respect to the demand shocks. What variable(s) will then be adjusted to ensure that output will be produced to satisfy the variations in demand? These are the questions that will be dealt with in this Chapter.

3.2 The Model

3.2.1 Assumptions

(A) In a fully set up model, one should have the producer – based on the present expectation – facing the simultaneous decisions on capacity, price, output, employment, wage, inventory and etc. In particular, because of the fixed cost per change in some of these decisions, a producer who experiences a demand shock will have to simultaneously consider among the following bang-bang decisions of

(1) changing the capacity or not;

(2) changing price or not;

(3) changing the employment or not;

(4) changing the wage promise or not;

(5) having unsatisfied demand or not.

However, this will - according to our usual formulation - imply a huge dimension of complexity. For example, even if we consider the simultaneous decisions of (2) and (3) only, we will have

Max of $f(\bar{P}, \bar{N}, \hat{\alpha})$=expected profit by

- (a) keeping the price at $\bar{P}$ and keeping employment at $\bar{N}$
- (b) changing price to $\bar{P}+\Delta P^{*}$ and keeping employment at $\bar{N}$
- (c) keeping the price at $\bar{P}$ and changing employment to $\bar{N}+\Delta N^{*}$
- (d) changing price to $\bar{P}+\Delta P^{**}$ and changing employment to $\bar{N}+\Delta N^{**}$

where ΔP^{*}, ΔN^{*}, ΔP^{**} and ΔN^{**} are respectively the optimal ΔP and ΔN with respect to their particular solution.

As will be shown later, unlike the case of pricing decision, it is necessary to sub-divide the expression of changing employment into that of raising employment (hiring) and that of reducing employment (layoff) so that there are indeed six expressions in the above formulation. Moreover, the graph of $f(\bar{P}, \bar{N}, \hat{\alpha})$ will also be three dimensional.

For the sake of tractability, we will try to make a few assumptions so that a simplified model can be formulated to highlight the response of producer's employment decision to demand shocks. We do so by assuming [1]:

(i) the planned excess capacity is sufficiently large so that the demand shocks we consider are well within this planned excess capacity;

(ii) there is no inventory;

(iii) the cost of having unsatisfied demand is relatively high when compared with the cost of raising production within the planned excess capacity so that the producer will choose not to have any unsatisfied demand; and

(iv) the nominal wage is bargained and revised every year, and the cost of paying below the agreed nominal wage is high enough to stop the producer from cutting wages in the face of a reduction in demand. We also assume the producer – in the case of positive demand shock – will choose to promise a higher bonus at the end of the year rather than raising the wage rate immediately.

With assumptions (i) and (iv), now we can consider the capacity and nominal wage at the beginning of most periods as predetermined for the employment decision. Assumptions (ii) and (iii) on the other hand will avoid the complexity arising from further choices on the levels of inventory and unsatisfied demand. The two assumptions also imply that the output decision is determined once the price decision is made. Thus, the only inter-dependent decisions left will be those on price and employment. Yet, as explained above, the problem will still be highly complicated. To avoid the complexity, we make the following assumptions so that price will also be predetermined for the employment decision :

(a) The case of a negative demand shock

Following the argument in Okun(1981), we assume the demand curve will have a kink at the present price P_0 and the price elasticity of demand at any price

[1] The results we derive in the subsequent sections may still hold even with the relaxation of some of these assumptions. A proper formulation of these will however be highly complicated.

below P_0 will be far smaller than unity. Such a demand curve is drawn as D_A or D_A' in diagram (3.1). The rationale for this assumption is as follows :

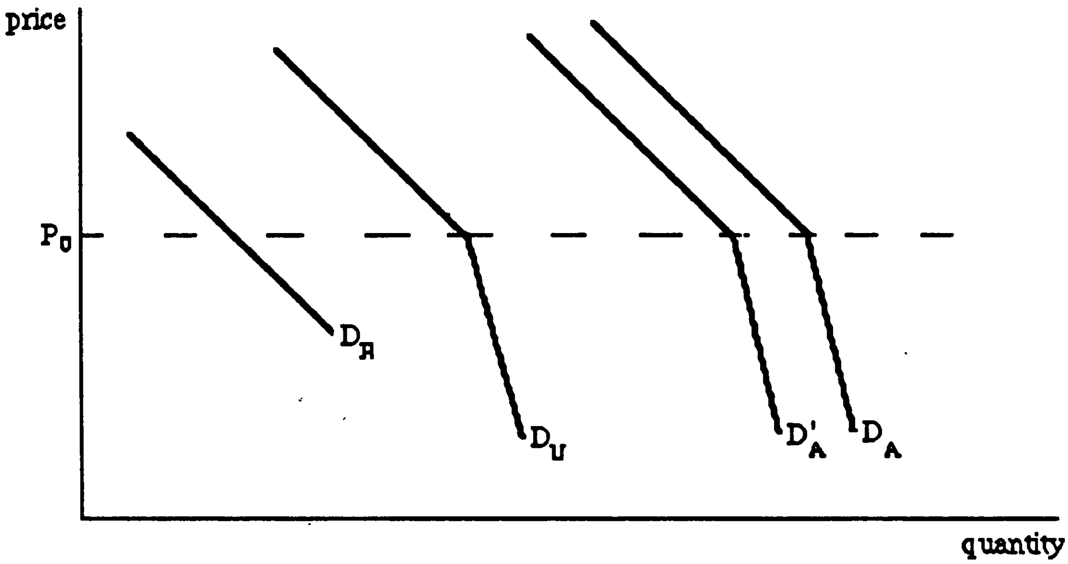


diagram (3.1)

Because of the cost of shopping, most of the customers of the firm who are satisfied with the previous purchase will come back to shop. If they find the price unchanged, they will become the buyers. If they find the price is raised, they may or may not decide to shop further. Thus, the demand from the "usual customers" will have a kink at P_0. Below P_0, the price elasticity will be low because presumably the repeat callers were ready to buy for the same price that they paid the last time. But the elasticity may be substantial at price above P_0 because the repeaters are responsive to price levels that exceed what they experienced previously. Such demand from the "usual customers" is drawn as D_U in diagram (3.1) in which the part of demand below P_0 is assumed to be highly inelastic [2]. In addition to the usual

[2]In the extreme case that the repeat customers only demand fixed amount of quantity (eg. a room in the hotel, a ticket for the film etc), the lower part of D_U will be vertical.

customers, we can also allow for some percentage of random shoppers who are new to the firm. As a group, the demand from the "random shoppers" is likely to be elastic and continuous. We draw this as D_R in diagram (3.1). Finally, the overall demand curve D_A can be obtained by the horizontal summation of D_U and D_R. Thus, we can assume the elasticity of demand from random shoppers is not too high and the percentage of random shoppers in the total demand is small enough so that the elasticity of overall demand at any price below P_0 is far smaller than unity.

As is well known in the standard theory of firm, the short run profit maximization point will never lie on the region of demand where elasticity is less than unity. Thus, if the producer were to start with demand D_A and price P_0, a negative demand shock shifting the demand to D_A' will imply the best short run price for the producer is still P_0.

However, the above reasoning – as noted by Okun(1981) – takes a myopic view of firms's profit maximization. It neglects the interdependence introduced by the customer relation between today's purchase and tomorrow's level of demand. That is, the value of obtaining the additional random shoppers as customers includes not merely the proceeds from their current purchases but also an additional benefit associated with the likelihood that they will return to buy in the future. With such potential gain in the long run, the producer might in the short run choose to reduce price in face of a reduction in demand. To avoid the complexity arise from such a possibility, we will – based on the explanations above – assume the elasticities of the overall demands (D_A and D_A') below P_0 are far smaller than unity so that the expected gain of the repeated purchase from random shoppers in the future is negligible when compared with the cost of charging a lower price at a very inelastic region. Such assumption guarantees that the producer will not reduce the price below

P_0 in face of a negative demand shock.

(b) The case of positive demand shock

We will assume the cost of hiring (such as advertisement and training cost) will be relatively small when compared with the cost of raising price [3,4]. This is not an unrealistic assumption because the absence of any reputation cost in hiring implies that the cost of hiring is at most related to the number of new employees, while most of the cost of raising price lies on the reputation loss to all customers. Thus, for those positive demand shocks that are not sufficient to cause a rise in price, we can assume price is predetermined for the employment decision [5].

We think that the above discussion is a more reasonable justification for assuming that price is predetermined for the employment decision. For example, if we drop the assumption in case (a), we might have to rely on the less realistic assumption of a lower cost of layoff than that of reducing price.

With this set of slightly restrictive assumptions, we can then rigorously develop the subsequent model and highlight the response of the employment decision to the range of demand shocks we have restricted to. It should however be emphasized that the response may be qualitatively the same for a larger demand shock

---

[3] Unlike the case of layoff, there is no reputation loss in hiring. Indeed, this is why we assume the cost of layoff is higher than the cost of reducing price on one hand, and assume the cost of hiring is lower than the cost of raising price on the other.

[4] The result may still be the same even if the cost of hiring turns out to be higher than the cost of raising price so that price is raised (instead of predetermined) for some large enough demand shock. In such case, the employment and price decision will be interdependent so that they have to be solved simultaneously.

[5] We also assume that the starting price is not too close to the margin of raising price. If this is not the case or the demand shock is too large, price will be raised and there will be interdependence between the price and employment decisions. We will return to the discussion of such possibility in section 3.5.

along which there might also be a rise in price as well [6]. We prefer the slightly restrictive assumptions only because we want to have a rigorous and simple model on employment decision.

(B) We assume that the cost of layoff, which include

(a) the actual payment to the worker being laid off; and

(b) the reputation cost (which imply a higher potential quit of existing workers and less willing applicant in the future [7]) due to the act of layoff,

can be approximated by $[B - \theta(\Delta N)]$[8], where $\Delta N<0$ is the amount of workers being laid off.

Similarly, the cost of hiring, which include

(a) the advertisement and interview cost;

(b) the training cost; and

(c) the production loss due to the distortion arising from on-the-job-training,

can be approximated by $[C + \psi(\Delta N)]$, where $\Delta N>0$ is the amount of new recruitment.

(C) As with efficiency wage models, we assume that output (Q) is a function of the efficiency unit of labour hours. That is, $Q=k(ehN)$, where (i) e is the effort; (ii) h is the working hours; and (iii) N is the amount of workers in the firm. For simplicity, we also assume that $k'(.)>0$; $k''(.)<0$; $k(0)=0$; and $k(ehN)\rightarrow\infty$ as $ehN\rightarrow\infty$.

---

[6]In such case, the employer may have to raise the wage to reduce the potential quits and attract more willing applicants when the economy recovers. Since there exists some high enough wage that can stop the potential quits and attract sufficient new applicants, the reputation cost would be finite.

[7]For the case of restaurants and hair salons, unsatisfied demand can be a possibility, we will assume that this is not very significant for the whole economy.

[8]The result will only be slightly more complicated in case the cost of layoff (and hiring) is quadratic rather than linear.

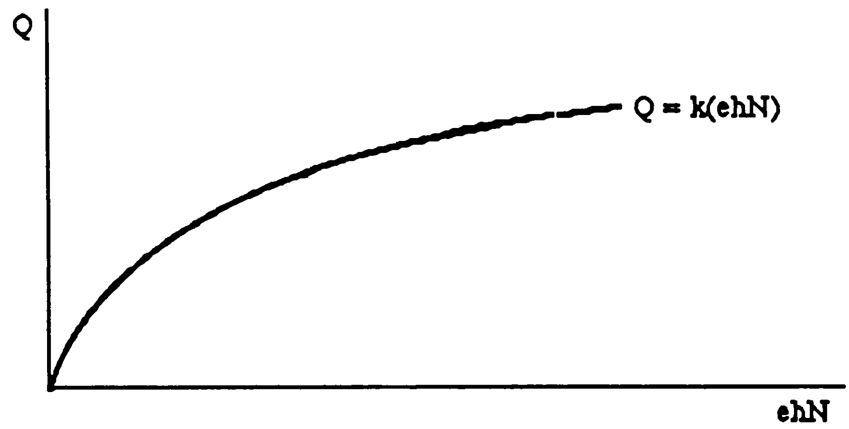


diagram (3.2)

For most jobs in the economy (which include those office/shop-keeping jobs in the banks, government offices, departmental stores, restaurants, hair salon, wholesaling offices, retail shops and etc), the number of working hours is fixed by usual practice. If the number of workers N is fixed as well, production effort will be the only variable of adjustment (i.e. if customers come in, the staff will serve them; if no customer comes in, the staff will just sit there)[9]. For the case of accounting, executive and administrative staffs, they might work overtime or bring work home during the high season. Nevertheless, as official hours are fixed and they only receive additional payment through the bonus at the end of year, we can still include this type of overtime or bring home work as effectively higher production effort instead of longer working hours h. However, for the others such as those in the manufacturing industry, the employers will ask their workers to increase production effort and possibly work overtime. In such case, both e and h will be raised. For simplicity, we will mainly present the case where h is institutionally fixed at $\bar{h}$. Nevertheless, the results we derive in the later sections will be basically the same as long as e is also raised with the rise of demand.

[9] In the case of hair salons where there is a temporary rise in demand, production effort is raised (or quality reduced) in the sense that the hair-dresser will speed up the hair cuts.

(D) We also assume, within a certain range of output-labour hour ratio (Q/hN), e can be adjusted up or down (without any additional cost of changing e) so that the required output can always be produced even if employment is fixed at $\bar{N}$. There is however a cost for higher effort per worker and the cost, g(e), takes the following shape (monotonically increasing and convex around a reasonable range of normal effort) :

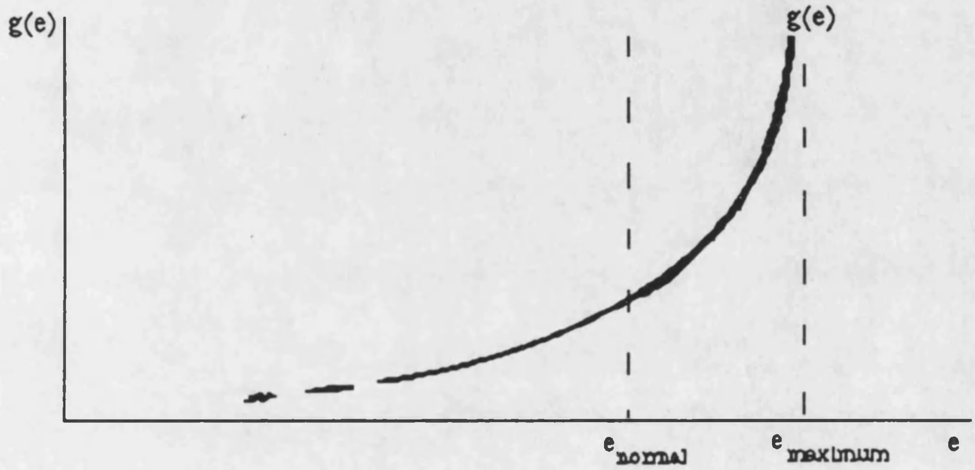


diagram (3.3)

Thus, we can conceptually distinguish two types of cost per worker :

(a) the normal wage wh; and

(b) the additional cost g(e) such as higher bonus at the end of year; the greater consumption of materials in case of tighter production schedule; or higher administrative cost in keeping labour at a higher level of production effort.

(E) Unlike the case of layoff or hiring, there is no additional (reputation or administrative) cost in changing effort from one level to another.

(F) For simplicity, we assume that the total cost other than whN (the wage bill) and

$g(e)\bar{h}N$ to be $a+b\hat{Q}$, where $\hat{Q}$ is the expected output at price $\hat{P}$ (eg. $\hat{Q}=\hat{\alpha}-\beta\hat{P}$).

Thus, by assumption (A), $\hat{Q}$ is also predetermined for the employment decision.

Moreover, the revenue minus such kind of cost will be $(\hat{P}-b)\hat{Q}-a$.

3.2.2 <u>Formulation</u>

(A) Similar to the case of pricing decision, the employer's problem – given $\hat{P}$ and $\hat{Q}$ – is to choose the optimal level of e and N to achieve the maximum expected profit $f(\bar{N},\hat{Q})$ where

$$f(\bar{N},\hat{Q})=\max\begin{cases}\sup\limits_{\Delta N>0}(\hat{P}-b)\hat{Q}-a-\bar{w}\bar{h}(\bar{N}+\Delta N)-g(e)\bar{h}(\bar{N}+\Delta N)-[C+\psi(\Delta N)]+\gamma f(\bar{N}+\Delta N,\hat{Q})\\(\hat{P}-b)\hat{Q}-a-\bar{w}\bar{h}\bar{N}-g(e)\bar{h}\bar{N}+\gamma f(\bar{N},\hat{Q})\\\sup\limits_{\Delta N<0}(\hat{P}-b)\hat{Q}-a-\bar{w}\bar{h}(\bar{N}+\Delta N)-g(e)\bar{h}(\bar{N}+\Delta N)-[B-\theta(\Delta N)]+\gamma f(\bar{N}+\Delta N,\hat{Q})\end{cases}\quad(3.1)$$

where $\bar{N}$ is the initial employment at the beginning of that period.

Unlike the case of price decision, layoff ($\Delta N<0$) is now associated with a linear reputation cost instead of gain. Hence, it is necessary to separate the cases of $\Delta N>0$ and $\Delta N<0$ as shown above.

(B) Given that (i) there is no extra cost in changing e from one level to the other; (ii) g(e) is convex; and (iii) labour is homogeneous, <u>it is always the best to share the output evenly among workers and keep the average e – given the level of output Q, employment N and working hours $\bar{h}$ – to the minimum.</u> From the production function $Q=k(e\,\bar{h}N)$, this implies $e = k^{-1}(\hat{Q}) / (\bar{h}N)$

Substituting this into equation (3.1), the employer's problem becomes choosing the level of N to achieve the maximum expected profit $f(\bar{N},\hat{Q})$:

$$f(\bar{N},\hat{Q})=\max\begin{cases}\sup\limits_{\Delta N>0} H(\bar{N}+\Delta N,\hat{Q})-[C+\psi(\Delta N)]+\gamma f(\bar{N}+\Delta N,\hat{Q})\\ H(\bar{N},\hat{Q})+\gamma f(\bar{N},\hat{Q})\\ \sup\limits_{\Delta N<0} H(\bar{N}+\Delta N,\hat{Q})-[B-\theta(\Delta N)]+\gamma f(\bar{N}+\Delta N,\hat{Q})\end{cases} \tag{3.2}$$

where

$$H[N,\hat{Q}] = (\hat{P}-b)\hat{Q} - a - \bar{w}\bar{h}N - g[\frac{k^{-1}(\hat{Q})}{\bar{h}N}]\bar{h}N \tag{3.3}$$

3.3 <u>Outline of the Solution</u>

(1) Similar to the case of pricing decision, solution of (3.2) and (3.3) can be shown to be equivalent to that of

$$f(\bar{N},\hat{Q}) = \max\begin{cases}\sup\limits_{\Delta N>0} \dfrac{H(\bar{N}+\Delta N,\hat{Q})}{1-\gamma} - [C+\psi(\Delta N)]\\ \dfrac{H(\bar{N},\hat{Q})}{1-\gamma}\\ \sup\limits_{\Delta N<0} \dfrac{H(\bar{N}+\Delta N,\hat{Q})}{1-\gamma} - [B+\theta(\Delta N)]\end{cases} \tag{3.4}$$

(2) Our next step is to <u>find the shape of $H(N,\hat{Q})$</u>. Given the shape of g(e) in diagram (3.2), it can be easily shown that the shape of $g[k^{-1}(\hat{Q})/(\bar{h}N)]$ and $g[k^{-1}(\hat{Q})/(\bar{h}N)]\bar{h}N$

will be as follows :

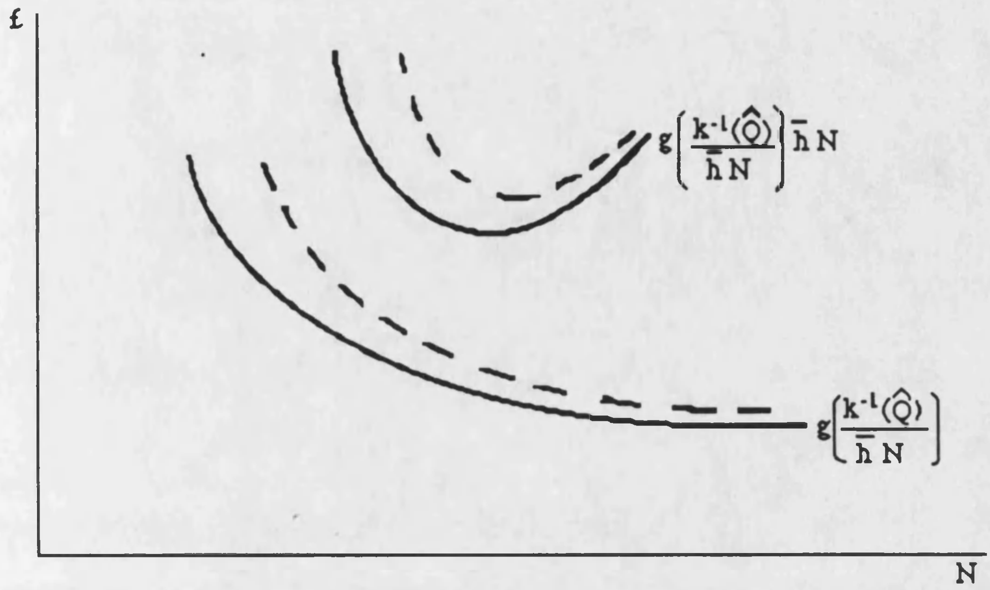


diagram (3,4)

With a rise in $\hat{Q}$· $g[k^{-1}(\hat{Q})/(\bar{h}N)]$ and $g[k^{-1}(\hat{Q})/(\bar{h}N)]\bar{h}N$ will shift to the dotted line.

Hence, $w\bar{h}N + g[k^{-1}(\hat{Q})/(\bar{h}N)]\bar{h}N$ and $H[N,\hat{Q}]$ can be drawn as follows :

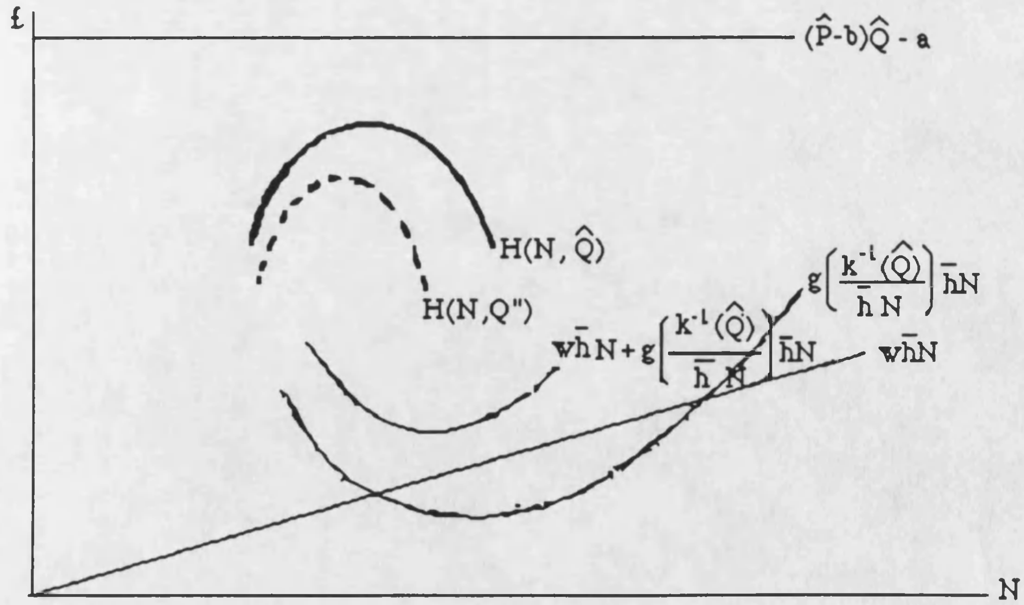


diagram (3,5)

With a reduction of $\hat{Q}$, $(\hat{P}\text{-b})\hat{Q}\text{-a}$ will shift downward and $w\bar{h}N + g[k^{-1}(\hat{Q})/(\bar{h}N)]\bar{h}N$ to the south east direction (not shown) so that H[N,Q] will shift to H[N,Q''], where $Q''<\hat{Q}$.

(3) After finding the shape of $H(N,\hat{Q})$, the value of $f(\bar{N},\hat{Q})$ in equation (3.4) can be found as follows:

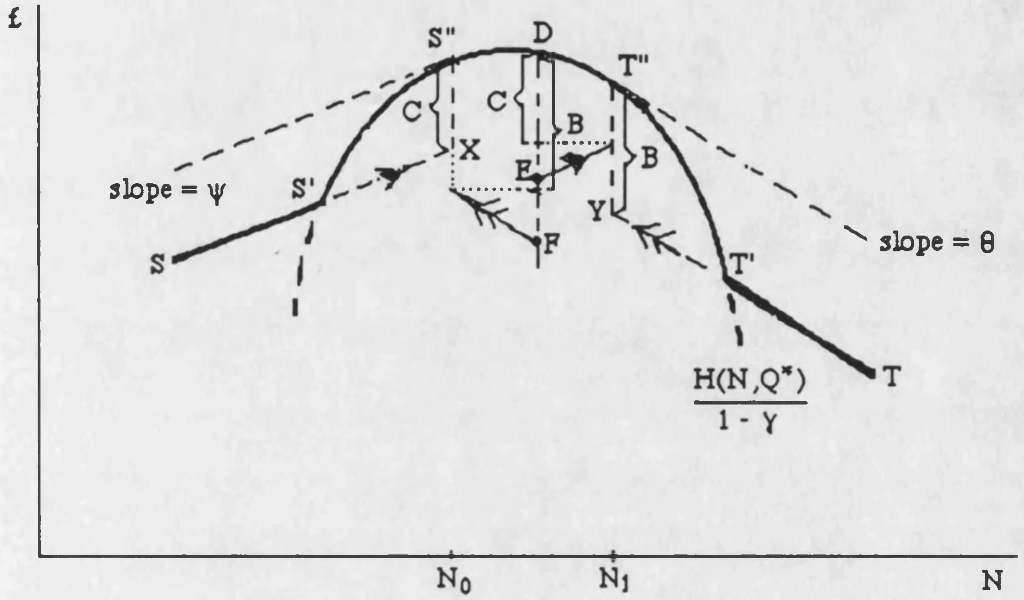


diagram (3.6)

(a) Knowing that the linear cost of hiring and layoff are $\psi(\Delta N)$ and $\theta(\Delta N)$ respectively, we locate S'' and T'' on $H[N, \hat{Q}]/[1-\gamma]$ so that the first and third expressions of equation (3.4) are respectively shown by the straight lines SS'X and YT'T. Hence, $f(\bar{N},\hat{Q})$ will take the shape of SS'S'' and T''T'T when the initial $\bar{N}$ is below N_0 or above N_1 (c.f. the mathematical appendix about the pricing decision).

(b) The remaining part of the analysis is for the case in which $\bar{N}$ lies between N_0 and N_1. Consider any point D on the curved section S''T''. If the employer keeps the employment at $\bar{N}$, his expected profit will be shown by the height of point D. On the other hand, if he reduces (raises) the employment to N_0 (N_1), his expected profit will only reach the height of point F(E). Clearly, the optimal policy for $N_1>\bar{N}>N_0$ is to keep the employment unchanged.

Hence, $f(\bar{N},Q^*)$ will be represented by the solid curve SS'S"DT"T'T which will shift to the south west (north east) with a reduction (rise) of $\hat{Q}$.

If (i) the initial $\bar{N}$ is below that of S', the employer should raise employment to N_0;

(ii) the initial $\bar{N}$ is between that of S' and T', the employer should keep the employment unchanged and only adjust the effort with respect to changes in $\hat{Q}$;

(iii) the initial $\bar{N}$ is above that of T', the employer should reduce employment to N_1.

It is interesting to note that, even if the fixed costs per action of hiring and layoff (B and C) are zero, there is still a range of N in which the employer would prefer to keep the employment unchanged. This is so because, unlike the linear cost of changing price, the "linear cost" of changing employment is indeed non-linear in the sense that there is a kink at $\Delta N=0$. In other words, a change in N in any direction is costly even if B=C=0.

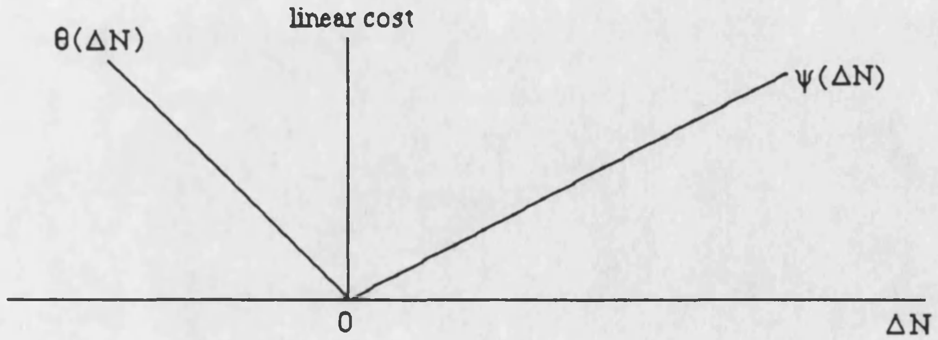


diagram (3.7)

3.4 Employment Response with respect to a Permanent Demand Shock

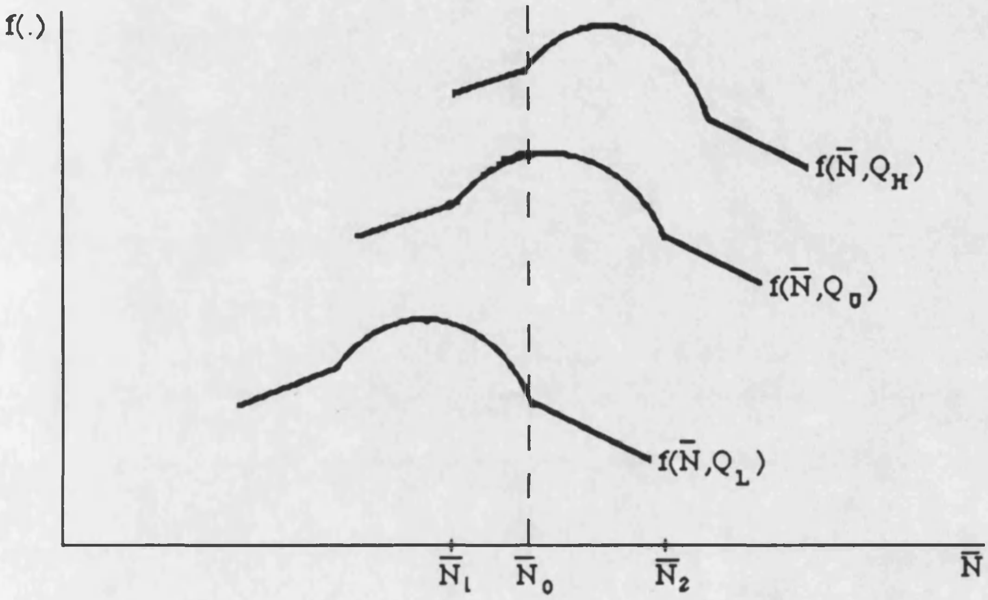


diagram (3.8)

(A) Mild Demand Shock

Suppose the employer starts with a value of $\overline{N}_0$ and $f(\overline{N},Q_0)$ as shown. If there is a mild increase in demand shifting $f(\overline{N},Q)$ slightly to the north-east of $f(\overline{N},Q_0)$, the employer will try to maintain the same employment and request a tighter production (or working) schedule [i.e. e is raised [10]]. Similarly, if there is a mild reduction of demand, the employer will keep the employment at $\overline{N}$ and allow the production effort (e) to fall automatically (i.e. let the workers sit around in the office). Hence we have shown that :

(a) In the presence of hiring and layoff cost, employment will be invariant to mild demand shocks. It should also be noted that such kind of sticky employment still

---

[10] For a larger demand shocks, employers in the manufacturing industry might also request overtime.

holds even if B and C are zero;

(b) Unlike the case of efficiency wage model [see the criticism by Yellen (1984)], productivity per head (Q/N) is procyclical as employment is invariant to mild demand shocks;

(c) For those jobs whose working hour is officially fixed, productivity per man hour will also be procyclical for mild demand shocks. Even in the case where overtime is introduced with the rise in demand, productivity per man hour may still be procyclical as long as e is raised with the higher demand. Only in the infrequent case where e falls with the overtime (eg. due to exhaustion), will productivity per man hour be counter-cyclical [11]. Because of these, the discussion behind assumption (C) implies that productivity per man hour for the whole economy is likely to be procyclical as well.

(d) Unlike the case of Azariadis(1975), even if there is a permanent reduction of demand, the employer might still hoard the "excessive" workers [12] as long as the reduction of demand is not too large. This is so because the reputation cost of layoff may be even higher than the discounted sum of the cost of holding a small amount of excessive workers.

(B) Large Demand Shock

If the change of demand is so large that $f(\overline{N},Q)$ shift beyond that enclosed by $f(\overline{N},Q_H)$ and $f(\overline{N},Q_L)$, the employer will then lay off (or hire) workers in significant amount. [Note, (i) if there is a sluggish revision of expectation about the persistence of

---

[11] Even so, the reduction will be small. Otherwise, the employer will rather not to have overtime.

[12] If there is a certain rate of retirement or quitting (which has not been explicitly incorporated in our model), the employer might prefer to wait (or raise wages below the trend value to encourage quitting) until the "over employment" falls gradually to zero.

the shock, the massive layoff will only occur after a while [13]; and (ii) as the fixed cost of hiring is relatively small when compared with the reputation cost of layoff, hiring may not be as massive as layoffs [14].] In that case, productivity(per head or per man hour) may rise, remain unchanged or fall – depending on the relative position of the initial $\bar{N}$ and $f(\bar{N},Q_0)$. For example, if the initial $\bar{N}$ is close to $\bar{N}_1$ of diagram (3.8), productivity will fall with the large demand shock. On the other hand, if initial $\bar{N}$ is close to $\bar{N}_2$, productivity will rise with the shock. In other words, productivity may not be procyclical when massive layoff occurs.

Thus, combining the results in the cases of mild and large demand shocks, we have shown the following bang-bang employment response to negative demand shocks[15]:

(a) within a certain range of reduction in demand, employment will be invariant; and

(b) beyond that range, layoffs will be massive.

With regard to the cyclical movement of productivity, we have shown that it is procyclical with respect to mild demand shocks. It is thus not surprising that empirical works will find some evidence of procyclical productivity. Nevertheless, as shown in case (B), productivity may not be procyclical when the demand shock is large enough to

---

[13]If we assume (i) demand will fall further with reduction in employment; and (ii) there is a substantial lag in the linkage between demand and employment, the economy will only reach the bottom after a substantial lag.

[14]In case C is zero, hiring will be gradual with gradual rise in expected demand.

[15]So far, we have only derived our result for the case of a specific and permanent demand shock. Nevertheless, the bang-bang result should also hold for the case of general demand shock where there is still a reasonably large fixed component of reputation cost of layoff. Perhaps the major difference is that, in the case of general demand shock, the reputation cost of layoff (in terms of higher quit rate or higher recruitment cost in the future) will be lower so that it takes a smaller reduction in demand to cause the massive layoff. The result should also hold for the case of temporary demand shock. This time, the threshold demand shock for the bang-bang turn will be higher than the case of permanent shock. We believe that research in this area will provide a better account than the traditional Keynesian models for the sharp rise in unemployment rate in recessions such as that in U.K. during 1982-1986.

induce a discrete change in employment. This also implies that previous empirical literature in this area should be refined or reformulated to allow for the possibility of case (B).

3.5 Implications for the effectiveness of stabilization policies

Having established the bang-bang employment response to negative demand shocks, let us

(a) check whether it makes much difference, from the point of view of keeping a lower unemployment rate, for the government to stimulate the economy before or after the massive layoff; and

(b) see why mild stimulation policies may not help reducing unemployment in the case of adverse demand shock [16].

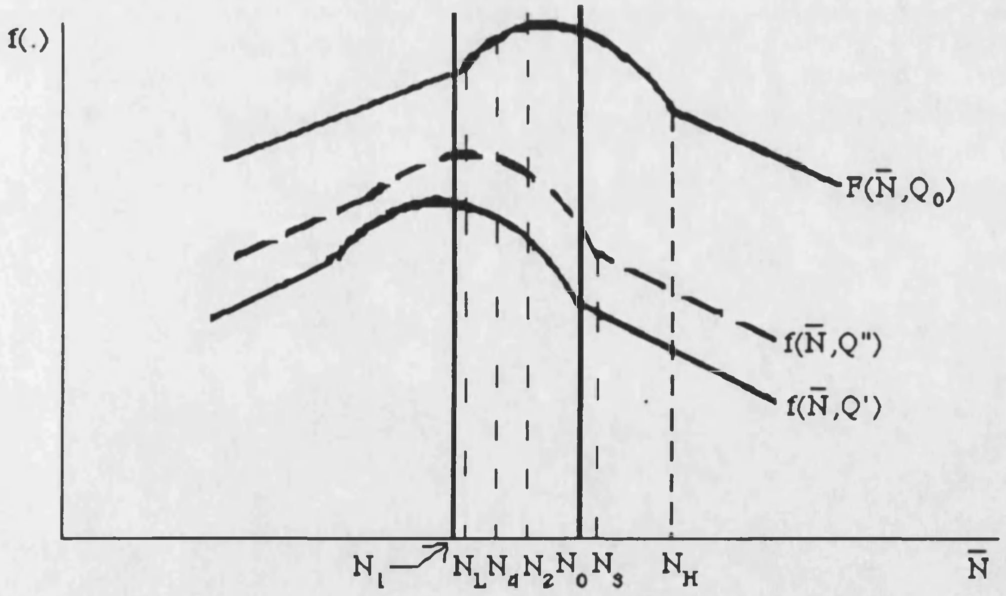


diagram (3.9)

[16]The logic below will help to suggest why, during severe recession (such as those in U.K. or Singapore in 1982-1986), many governments had found their stimulation policies relatively ineffective in reducing unemployment. [Note, because we have not solved for the case of temporary shock, the subsequent discussion is not a formal proof because recession refers to temporary instead of permanent demand shock.]

Consider the hypothetical case in which (i) all producers are expecting the demand to be Q_0 over the future so that their expected discounted profit is shown by $f(\bar{N},Q_0)$ in diagram (3.9); and (ii) the starting levels of employment are somehow (eg symmetrically or triangularly) distributed between N_L and N_H. Suppose there comes a negative demand shock so that every producer expects a reduction of demand to Q' and hence a shift of f(.) to $f(\bar{N},Q')$. Thus,

(a) those producers whose starting level of employment are between N_0 and N_H will cut their employment to N_1 through massive layoff; and

(b) those producers whose starting level of employment are between N_L and N_0 will keep their employment unchanged.

Suppose the government, after seeing the massive layoff, starts stimulating the economy and succeed in shifting f(.) to $f(\bar{N},Q_0)$. In such case, the group of producers who had cut their employment will now raise their employment to N_2. Nevertheless, when compared with the initial situation, their employment is still lower. Now, consider an alternative scenario in which the government succeeds in stimulating the economy back to $f(\bar{N},Q_0)$ <u>before</u> the massive layoff, all producers will then keep their employment unchanged. In this case, the total level of employment will be higher than the case where the stimulation only occurs <u>after</u> the massive layoff. <u>Hence, we propose that it would be preferable – from the point of view of keeping a lower unemployment rate – to stimulate the economy before layoffs occur</u> [17]. The result will still hold for large negative demand shocks (not shown) as long as the fixed cost of layoff is higher than the fixed cost of hiring and the starting levels of employment is symmetrically distributed around the

[17]This is so because stimulation before the massive layoff will induce those employers with excessive labour hoarding to maintain the hoarding instead of making the massive layoff.

middle of N_L and N_H [18].

Let us now discuss the effectiveness of mild stimulation policies after a massive layoff. As illustrated in diagram (3.9), the mild stimulation policy will cause a shift of f(.) from $f(\bar{N},Q')$ to $f(\bar{N},Q'')$. However, no producer will raise his employment because N_1 is still above the lower threshold of $f(\bar{N},Q'')$. Thus, the unemployment rate is not reduced by the mild stimulation policy. Again, let us consider a mild stimulation <u>before</u> massive layoff. In such case, the recession and the mild stimulation will cause a shift of f(.) directly from $f(\bar{N},Q_0)$ to $f(\bar{N},Q'')$. Unlike the stimulation <u>after</u> the massive layoff, only those producers whose starting level of employment is between N_3 and N_H will cut the level of employment to N_4. As a result, there are less producers cutting employment and the amount by which they cut are smaller. Thus, it is again better to have the mild stimulation – say, due to government's worry about inflation – before rather than after the massive layoff.

---

[18]Indeed, the result will hold as long as the mean of $\bar{N}$ lies to the right of N_2.

3.6 Sensitivity of the result with respect to Assumption (A) - interdependence between the other decisions and the employment decision

So far, the results in section 3.4 are derived with the simplified assumption that the price, wage and capacity decisions are determined before the employment decision. Let us now explore the sensitivity of our results to such simplifying assumptions. For example, if the pricing decision is simultaneously determined with the employment decision, the producer's problem is to choose the optimal ΔP and ΔN so that the maximum expected profit $f(\overline{P},\overline{N},\hat{Q}) =$

$$
\max\begin{cases}
H(\overline{P},\overline{N},\hat{Q}) + \gamma f(\overline{P},\overline{N},\hat{Q}) & \text{(maintain P and N)} \\
\sup\limits_{\Delta P\neq 0} H(\overline{P}+\Delta P,\overline{N},\hat{Q}) - [A + \xi(\Delta P)] + \gamma f(\overline{P}+\Delta P,\overline{N},\hat{Q}) & \text{(change P and maintain N)} \\
\sup\limits_{\Delta N<0} H(\overline{P},\overline{N}+\Delta N,\hat{Q}) - [B-\theta(\Delta N)] + \gamma f(\overline{P},\overline{N}+\Delta N,\hat{Q}) & \text{(reduce N and maintain P)} \\
\sup\limits_{\Delta N>0} H(\overline{P},\overline{N}+\Delta N,\hat{Q}) - [C+\psi(\Delta N)] + \gamma f(\overline{P},\overline{N}+\Delta N,\hat{Q}) & \text{(raise N and maintain P)} \\
\sup\limits_{\Delta P\neq 0}\sup\limits_{\Delta N<0} H(\overline{P}+\Delta P,\overline{N}+\Delta N,\hat{Q}) - [A+\xi(\Delta P)] - [B-\theta(\Delta N)] + \gamma f(\overline{P}+\Delta P,\overline{N}+\Delta N,\hat{Q}) & \text{(reduce N and change P)} \\
\sup\limits_{\Delta P\neq 0}\sup\limits_{\Delta N>0} H(\overline{P}+\Delta P,\overline{N}+\Delta N,\hat{Q}) - [A+\xi(\Delta P)] - [C+\psi(\Delta N)] + \gamma f(\overline{P}+\Delta P,\overline{N}+\Delta N,\hat{Q}) & \text{(raise N and change P)}
\end{cases}
$$

where $H(P,N,\hat{Q}) = (P-b)[\hat{\alpha}-\beta P]-a-w\overline{h}N-g[k^{-1}(\hat{\alpha}-\beta P)/\overline{h}N]\overline{h}N$
(ie the expected profit at period t)

Hence, we can separate demand shocks into three different ranges :

(a) "small" demand shocks

For a reasonably mild demand shock, we expect the producer will choose to keep

P and N unchanged [19]. In such a case, the invariance of employment with respect to a mild demand shock holds.

(b) "large" demand shock

In the case of a reasonably large reduction of demand, the producer would find it worthwhile to change both P and N. Thus, our "bang-bang" result still holds for the case of a very large demand shock.

(c) "medium" demand shocks

In case of a medium-sized reduction of demand, the result will depend on

(i) the relative size of the cost of reducing price and the cost of layoff; and

(ii) the relative contribution between a reduced price and reduced employment to the expected stream of profit $H[P,N,\hat{Q}]/[1-\gamma]$.

Suppose demand is very elastic (i.e. an optimal reduction of price with respect to the demand shock will contribute a lot to $H[P,N,\hat{Q}]/[1-\gamma]$) and the cost of layoff is greater than the cost of reducing price so that the second expression of equation (3.5) always dominates the third expression, the decision of reducing price will always precede the decision of layoff [20]. If price is reduced, output will rise which will in turn make it less worthwhile to cut employment (as the problem of redundancy is less urgent). In such a case, the possibility of a "bang-bang" decision on price raises the stickiness (or invariance) of employment[21] towards negative demand shocks [22].

---

[19]We assume the starting price and employment ($\overline{P},\overline{N}$) are not too close to the trigger points.

[20]This can be consistent with the possibility that the decision of hiring always precede the decision of raising price with respect to a favourable demand shock.

[21]There would also be repercussions from the employment decision to the pricing decision.

[22]Or more correctly, the possibility of a "bang-bang" decision of price raise the critical size of adverse demand shock that would cause the "bang-bang" reduction of employment.

As changing the capacity also involves a sunk cost, the joint decision of planned capacity and employment will be qualitatively similar to the pricing-employment decision mentioned above. Indeed, we suspect that

> the need to simultaneously consider the employment decision with the others will only change the critical size of demand shock, below which employment will be invariant to demand shocks.

Nevertheless, a rigorous proof of such hypothesis is beyond the scope of this thesis.

3.7 Remarks [23]

With the help of the model, let us explore the sources of cyclical unemployment and the effect of other policy options on the employment decision. It should be emphasized that, because of the complexity of the problem, all the remarks here remain tentative.

3.7.1 Source of Cyclical Unemployment

Unlike Akerlof and Miyazaki (1980), we believe cyclical variations in unemployment rate can result despite the implicit guarantee of employment by employers.

The reasoning is as follows. First, the difference in risk aversion between employers and workers may not be that great for the state of very low demand [24]. This

[23] In this section, we sometimes refer our discussion to cyclical demand shocks instead of just permanent negative demand shocks. It must be emphasized that we have not yet solved for the case of temporary (cyclical) demand shocks. Nevertheless, it appears to the writer that the solution procedure, though somewhat more tedious, is basically similar to that with permanent demand shocks.

[24] Schultz(1985) suggests that guarantee of employment can only be partial because risk aversion may be reversed in case of very adverse cyclical demand shocks. While it is not sure whether the risk aversion is reversed when employers are making heavy loss, we think that it is not necessary (though sufficient) to rely on the risk reversal to explain why guarantee of employment is partial. Indeed, (i) employers' preference on flexibility, and (ii) workers distrust on employer's ability and incentive to honour the insured

makes the potential gain from insured employment against the states of very adverse demand shocks rather small. Because of the preference for flexibility, employers may not – at the very beginning – find the small reduction of wage worthwhile for the promise of fully insured employment against the states of very adverse demand shocks. Even if they do, they might find it worthwhile to break the promise in case of general adverse demand shock because he (rightly) expects many other employers will do the same thing as he does. Workers, seeing this as a possibility, will not accept a wage cut even if some employers "claim" to provide the insured employment against very adverse demand shocks [25]. In other words it is not the risk shifting between workers and employers, but rather (i) the worker's fear that the employer may not honour the guarantee of employment; (ii) the employer's preference for flexibility; and (iii) the small potential gain form insured employment against adverse demand shock that make the scheme of insured employment only partial instead of full.

Having this in mind, our result in section 3.5 suggests that massive layoff may happen in the case of an adverse demand shock. Thus, a large increase of unemployment (in the case of adverse demand shock) can occur even though there is apparently an implicit guarantee of employment simply because the guarantee we can have in this world is usually unenforceable and therefore partial.

The next thing we have to explain is the fact that cyclical variations of unemployment also appear to coexist with insured employment even in the case of mild variations in demand. This should not be too surprising if one recognizes the fact that the

---

employment in case of adverse demand shock may be sufficient to make it not worthwhile to have a full guarantee of employment.

[25]Government may be the only exception because civil servants know, explicitly or implicitly through their previous observations, that the government have the incentive to maintain employment so as to avoid further reduction in aggregate demand. In other words, the government is "rational" in the sense that it can achieve its stabilization aim and can afford to pay lower wage by providing a full guarantee of employment to the civil servants.

guarantee of employment is provided only to those already employed in the firm but not to those outside the firm. Consider an economy in static equilibrium where there are some old workers retiring from the labour force and some new workers entering the labour force. Suppose there comes a mild negative demand shock. The employer, due to the reputation cost of layoffs, will try to honour his promise of insured employment to those already in the firm. However there is no need for the employer to provide insurance to those outside the firm. Hence there is no reputation cost in delaying (or suspending) new recruitment in case of a mild negative demand shock. What the employer can do is to let the old workers retire and delay new recruitment. This explains why most of the burden of recession will fall on school leavers or new applicants. Similarly, if we have a pool of involuntary unemployment [26], the employer can speed up his recruitment in the case of mild positive demand shock. These explain why, even if the guarantee of employment is perfectly honoured in the case of a mild demand shock, we still have cyclical variations in the level of employment or the unemployment rate.

Nevertheless, it must be noted that the variations of employment here will be limited by the quit rate in the case of a mild reduction of demand. In other words, the extent of cyclical variations of the unemployment rate here is far from proportional to that in the case of a very adverse demand shock. Indeed, the difference is so great that it is preferable to emphasize the difference between mild and adverse demand shocks rather than the difference of cyclical or sticky variations of employment in case of mild demand shock.

Extending the above analysis to the case where there are some firms near the margin of layoff (eg. declining industries) and some firms near the margin of hiring (eg. expanding industries), we get another explanation for the coexistence of cyclical

---

[26]The traditional efficiency wage, labour turnover and shirking models reviewed in chapter 1 can be used to explain the existence of a pool of involuntary unemployment.

variations in unemployment with the insured employment. The reasoning is as follows. With a moderate positive demand shock, those firms near the margin of hiring will be induced to hire (or speed up the expansion) while the other firms will keep employment unchanged. As a result, the overall unemployment rate will show a small reduction with the positive demand shock. Similarly, a moderate negative demand shock will induce those firms near the margin of layoff to start the layoff (or speed up the contraction), leading to a small rise in the overall unemployment rate.

Finally, cyclical variations of the unemployment rate can also arise from the casual labour market where there is little cost of training or layoff. Nevertheless, it must be emphasized that, given the size of career market such as that in the world, the cyclical variations in the overall unemployment rate with respect to mild demand shocks is far from proportional to that with respect to very adverse demand shocks.

3.7.2 Presence of Other Policy Options

So far, the model built in section 3.2 – 3.5 only allows the employer to choose between (i) changing the production effort; and (ii) changing the employment level. In reality, the employers will have more policy options other than the two mentioned above. For example, if we introduce inventories, the producer might – in the case of a recession – decide (at least at the beginning of the recession) to build up inventories instead of making massive layoffs. In such a case, the presence of inventories may increase the stickiness/invariance of employment to negative demand shocks [27,28].

---

[27]It would be interesting to show that the possibility of inventory will increase the threshold of the bang-bang employment response. Such kind of formulation will however be highly complicated and is beyond the scope of this thesis.

[28]Of course, if the negative demand shock is persistent enough, the cost of holding the continuously increasing inventory may be so high that employer will make the layoff eventually.

In addition to the building up of inventories, the producer also has the choice of asking his workers to do some maintenance work (such as cleaning, repairing and painting). The presence of such options is again likely to increase the stickiness of employment to negative demand shocks.

In setting up equation (3.1), we have assumed that :

(i) there is no retirement and quitting of workers; and

(ii) workers are homogeneous.

Suppose we relax assumption (i) by assuming a fixed amount of retirement (and quits) R_N every period, we will then have to replace all the $\overline{N}$ on the right hand side of equation (3.4) by $(\overline{N} - R_N)$. Thus, when the demand is at the normal level, the producer will regularly replenish his pool of workers by a discrete amount that depends on the fixed cost of hiring. In case of negative demand shock, the producer can however halt the recruitment of new workers and let the old workers retire [29]. When demand recovers, the producer can then pick up the previous lag by making larger recruitment [30]. As there is no additional cost for not recruiting, our theory suggests that the producer will have a high tendency to defer recruitment in case of recession. Indeed, this appears to be what is happening in the world.

We can also relax assumption (ii) by assuming heterogenous labour. Indeed, even if all the workers (after training) have the same productivity and they only

---

[29] As noted by Okun(1981), quitting will be low in such case.

[30] By the same argument, producers will speed up the recruitment and slow it down subsequently when the demand is at the expansion part of the cycle.

(apparently) differ from each other by their number of years with the firm [31], producers can – in the case of negative demand shock – choose the less costly last-in-first-out policy of layoff. That is, producers can ensure established worker's attachment by developing the practice that only those new workers will be laid off in case of adverse demand shocks. The advantage of this policy over random layoff is that it avoids cost of quitting of established workers when the economy recovers. Although it might increase (i) the cost of recruitment; or (ii) the promise of legal compensation in future layoff [32] when potential applicants get to know such policy, the advantage of avoiding the turnover cost from the quitting of the established workers is likely to be dominating. This explains why, in the case of very adverse demand shocks that some kind of layoffs have to be made, many producers will choose to retrench the new workers. Mathematically, if the representative producer has established [33] a last-in-first-out employment policy, we have to replace $[B-\theta(\Delta N)]$ and $[C+\psi(\Delta N)]$ of equation (3.4) respectively by $[B'-\theta'(\Delta N)]$ and $[C'+\psi'(\Delta N)]$, where $[B'-\theta'(\Delta N)] < [B-\theta(\Delta N)]$ and $[C'+\psi'(\Delta N)] < [C+\psi(\Delta N)]$.

In addition to the above, producers – at the beginning – can also choose to (i) have some optimal percentage of casual workers who will receive a higher wage in compensation for the absence of any guarantee of employment; or (ii) subcontract part of its production to other firms. Thus, despite the higher payment in case of normal

[31]In case the productivity of workers differ, producers might develop the practice of retrenching the least productive worker so that the more productive workers have little incentive to quit when the economy recovers. Nevertheless such practice might have to be limited to the very unproductive worker. Otherwise, the productive workers and the new applicants who believe they are reasonably productive will still worry about a layoff in the case of extremely adverse demand shocks.

[32]Alternatively, producers with an established last-in-first-out employment policy can offer higher wage or additional payment to the new workers. This implies we should replace $[B - \theta(\Delta W)]+WN$ by $[(\Delta W)N_n + WN]$,where ΔW is the additional payment to the new workers and N_n is the number of new workers.

[33]Before the producer's first attempt to retrench the new workers there will not be any rise in recruitment cost in case of recession.

demand, the producer can avoid the cost of layoff or excessive payroll in the case of recession. In general, the greater the variations of demand, the greater the percentage of casual workers and/or subcontraction [34].

Finally, our model and hence equation (3.4) can also be extended to include, in addition to the choice of massive layoffs, the choice of part time working schedule whose introduction will also involve a fixed reputation cost. Thus, in the case of adverse recession, the producer can have the choice of massive layoff and part time working schedule. Which one the producer chooses will depend on whether the cost of layoff is greater or smaller than the cost of a part-time working schedule, and this will in turn depend on the nature of his business and the expected length of the recession. This explains why some producers choose to retrench workers while some others choose to have a part-time working schedule. In addition, if the recession is really adverse so that it is worthwhile to close a plant, a producer might announce an overall dismissal subjected to recall.

Having reviewed the few policy options other than hoarding (with production effort changing) and layoffs, we believe that producers – in the case of negative demand

[34] One very good example of this is Hong Kong where the manufacturing industry is very much export-oriented. When there is no order from overseas, there exist redundancy in capacity and labour. However, when the order comes, it will generally be a big one. To find somebody sharing the risk of excessive payroll in case of no order from overseas, the exporter (usually the trading company) will try to subcontract the production to some manufacturing firms who might further subcontract part of their production to smaller firms. Thus, a typical trading company is effectively working as the marketing department while the manufacturing firms are effectively working as the various production lines of a big firm [in addition to the advantage of spreading the risk arising from the large variations in demand, such a kind of arrangement is very effective in cutting the cost, because each firm is very specialized in its production and knows very well about the various opportunities in cutting the costs (eg. distributing the sewing to the housewives in the housing estate and those in China with a much lower labour cost). In addition to the subcontraction, the manufacturing firms will as well hold only a few regular workers in the firm. When demand is high, the producer will try to ask the regular workers to work overtime. If the order is really big, the producer will also employ some casual workers with a higher wage. When demand is low, the producer will stop hiring the casual labour and the amount of labour hoarding is only limited to the few regular workers.

shocks – will try to avoid the reputation cost of layoff by

(i) stopping new recruitment ;

(ii) stopping or reducing subcontraction;

(iii) retrenching casual labour;

(iv) building up inventory;

(v) asking the workers to do some maintenance work such as cleaning, repairing and painting; or

(vi) simply keeping the insured workers in a state of on-the-job unemployment.

If the demand shock is really adverse, they might start to:

(i) retrench the new or least productive workers;

(ii) introduce a part-time working schedule;

(iii) close a plant and announce overall dismissal subjected to recall; or

(iv) make massive layoffs.

Which option they choose will depend on the cost and benefit of the policies, which in turn depend on the nature of the firm.

3.8 Conclusions

In the early version of Implicit Contract Theory [Bailey (1974), Gordon (1974) and Azariadis (1975)], real wage rigidity was considered/shown to be a result of implicit contractual arrangements between risk averse employees and less risk averse employers. The result of such insured (sticky) wages were then used by macroeconomists [such as Gray (1976,1978), Poole (1976) and Fischer (1977) to account for non-neutrality of monetary policies on aggregate output and employment. Akerlof and Miyazaki (1980), however, challenged such an explanation of Keynesian unemployment by arguing that

> *" it pays both parties if the less risk averse employers insure the risk averse employees against layoff; and in return the employers are compensated by risk premia in the form of lower average wages which workers are implicitly willing to pay for such employment insurance [Akerlof and Miyazaki (1980)]."*

They then suggest that

> *" Because both firms and workers prefer full-employment to layoffs in any given state, it follows that unemployment cannot occur in an equilibrium with rationally negotiated contracts."*

The contribution of Akerlof and Miyazaki (1980) is that it highlights and explains the phenomenon of insured employment to those employed within the firms. Their second argument (from insured employment within the firms to full employment for the whole economy) is however subject to debate.

There are at least two important lines of criticism that cause us to believe that involuntary unemployment might coexist with insured employment. The first is due to the possibility of involuntary unemployment in the traditional efficiency wage, shirking and turnover cost models discussed in Chapter 1. Thus, although employers find it beneficial to provide insured employment to those within the firm, they might not reduce the wage to the full employment level as such a reduction in the wage may reduce production effort, raise shirking and increase turnover cost (ie the existence of involuntary unemployment is due to the high wage policy of every employer who initially attempts to offer a higher relative wage to encourage high production effort, discourage shirking and reduce turnover which is in fact achieved by the pool of involuntary unemployment associated with the high wage policy of all employers). In such a case, a guarantee of employment to those employed (within the firms) may coexist with a (static) pool of involuntary unemployment (in the economy).

The second criticism of Akerlof and Miyazaki (1980) is that the implicit guarantee of employment may only be partial instead of full. That is, in the case of very adverse demand shocks, employers may find the cost of layoff much lower than the cost of maintaining the excessive wage bill so that the decision of layoff is preferred. The aim of this chapter is to build an explicit model on the employment decisions of firms where the reputation cost of layoff is finite. The following results were derived:

(1) The employment decision of a representative employer will be a bang-bang solution:

> In the case of a moderate reduction in demand, employers – because of the cost of layoff arising from the implicit guarantee of employment – will attempt to hoard the excessive amount of labour. However, if the reduction in demand exceeds a certain threshold, there will be massive layoffs.

(It also appears that, if the demand shock is general, the reputation cost of layoff for each employer will be lower and hence the more likely such significant layoffs will be. This is probably why general negative demand shocks are usually accompanied with significant layoff in many firms.);

(2) In the case of a moderate demand shock with labour hoarding, production effort – instead of wage and employment – will be the variable of adjustment. This suggests that the Keynesian and Classical debate of Price versus Quantity Adjustment in the

labour market may be misleading. The results also suggest that efficiency wage models and the traditional Keynesian models can be refined to produce procyclical productivity; and

(3) In the case of a very adverse demand shock, it is always better, from the view of employment, to stimulate the economy before rather than after employers layoff their workers.

Thus, in contrast to Akerlof and Miyazaki (1980), our bang-bang solution suggests that unemployment can result despite the implicit guarantee of employment by employers. With these results in mind, the Remarks in section 3.7.1 suggests how cyclical unemployment can evolve despite the presence of an implicit guarantee of employment to those within the firms.

Chapter 4

4.1 Introduction

In Chapter 2, we have built a model in which the degree of price stickiness/ sluggishness depends on the size of the cost of changing price. In this chapter, we will try to provide a rough estimate of the cost. The magnitude of such an estimate may also help to ascertain whether reputation cost or menu cost is the more important component in the total cost of changing price. In Akerlof and Yellen (1985a,b), Mankiw (1985) and Blanchard and Kiyotaki (1987), it was argued that a "second order" menu cost may be sufficient to explain a "first order" degree of price stickiness. In Chapter 1, we have criticized their underlying assumptions and suggested that it is the much higher reputation cost that causes the extensive degree of price stickiness/sluggishness in the world. To find an empirical support for our argument, we will try to show that the estimated size of the cost of changing price is much higher than that could be explained by the menu cost. Along with the estimation, we would also like to check whether the planned profit mark up over cost is constant (such as that suggested by Nordhaus and Godley (1972), who claim to have found empirical support for the normal cost hypothesis) or an increasing function of demand (such as Gordon (1975)).

4.2 Derivation of the Price Equation

As data on expected profit in the future is not available, a test of equation (2.6) of Chapter 2, rewritten as equation (4.1) here [1],

[1] As (i) the linear component will not affect the gap between the thresholds of changing price, and (ii) the distance between the P_t^* with the linear component and that without the linear component will possibly be picked up a constant term, we will omit the linear component for simplicity.

$$f(\bar{P}_t,\hat{\alpha}_t) = \max \begin{cases} \sup\limits_{P_t \neq \bar{P}_t} (P_t - b_t)Q(P_t, \hat{\alpha}_t) - a_t + \rho\, f(P_t, \hat{\alpha}_t) - A_t \\ (\bar{P}_t - b_t)Q(\bar{P}_t,\hat{\alpha}_t) - a_t + \rho\, f(\bar{P}_t,\hat{\alpha}_t) \end{cases} \tag{4.1}$$

is not possible[2]. Thus, we will first derive the price equation from equation (4.1) and test the model through such an equation.

4.2.1 The Price Equation without inflation

As inflation is fairly high within the sample period chosen, one should derive the price equation from a model with explicit consideration of expected inflation. However, as mentioned in Chapter 2, the explicit solution of such a model is in itself a difficult problem. Instead, we will approximate it by solving the model without inflation in this section and then add the effects of inflation in section 4.2.2.

Following the solution procedures in the Appendix of Chapter 2, equation (4.1) can be simplified to

$$(P_t - b_t)Q(P_t, \hat{\alpha}_t) - a_t$$

$$f(\bar{P}_t, \hat{\alpha}_t) = \max \begin{cases} \sup\limits_{\Delta P \neq 0} \dfrac{(P_t - b_t)Q(P_t, \hat{\alpha}_t) - a_t}{1-\rho} - A_t \\ \dfrac{(\bar{P}_t - b_t)Q(\bar{P}_t, \hat{\alpha}_t) - a_t}{1-\rho} \end{cases} \tag{4.2}$$

(A) First order condition for the Raise Price Regime

To solve for equation (4.2), the first step is to obtain the optimal P_t, denoted as

---

[2] Another difficulty of estimating equation (4.1) is that only one of the first and second expressions on the right hand side of (4.1) will be observable at any point of time: if the producer has chosen to raise price, only the first expression would be observable; if the producer has chosen to keep the price, only the second expression would be observable. Such a problem does not exist for the corresponding price equation.

P_t^*, for the "raise price regime" expression on the right hand side of equation (4.2). Differentiating

$$H(.) = \frac{(P_t - b_t)Q(P_t, \hat{\alpha}_t) - a_t}{1 - \rho} - A_t \tag{4.3}$$

with respect to P_t and setting the derivative to zero, we have

$$\frac{(P_t^* - b_t)\left.\frac{\partial Q(.)}{\partial P_t}\right|_{P_t = P_t^*} + Q(P_t^*, \hat{\alpha}_t)}{1 - \rho} = 0 \tag{4.4}$$

$$\Rightarrow \quad \frac{P_t^* - b_t}{P_t^*} = \frac{1}{\frac{P_t^*}{Q(P_t^*, \hat{\alpha}_t)}\left(-\left.\frac{\partial Q(.)}{\partial P_t}\right|_{P_t = P_t^*}\right)} = \frac{1}{\theta_t} \tag{4.4'}$$

where $\theta_t = \theta(P_t^*, \hat{\alpha}_t)$ is the elasticity of demand. Rearranging, we have

$$P_t^* = \frac{\theta_t}{\theta_t - 1} b_t \tag{4.4"}$$

Defining

$$1 + m_t \equiv \frac{\theta_t}{\theta_t - 1} \tag{4.5}$$

equation (4.4") becomes

$$P_t^* = (1 + m_t)\, b_t \tag{4.6}$$

Thus, the planned profit mark-up m_t will be an increasing function of the expected level of demand $\hat{\alpha}_t$ if θ_t is a decreasing function of $\hat{\alpha}_t$. On the other hand, if θ_t is insensitive to the level of demand, the planned profit mark-up will be fixed – which is favoured by the normal cost hypothesis.

Equation (4.6) is almost the same as the equation in the theory of mark-up pricing, except that our theory suggests that equation (4.6) only applies when the condition for raising price is satisfied (i.e. when the first regime expression of equation

(4.2) is greater than the second regime expression). When the condition of raising price is not satisfied, the producer's choice of price will be $\overline{P}_t$ instead of $(1 + m_t)\, b_t$. This explains why, as we proposed in Chapter 1, the theory of mark up pricing is at most an approximation of actual pricing behaviour. Indeed, we believe this also explains why some empirical works such as Gordon (1975), have found that the planned profit mark-up m_t is an increasing function of the level of demand $\hat{\alpha}_t$ whilst others such as Nordhaus and Godley (1972) have found that m_t is invariant to $\hat{\alpha}_t$. We will return to this at the end of the section.

(B) <u>Condition for raising price</u>

The second step in the solution of equation (4.2) is to derive the conditions under which the first regime expression of equation (4.2) will be greater than the second regime expression. If

$$\frac{(P_t^* - b_t)Q(P_t^*, \hat{\alpha}_t) - a_t}{1 - \rho} - A_t > \frac{(\overline{P}_t - b_t)Q(\overline{P}_t, \hat{\alpha}_t) - a_t}{1 - \rho}$$

the producer will raise price to P_t^*. As

$$\begin{aligned}(P_t^* - b_t)Q(P_t^*, \hat{\alpha}_t) &= (\overline{P}_t + \Delta P_t^* - b_t)Q(\overline{P}_t + \Delta P_t^*, \hat{\alpha}_t) \\ &= (\overline{P}_t - b_t)Q(\overline{P}_t, \hat{\alpha}_t) + (\overline{P}_t - b_t)[Q(\overline{P}_t + \Delta P_t^*, \hat{\alpha}_t) - Q(\overline{P}_t, \hat{\alpha}_t)] \\ &\quad + \Delta P_t^* Q(\overline{P}_t + \Delta P_t^*, \hat{\alpha}_t)\end{aligned}$$

the condition for raising the price can be simplified to

$$\frac{(\overline{P}_t - b_t)[Q(\overline{P}_t + \Delta P_t^*, \hat{\alpha}_t) - Q(\overline{P}_t, \hat{\alpha}_t)] + \Delta P_t^* Q(\overline{P}_t + \Delta P_t^*, \hat{\alpha}_t)}{1 - \rho} > A_t \qquad (4.7)$$

Substituting equation (4.4), the first order condition for P_t^*, into the above inequality, we have

$$\frac{(\bar{P}_t-b_t)\{[Q(\bar{P}_t+\Delta P_t^*,\hat{\alpha}_t)-Q(\bar{P}_t,\hat{\alpha}_t)] - \frac{\partial Q(.)}{\partial P_t}\Big|_{P_t=P_t^*}\Delta P_t^*\} + \Delta P_t^*\,[-\frac{\partial Q(.)}{\partial P_t}\Big|_{P_t=P_t^*}]\Delta P_t^*}{1-\rho} > A_t \quad (4.8)$$

To simplify the above inequality, we assume the demand curve is approximately linear between $\bar{P}_t$ and $\bar{P}_t+\Delta P_t^*$ so that

(a) $Q(\bar{P}_t+\Delta P_t^*,\hat{\alpha}_t) - Q(\bar{P}_t,\hat{\alpha}_t) \approx \frac{\partial Q(.)}{\partial P_t}\Big|_{P_t=\bar{P}_t+\Delta P_t^*/2}(\Delta P_t^*)$

(b) $\frac{\partial Q(.)}{\partial P_t}\Big|_{P_t=\bar{P}_t+\Delta P_t^*} \approx \frac{\partial Q(.)}{\partial P_t}\Big|_{P_t=\bar{P}_t+\Delta P_t^*/2}$

Substituting these into inequality (4.8), the condition for raising the price becomes

$$\frac{\Delta P_t^*\,(-\frac{\partial Q(.)}{\partial P_t}\Big|_{P_t=P_tA})\,\Delta P_t^*}{1-\rho} > A_t \quad (4.9)$$

where P_tA is the average of $\bar{P}_t$ and $\bar{P}_t+\Delta P_t^*$.

Suppose (i) the menu cost is relatively insignificant compared with the reputation cost of changing price; and (ii) irregular rise of price will cause an inward shift of demand, we can approximate A_t by

$$A_t \approx A_o\ \bar{m}\ \frac{P_tAQ(P_tA,\hat{\alpha}_t)}{1-\rho} \quad (4.10)$$

where A_o is some percentage <<1;

$\bar{m}$ is the average mark up; and

$P_tAQ(P_tA,\hat{\alpha}_t)$ is the revenue evaluated at the "average" price P_tA

so that $\frac{P_tAQ(P_tA,\hat{\alpha}_t)}{1-\rho}$ is a proxy of discounted revenue over the future.

Thus, the specification assumes the cost of changing price is a percentage of the "average" discounted profit over the future. The presence of the term $P_tAQ(P_tA, \hat{\alpha}_t)$ also implies that the size of a firm will not affect its frequency of changing price. We believe this to be a more satisfactory specification than that of fixed A_t (such as that in the literature of menu cost) which implies firms with greater revenue will have a higher frequency of changing price. This is so because the menu cost in a large firm will be relatively insignificant compared with the change in revenue associated with a change of price to the optimum level.

Substituting specification (4.10) into inequality (4.9),

$$\Rightarrow \frac{\Delta P_t^* \left(-\left.\frac{\partial Q(.)}{\partial P_t}\right|_{P_t=P_tA}\right) \Delta P_t^*}{1-\rho} > A_o \, \bar{m} \, \frac{P_tAQ(P_tA, \hat{\alpha}_t)}{1-\rho}$$

$$\Rightarrow \left(\frac{\bar{P}_t}{P_tA}\right)^2 > A_o \quad \bar{m} \, \frac{1}{\frac{P_tA}{Q(P_tA,\hat{\alpha}_t)} \left(-\left.\frac{\partial Q(.)}{\partial P_t}\right|_{P_t=P_tA}\right)}$$

As $\frac{P_tA}{Q(P_tA,\hat{\alpha}_t)} \left(-\left.\frac{\partial Q(.)}{\partial P_t}\right|_{P_t=P_tA}\right)$ is the elasticity at the "average" price of $\bar{P}_t$ and P_t^*, we approximate this as the average price elasticity for the whole sample $\bar{\theta}$. Using definition (4.5), this elasticity can be rewritten as $(1 + \bar{m})/\bar{m}$. Thus,

$$\frac{P_tA}{Q(P_tA,\hat{\alpha}_t)} \left(-\left.\frac{\partial Q(.)}{\partial P_t}\right|_{P_t=P_tA}\right) \approx \bar{\theta} = \frac{1+\bar{m}}{\bar{m}}$$

Substituting this into the inequality,

$$\Rightarrow \quad \frac{\Delta P_t^*}{P_tA} > \gamma \qquad \text{where } \gamma = \bar{m}\sqrt{A_0/(1+\bar{m})}$$

If we further approximate $\frac{\Delta P_t^*}{P_t A}$ by $(\ln P_t^* - \ln \overline{P}_t)$, the condition for changing price can be written as

$$\ln P_t^* - \ln \overline{P}_t > \gamma \quad \text{where } \gamma = \overline{m}\sqrt{A_0/(1+\overline{m})} \qquad (4.11)$$

(C) The Price Equation with Switching Regimes

Combining the results in sections (A) and (B) [i.e. equations (4.6) and (4.11)], the solution to equation (4.2) will be as follows:

$$P_t^* = (1 + m_t)\, b_t$$
$$P_t = \begin{cases} P_t^* & \text{if } \ln P_t^* - \ln \overline{P}_t > \gamma \\ \overline{P}_t & \text{otherwise} \end{cases} \qquad (4.12)$$

The equation is not yet suitable for estimation because we do not have data on m_t. One way to circumvent the problem is to assume

$$m_t = \ln k_o + k_1 \ln \hat{\alpha}_t \qquad (4.13)$$

where k_o, k_1 are constant and $\hat{\alpha}_t$ is the expected level of demand (to be generated from appropriate data). The intuition of the assumption in equation (4.13) is that 1% change in the level of demand will cause k_1% change in planned profit margin. For those economists who favour a constant mark up, k_1 will be zero. [Note: the expected level of demand $\hat{\alpha}_t$ is different from the actual quantity demanded, $Q(P_t, \hat{\alpha}_t)$. Later in the empirical work, we will have to generate $\hat{\alpha}_t$ from $Q(P_t, \hat{\alpha}_t)$ – see section 4.5.1 for the details].

In order to derive a simple log-linear form price equation, we have to make a linear approximation of $\ln(1+m_t)$. Thus, using a Taylor series expansion around the average mark up $\bar{m}$[3],

$$\ln(1+m_t) = k_3 + k_4 m_t \qquad \text{where } k_3 = \ln[(1+\bar{m}) - \bar{m}/(1+\bar{m})]\,;\ k_4 = k_1/(1+\bar{m}).$$

Substituting equation (4.13) into the above equation, we have

$$\ln(1+m_t) = k_5 + k_6 \ln \hat{\alpha}_t \tag{4.14}$$

where $k_5 = k_3 + k_4 \ln k_0$; and $k_6 = k_1 k_4$.

Taking the logarithm of equation (4.6), we have

$$\ln P_t^* = \ln(1+m_t) + \ln b_t \tag{4.15}$$

As we only have <u>index</u> data on material cost and wage cost, we assume

$$\ln b_t = K + \lambda \ln c_t + (1-\lambda) \ln w_t \tag{4.16}$$

where c_t is the index of material cost per unit output;
w_t is index of wage cost per unit output; and
K is a constant[4].

Substituting equations (4.14) and (4.16) into equation (4.15) and adding the stochastic error term u_t, the price equation for the raise price regime can be written as

$$\ln P_t^* = \zeta_0 + \zeta_1 \ln \hat{\alpha}_t + \lambda \ln c_t + (1-\lambda) \ln w_t + u_t \tag{4.17}$$

where $\zeta_0 = k_5$; and $\zeta_1 = k_4 = 1/(1+\bar{m})$[5].

Replacing the first part of equation (4.12) by equation (4.17) implies that the empirical

---

[3] As m_t can be as high as 50% so that $\ln(1+m_t) \neq m_t$, we expand $\ln(1+m_t)$ around $m = \bar{m}$ instead of $m = 0$.

[4] K will be zero if c_t and w_t are the actual cost instead of index numbers.

[5] If we have actual data instead of index numbers on c_t and w_t (ie K=0), there will be a restriction on γ_0, γ_1, ζ_0 and ζ_1. However, as only index numbers are available, we have to give up such restriction and proceed to get free estimates of the four parameters.

version of the price equation will be

$$\ln P_t^* = \zeta_0 + \zeta_1 \ln \alpha_t + \lambda \ln c_t + (1-\lambda) \ln w_t + u_t$$
$$P_t = \begin{cases} P_t^* & \text{if } \ln P_t^* - \ln \overline{P}_t > \gamma \\ \overline{P}_t & \text{otherwise} \end{cases} \tag{4.18}$$

Thus, if the estimate of $\zeta_1 = k_1/(1+\overline{m})$ is significantly positive, the hypothesis of a constant mark up is rejected and there will be empirical evidence that demand affects the pricing decision through the planned profit margin. It must be noted, however, that a significant $\hat{\zeta}_1$ does not imply that the price equation $P_t = (1+m_t)b_t$ with m_t as an increasing function of demand is the correct one. As our theory and equation (4.18) suggest, the correct price equation is one with switching regimes and $P_t = (1+m_t)b_t$ only holds when the condition of raising price is satisfied. As the debate over a variable or constant mark up neglects the possibility of switching regimes, one can conclude that the debate has a mistaken theoretical foundation and hence, it is not surprising that the associated empirical work has contrasting results for different sample periods and different countries.

Our theory also suggests that the sum of coefficients of $\ln c_t$ and $\ln w_t$ in equation (4.18) should be close to one.

4.2.2 Adjustments for expected inflation

In formulating our model in equation (4.1), it is assumed that there is no expected inflation. However, if expected inflation significantly affects the pricing decision, empirical work with equation (4.16) will not be satisfactory without appropriate adjustments. The reasoning for the adjustment is as follows.

First, suppose that there is no expected inflation, the derivation in section 4.2.1 suggests that price should be raised to $(1+m_t)b_t$ when the condition for raising the price is satisfied. With expected inflation, the producer will raise the price to $(1+m_t)b_t$ plus some "preadjustment" associated with such expected inflation [i.e. to $(1+m_t)b_t(1+\text{preadjustment})$]. Assuming the preadjustment is proportional to the expected inflation [6] (i.e. preadjustment = $\varphi\, \Delta\ln(\mathbf{P}_{wt}{}^e)$, where $\Delta\ln(\mathbf{P}_{wt}{}^e)$ is the expected inflation rate), we have

$$P_t^* = (1+m_t)\, b_t\, (1+\text{preadjustment}) = (1+m_t)\, b_t\, [1+\varphi\, \Delta\ln(\mathbf{P}_{wt}{}^e)] \qquad (4.19)$$

where P_t^* is reinterpreted as the price that would be chosen by the producer whenever the condition of raising price is satisfied.

In addition to the effect on the price chosen in regime 1, expected inflation will also affect the condition of raising price. Instead of waiting for

$$(\beta^0 \ln X_t{}^0 + u_t) - \ln \overline{P}_t > \gamma$$

where $\beta^0 \ln X_t{}^0 = \zeta_0 + \zeta_1 \ln \alpha_t + \lambda \ln c_t + (1-\lambda) \ln w_t$, the producer will raise price whenever

$$(\beta^0 \ln X_t{}^0 + \text{preadjustment} + u_t) - \ln \overline{P}_t > \gamma$$

That is, a producer will raise price whenever changes in demand and cost plus the preadjustment exceed the threshold. If we also believe the cost of changing price will be higher with higher inflation rate,[7] the above condition should be replaced by

$$(\beta^0 \ln X_t{}^0 + \text{preadjustment} + u_t) - \ln \overline{P}_t > \gamma_0 + \gamma_1\, \Delta\ln(\mathbf{P}_{wt}{}^e) \qquad (4.20)$$

---

[6] We assume a higher inflation rate will at most cause a moderate increase in frequency of changing price so that the size of preadjustment rises with the higher inflation rate.

[7] If γ_0 is negligible so that $\gamma = \gamma_1\, \Delta\ln(\mathbf{P}_{wt}{}^e)$, a one percentage point rise in expected inflation will cause a one percentage point rise in threshold. In this case, the frequency of price adjustment will be independent of the inflation rate. On the other hand, a significantly positive γ_0 will imply that price adjustment will be more frequent with a higher inflation rate.

Thus, replacing equation (4.6) by equation (4.19) and inequality (4.11) by inequality (4.20) respectively, the appropriate price equation in a world with expected inflation will be

$$\ln P_t^* = \beta \ln X_t + u_t$$
$$\ln P_t = \begin{cases} \ln P_t^* & \text{if } \beta \ln X_t - \ln \overline{P}_t + u_t > \gamma \ln Z_t \\ \ln \overline{P}_t & \text{otherwise} \end{cases} \qquad (4.22)$$

where $\beta \ln X_t = \beta^0 \ln X_t^0 + \text{preadjustment}$
$= \zeta_0 + \zeta_1 \ln \hat{\alpha}_t + \lambda \ln c_t + (1-\lambda) \ln w_t + \varphi\, \Delta \ln(P_{wt}^e)$; and
$\gamma \ln Z_t = \gamma_0 + \gamma_1\, \Delta \ln(P_{wt}^e)$

4.3 The Likelihood Function

4.3.1 Without stochastic variations in the cost of changing price

We now come to the likelihood function for equation (4.22). Assume that the distribution of u_t is normal with variance σ_u so that the density function is denoted as $f(u_t,\sigma_u)$. For an observation $\ln P_t$ in regime 1, the likelihood of such an observation, $l_t(.)$, is:

$$f(\ln P_t - \beta \ln X_t \mid u_t > \gamma \ln Z_t - \beta \ln X_t + \ln \overline{P}_t) * \text{Prob}(u_t > \gamma \ln Z_t - \beta \ln X_t + \ln \overline{P}_t)$$

with the constraint that

$$\ln P_t - \beta \ln X_t > \gamma \ln Z_t - \beta \ln X_t + \ln \overline{P}_t$$

because equation (4.22) implies $u_t = \ln P_t - \beta \ln X_t$.

Noting that $l_t(.)$ can be simplified to

$$l_t(.) = \frac{f(\ln P_t - \beta \ln X_t, \sigma_u)}{\text{Prob}(u_t > \gamma \ln Z_t - \beta \ln X_t + \ln \overline{P}_t)} * \text{Prob}(u_t > \gamma \ln Z_t - \beta \ln X_t + \ln \overline{P}_t)$$

$$= f(\ln P_t - \beta \ln X_t, \sigma_u)$$

the likelihood for an observation in regime 1 can be written as:

$$\frac{1}{\sigma_u}\phi\left(\frac{\ln P_t - \beta \ln X_t}{\sigma_u}\right)$$

$$\text{s.t. } \ln P_t - \ln \overline{P}_t > \gamma \ln Z_t$$

where $\phi(.)$ is the standard normal density function.

For an observation $\ln P_t$ in regime 2, all we know is that

$$\text{Prob}(\ln P_t = \ln \overline{P}_t) = \text{Prob}(u_t < \gamma \ln Z_t - \beta \ln X_t + \ln \overline{P}_t)$$

$$= \Phi\left[\frac{\gamma \ln Z_t - (\beta \ln X_t - \ln \overline{P}_t)}{\sigma_u}\right]$$

where $\Phi(.)$ is the standard cumulative normal density function.

Hence, the likelihood for equation (4.22) is

$$L_f(\beta,\gamma,\sigma_u|P_t,X_t,Z_t\overline{P}_t) = \prod^{t\in R_1} \frac{1}{\sigma_u}\phi\left(\frac{\ln Z_t - \beta \ln X_t}{\sigma_u}\right) \prod^{t\in R_2} \Phi\left[\frac{\gamma \ln Z_t - (\beta \ln X_t - \ln \overline{P}_t)}{\sigma_u}\right]$$

$$\text{s.t. } \ln P_t - \ln \overline{P}_t > \gamma \ln Z_t \quad \forall\, t\in R_1 \qquad (4.23)$$

Thus, our exercise will be a maximum likelihood estimation subjected to a set of inequality constraints [8].

To highlight the importance of the constraints and see how maximization of equation (4.23) can be simplified, consider the simple case $\ln Z_t = 1$ so that we have a fixed threshold to be estimated. In equation (4.23), there are two forces at play in the estimate of γ:

(a) In the non-constraint part of the equation, the term $\Phi(.)$ will tend to make $\hat{\gamma}$ as high

[8] In earlier attempts to write down the likelihood function, we committed the same mistake as Maddala (1983) (page 164) by neglecting the constraints. As will be explained later, the constraints are important for a proper estimate of γ.

as possible. Indeed, in the absence of the constraints, $\hat{\gamma}$ will tend towards infinity so that all $\Phi(.)$ approach unity;

(b) The constraints of equation (4.25) however state that $\Phi(.)$ has to be less than all, including the smallest, percentages of price ($\ln P_t - \ln \overline{P}_t$) found in the first regime of the sample. Thus, the constraints are effectively giving an upper boundary for $\hat{\gamma}$.

Combining (a) and (b), we know that the maximum likelihood estimate of γ will be the minimum of ($\ln P_t - \ln \overline{P}_t$) found in the first regime of the sample (ie min.{($\ln P_t - \ln \overline{P}_t$), $\forall\ t \in R_1$ }). We can then find the maximum likelihood estimate of β and σ_u by setting the $\tilde{\gamma}$ of the following likelihood to the above estimate of γ:

$$L(\beta,\sigma_u|\tilde{\gamma},P_t,X_t,Z_t\overline{P}_t)=\prod^{t\in R_1}\frac{1}{\sigma_u}\phi\left(\frac{\ln Z_t - \beta \ln X_t}{\sigma_u}\right)\prod^{t\in R_2}\Phi\left[\frac{\tilde{\gamma}\ln Z_t - (\beta \ln X_t - \ln \overline{P}_t)}{\sigma_u}\right] \quad (4.23')$$

Thus, for the simple case where $\ln Z_t = 1$, the constrained maximization of equation (4.23) can be partitioned into two simple parts described above [9].

To see why our model requires explicit specifications of the constraints while the usual tobit model does not, consider the hypothetical case where $\ln \overline{P}_t = 0$ and $\gamma = 0$. Substituting these into equation (4.22), we have

$$\ln P_t = \begin{cases} \beta \ln X_t + u_t & \text{if } \beta \ln X_t + u_t > 0 \\ 0 & \text{otherwise} \end{cases}$$

which is a typical tobit model. Substituting $\ln \overline{P}_t = 0$ and $\gamma = 0$ into the constraints of equation (4.23), we have

$$\ln P_t > 0 \qquad \forall\ t \in R_1$$

---

[9] For the case where $\ln Z_t$ includes other variables, we would favour the single step constrained maximization.

which is <u>automatically satisfied</u>. This is why constraints are not required for tobit models and the likelihood function is usually written as

$$L_T(\beta,\sigma_u|P_t,X_t) = \prod^{t\in R_1} \frac{1}{\sigma_u}\,\phi\left(\frac{\ln P_t - \beta \ln X_t}{\sigma_u}\right) \prod^{t\in R_2} \Phi\left(\frac{\beta \ln X_t}{\sigma_u}\right)$$

For the case where γ has to be estimated, the constraints are no longer automatically satisfied and explicit constraints are required to give an upper boundary of the threshold.

4.3.2 With stochastic variations in cost of changing price

In section 4.3.1, we have written down the likelihood function for the case where γ is assumed to be fixed. In the simple case where $\ln Z_t = 1$, the estimate of γ will depend on one observation only – the observation with the smallest non-zero ($\ln P_t - \ln \overline{P}_t$) in the first regime of the sample. If we have reason to believe that there is stochastic variation in the cost of changing price (eg. a temporarily lower cost of changing price due to metricfication or oil surcharge), this estimate will be biased downward even in large samples. To see this, suppose we have a large enough sample size so that there is an approximately continuous spectrum of ($\ln P_t^* - \ln \overline{P}_t$). Without stochastic variations in γ, the recorded minimum non-zero ($\ln P_t - \ln \overline{P}_t$) will be equal to the true γ. However, with stochastic variations in γ, some observations with ($\ln P_t^* - \ln \overline{P}_t$) $< \gamma$ may still be characterized by an observed rise in price simply because the cost of changing price at these periods turned out to be lower than usual. As a result, the recorded minimum non-zero ($\ln P_t - \ln \overline{P}_t$) will be lower than the true expected value of γ. [Of course, there are also some observations where ($\ln P_t^* - \ln \overline{P}_t$) $> \gamma$ and yet there is no observed rise in price simply because the cost of changing price at these periods turned out to be larger than usual. Nevertheless, the existence of these observations will have no effect on the estimate of γ described in section 4.3.1].

Having recognized that the estimate in section 4.3.1 will be biased downward even in large samples, let us obtain the likelihood function when there is stochastic variations in the threshold. Adding a disturbance term $-v_t$ [10] to $\gamma \ln Z_t$ of equation (4.24) and noting that ($\beta \ln X_t + u_t$) in the inequality of the first regime is equal to $\ln P_t^*$ or $\ln P_t$, we have

$$\ln P_t^* = \beta \ln X_t + u_t$$

$$\ln P_t = \begin{cases} \ln P_t^* & \text{if } \quad \ln P_t \quad - \quad \ln \overline{P}_t > \gamma \ln Z_t - v_t \\ \ln \overline{P}_t & \text{if } \ (\beta \ln X_t + u_t) - \ln \overline{P}_t \leq \gamma \ln Z_t - v_t \end{cases} \tag{4.24}$$

Rearranging, we have

$$\ln P_t = \begin{cases} \beta \ln X_t + u_t & \text{if } \quad v_t \quad > \gamma \ln Z_t - (\ln P_t - \ln \overline{P}_t) \\ \ln \overline{P}_t & \text{if } \ u_t + v_t \ \leq \gamma \ln Z_t - (\beta \ln X_t - \ln \overline{P}_t) \end{cases} \tag{4.25}$$

It can be easily shown that the likelihood function $L_s(.)$ for this equation will be as follows:

$$L_s(.) = \prod^{t \in R_1} \int_{K_t}^{\infty} g(\ln P_t - \beta \ln X_t, v_t)\, dv_t \prod^{t \in R_2} F[u_t + v_t \ \leq \gamma \ln Z_t - (\beta \ln X_t - \ln \overline{P}_t)]$$

where $g(.)$ is the joint density function of u_t and v_t;

$F(.)$ is the cumulative distribution for $u_t + v_t \ \leq \gamma \ln Z_t - (\beta \ln X_t - \ln \overline{P}_t)$

$K_t = \ \gamma \ln Z_t - (\ln P_t - \ln \overline{P}_t)$

Assuming that (i) $v_t \sim N(0, \sigma_v^2)$; (ii) $u_t \sim N(0, \sigma_u^2)$; (iii) u_t and v_t are independent of each

10: As we will assume that v_t is symmetrically distributed around zero, it does not matter whether we add the disturbance term in the form of v_t or $- v_t$.

other; and writing $\sigma_w^2 = \sigma_u^2 + \sigma_v^2$,we have

$$L_s(.) = \prod^{t\in R_1} f(\ln P_t - \beta \ln X_t, \sigma_u^2) \int_{K_t}^{\infty} f(v_t, \sigma_v^2)\, dv_t \prod^{t\in R_2} \Phi\left[\frac{\gamma \ln Z_t - (\beta \ln X_t - \ln \bar{P}_t)}{\sigma_w}\right]$$

where $\Phi[.]$ is the cumulative normal density function;

$f(u_t, \sigma_u^2)$ and $f(v_t, \sigma_v^2)$ are the normal density functions with variance σ_u^2 and σ_v^2 respectively.

$$\Rightarrow L_s(.) = \prod^{t\in R_1} \frac{1}{\sigma_u} \phi\left(\frac{\ln P_t - \beta \ln X_t}{\sigma_u}\right)\left[1 - \Phi\left(\frac{\gamma \ln Z_t - (\beta \ln X_t - \ln \bar{P}_t)}{\sigma_v}\right)\right] \prod^{t\in R_2} \Phi\left[\frac{\gamma \ln Z_t - (\beta \ln X_t - \ln \bar{P}_t)}{\sigma_w}\right] \tag{4.26}$$

Unlike the case in section 4.3.1, the estimate of γ of the above likelihood function will depend on all observations [with heavier weights for small $(\ln P_t - \ln \bar{P}_t)$ in the case of regime 1 and large $(\beta \ln X_t - \ln \bar{P}_t)$ in the case of regime 2]. We expect such an estimate to avoid the problem of downward biasedness mentioned above.

Maximization Routine

In carrying out the maximum likelihood estimations of equations (4.25) and (4.28), the Newton's method with analytic second order derivatives is employed. This routine is available in commercial packages such as the Time Series Processor, Version 4.1 (TSP 4.1).

4.4 Availability of Data and Choice of Product

Unlike most empirical work on aggregate price sluggishness, the testing of our theory requires disaggregate data, preferably that of a single product [11]. The product chosen must have reasonably well published data on price, cost and demand for a

[11]: As mentioned in Chapter 1, aggregation of individual prices tends to smooth out the discrete jumps predicted by our theory.

reasonably long period. Also there must have been little change in quality, definition or description during the sample period. No product appears to satisfy all these requirements and we have to choose among a few products which, after suitable assumptions or approximations, are best suited for empirical work. After careful consideration, we decided to choose the steel beams – one of the many products by the British Steel Corporation[12]. The sample period is from 1970:1 to 1979:12 and from 1980:8 to 1982:12 because (i) the period 1970:1 to 1982:12 is the only one in which data on sales (net delivery) of steel beams is available; and (ii) the steel worker strike from 1980:1 to early 1980:4 caused exceptionally low delivery in this period and somewhat erratic movements of delivery between early 1980:4 and 1980:7. In most of the regressions reported later, the sample period will be as mentioned. However, as it is necessary to use regressions to generate data for the estimation of (4.14) – (4.26), the sample period of these regressions may be slightly different due to the presence of leaded and lagged variables.

As steel beams do not quite possess the "perfect" features listed above, the following assumptions or approximations are made:

(a) We assume (i) the material-output ratio for steel beams has remained unchanged within the sample period; and (ii) the cost index for steel beams moves fairly closely with that of the iron and steel industry [13]. These assumptions imply c_t (the unit material cost of steel beams) can be approximated by the material cost index for the

12: This implicitly assumes the pricing decision of steel beams by the British Steel Corporation is independent of the pricing decision of other steel products. While it is not certain how strong this assumption is, the absence of well published data in other industries dictate a choice within the iron and steel industry.

13: A trend variable can be added to equation (4.24) to pick up (i) any trend divergence between the two cost indice; and (ii) and trend difference between the two material–output ratios. However, as reported later, addition of a trend variable in equation (4.24) does not appear to be significant. This implies the material cost per unit of steel beam can be approximated by the material cost index of steel beam which can be approximated by the material cost index of the iron and steel industry.

iron and steel industry.

(b) As data about the unit wage cost of steel beams is not available, data for the metal industry, for which the iron and steel industry is the main provider, is used to proxy that for steel beams. In section 4.5.2, we describe the generation of the unit labour cost series.

(c) Expected inflation is proxied by the annual inflation rate of the producer price index between t-1 and t-13 [i.e. $\Delta \ln(P_{wt}^e) = (P_{wt-1} - P_{wt-13})/P_{wt-13}$].The only exceptions are those at 1974:4 and 1979:6-1979:12. During these periods, there were sharp rises in the oil price and hence expected inflation rates at these periods are likely to be higher than the lagged inflation rates. Seeing this, we will approximate the expected inflation rate at 1974:4 by the actual inflation rate of 1974:4; and the expected inflation rate at 1979:6-1979:12 by the actual rate of 1979:12.

(d) Delivery data of steel beams for each month will be used to generate the expected demand.

In addition, because the cost index and inflation rate are monthly averages and the revisions of prices did not occur at a fixed date of the month, the following adjustments are made:

(i) If the revision of price happened during the first 15 days of the month t, the unadjusted data X_t are replaced by

$$\tilde{X}_t = ((30 - DAY_t)*X_t + DAY_t*X_{t+1})/30$$

while the data X_{t-1} are replaced by

$$\tilde{X}_{t-1} = ((30 - DAY_t)*X_{t-1} + DAY_t*X_t)/30$$

where: X_t is the vector of unadjusted data;
$\tilde{X}_t$ is the vector of adjusted data
DAY_t is the number of days from the beginning of the month t when the price is raised.

(ii) If the price revision happened during the last 15 days of the month, the cost index and inflation rate at t and t + 1 are replaced by

$$\tilde{X}_t = ((30 - DAY_t)*X_{t-1} + DAY_t*X_t)/30$$

$$\tilde{X}_{t+1} = ((30 - DAY_t)*X_t + DAY_t*X_{t+1})/30$$

where X_t, $\tilde{X}_t$ and DAY_t are the same as that defined in (i)

Last but not least, when there were sharp rises in cost such as that in the oil crisis, administrative or decision lags in raising prices may cause a serious overestimation of the cost of changing price. To avoid such overestimation, we will also report the result that excludes some of the observations which may be subjected to decision lags. This exclusion, if inappropriate, might on the other hand, cause an underestimation of the cost. However, if the estimate is significantly larger than that suggested by the menu cost hypothesis, the exercise can still be interpreted as providing a lower boundary estimate of the cost. Moreover, a check with equation (4.26) will reveal that the problem of underestimation is likely to be less serious than the problem of overestimation [14]. Thus, given our aim of searching for empirical support for the hypothesis of a significant cost to changing price and the trade off between overestimation and underestimation mentioned above, the results with the smaller sample are preferred.

[14]: According to the functional form of $\Phi(.)$ in equation (4.26), a large βlnX_t in regime 2 will have a much greater weight than a small βlnX_t. As a result, inclusion of an inappropriate βlnX_t (eg due to decision lags will cause a rise in the estimate of γ. On the other hand, an inappropriate exclusion of βlnX_t, in the case of a large sample, will have little effect on the estimate of γ.

4.5 The Empirical Result

Before the maximum likelihood estimates of equation (4.23) and (4.26) can be reported, we need to obtain the proxy for expected demand, $\hat{\alpha}_t$ and unit labour costs, w_t. These are made by a series of regressions and approximations reported in sections 4.5.1 and 4.5.2.

4.5.1 Proxy for the expected demand $\hat{\alpha}_t$

When making his price decision, the producer will rely on the expected level of normal demand that will prevail until the next revision of price. This implies current delivery data will not be a good proxy for such demand variable because

(a) the delivery data, even under the assumption of demand determined output, is only the quantity demanded and not the level of demand (ie the price effect has to be removed from the delivery data to give a measure of the level of demand);

(b) Strikes and seasonal effects have to be removed to give a measure of the normal demand; and

(c) Current demand may differ from the demand that is expected to prevail until the next revision of price.

In what follows, we will first remove the effect of strike, price and seasonal variations in section(A); and then generate the expected demand in section (B).

(A) Depriced-destrike-and-deseasonalized demand in the current period

To obtain the current depriced-destrike-and-deseasonalized demand for steel beams, our recommended procedure is to

(i) estimate the demand function with relative prices, industrial demand and aggregate demand as explanatory variables; and

(ii) with the help of the estimated coefficients obtained in (i), remove the price and strike effects from the delivery data on steel beams.

However, the delivery data for the iron and steel industry is also subjected to variations in relative price. To generate the industrial demand for the estimation mentioned in (i), we have to repeat the above procedures (i.e. estimate the demand function and obtain the depriced and destrike demand) for the industry.

Bearing the above procedures in mind, we assume the desired demand for iron and steel ($\ln QM_t^*$) is described by

$$\ln QM_t^* = K_t - \beta \ln (e_t P_t^I / P_t^f) + \text{Strikes} + u_t \qquad (4.27)$$

where K_t is the vector of variables that affect the level of demand for the iron and steel industry, examples of which include aggregate demand and trend variables;

P_t^I is the wholesale price index of the iron and steel industry;

e_t is the exchange rate so that P_t^f/e_t is the European export price of Wide Flange Beam express in sterling;

Strikes is a vector of dummy variables, including the Road Haulage Strike, Coal Miners Strikes, Engineering Strike etc. within the sample period, multiplied by the associated vector of estimated coefficients). As our aim is to obtain a "normal" demand variable, free from the short run effect of strikes, strike dummies with not very significant coefficients are still included if we theoretically believe that the strike has some effect, no matter how small, on observed demand.

We also assume a partial adjustment [15] of $\ln QM_t$ from $\ln QM_{t-1}$ to $\ln QM_t^*$:

$$\ln QM_t - \ln QM_{t-1} = \delta (\ln QM_{t-1} - \ln QM_t^*) \qquad (4.28)$$

Substituting equation (4.27) into equation (4.28), we have

$$\ln QM_t = \delta K_t - \delta\beta \ln (e_t P_t^I / P_t^f) + (1-\delta) \ln QM_{t-1} + \delta(\text{Strikes}) + \delta u_t$$

Using instrumental variable estimation for 1969:10 to 1979:12 and 1980:8 to 1982:12,

---

15: Partial adjustment can arise whenever there are lags for (i) the change in aggregated demand to cause a change in demand for iron and steel; (ii) the change in demand to affect the order and then delivery of iron and steel. This includes the signal extraction problem between persistent and transitory demand shocks. For example, producers who demand the iron and steel may have a lag in recognizing a non-transitory change in their demand. During this period, they will temporarily run down their inventory. Only after they recognize the change is non-transitory will they start making the order for delivery. If we have higher levels of hierarchy between the British Steel Corporation and the final users (eg. existence of wholesalers etc), the lag will be longer.

the best equation we obtain is [16]

$$\ln QM_t = \underset{(-3.72)}{-2.09} \underset{(-2.10)}{- 0.0728} \ln (e_t P_t^I / P_t^f) + \underset{(5.15)}{0.357} \ln QM_{t-1}$$

$$+ \underset{(7.25)}{1.18} \ln y_t \underset{(-8.29)}{- 0.003} t + \delta \text{ (Strikes)} + \delta u_t \qquad \text{(EQ1)}$$

with (i) $R^2 = 0.912$;
(ii) $F(16,135) = 106.07$;
(iii) standard error of regression = 0.059;
(iv) Durbin-Watson statistic =1.805.

where y_t is the seasonally adjusted industrial output index which is used to proxy the aggregate demand [17];
t is the trend variable;
QM_t is the seasonally adjusted output index of the metal industry [18]; and

the values below the estimated coefficients are the t-statistic.

With the estimated coefficients, we obtain the normal level of demand for the metal industry $QMDSDP_t$ by removing the price, strike and lag effects according to the following formula:

$$QMDSDP_t = \frac{1}{0.643} \ln QM_t - \frac{0.0728}{0.643} \ln (e_t P_t^I / P_t^f) - \frac{0.357}{0.643} \ln QM_{t-1} - \text{Strikes}$$

$QMDSDP_t$ will then be used as a proxy of industrial demand in the next two regressions.

16: Note that our only aim here is to generate the industrial demand variable for the regression of the demand function of steel beams which will then be used to generate the expected demand for steel beams. Such an expected demand will then be used in the price equation. With such a diffused linkage, insufficiency in the removal of price and strike effects from the present regression is likely to have weaker effect on the later regressions (i.e. the insufficiency may be absorbed by error terms or other variables in subsequent regressions). For this reason, we will not go into the detail mis-specification tests for those regressions which are primarily used for the generation of variables in subsequent regressions.

17: For those periods such as 1972:1-1972:3, 1973:11-1974:3, 1979:1-1979:2 and 1979:8-1979:10 when the industrial output index are affected by strikes, y_t is taken as the average of output one period before or after the periods of strikes.

18 As the iron and steel industry contain most of the weights in the metal output index, QM_t can be considered as a proxy of the iron and steel output index.

The next step is to estimate the demand function for steel beams with the industrial demand variable generated from EQ1. As the regressor contains the endogenous variables P_t (the price of steel beams), P_{t-1} and the price of another steel product – heavy angle – is used as an instrument of P_t. After some experiments, the best regression for the sample period 1970:1 to 1979:12 and 1980:8 to 1982:12 was found to be

$$\ln Q_t = \underset{(-5.73)}{-6.14} \; \underset{(-7.06)}{-0.726} \ln (P_t/P_{wt}) + \underset{(6.98)}{1.63} \ln y_t + \underset{(7.41)}{0.418} \ln QMDSDP_t + \text{Strikes} + \text{Seasonals} \qquad \text{(EQ2)}$$

with (i) $R^2 = 0.672$;
(ii) $F(25,123) = 10.09$;
(iii) standard error of regression = 0.124;
(iv) Durbin Watson statistic =1.57.

where Q_t is the delivery index of steel beams;
P_t is the price of steel beams per metric ton;
P_{wt} is the wholesale price index for the whole economy; and
Seasonals is the vector of seasonal dummies multiplied my the associated vector of estimated coefficient.

The estimated coefficients of EQ2 have the expected sign [19]. We then, as before, obtain the depriced-destrike-and-deseasonalized demand for steel beam $AQDS_t$ by the following formula:

$$\ln AQDS_t = \ln Q_t + 0.726 \ln (P_t/P_{wt}) - \text{Strikes} - \text{Seasonals}$$

(B)Expected demand

In section (A), we have obtained the normal level of demand $AQDS_t$ at each period. However, what really affects the pricing decision is the expected demand that will

[19]The estimated coefficient of relative price is only -0.726 which is somewhat higher than -1, the maximum level required for equation (4.5) or (4.6) as a profit maximization condition. One explanation for this is that the price effect in EQ2 is a short run relationship and this is what we want to remove from the observed demand. On the other hand, pricing decision (i.e. equation (4.6)) should be based on long run price elasticity which is greater than the short run elasticity under the assumption of sticky customer-supplier relationships. Thus, the fact that short run elasticity so estimated is less than 1 does not necessarily imply any theoretical embarrassment in deriving equation (4.6).

prevail until the next price decision. To generate this demand variable, we first note that the average time between each price revision is 10.4 months. We then generate the variable $AQDSA10_t$ by

$$AQDSA10_t = (\sum_{i=0}^{9} AQDS_{t+i} + 0.4\ AQDS_{t+10})/10.4$$

We then regress $AQDSA10_t$ on lagged values of $AQDS_t$ (the depriced and destrike demand), industrial demand and aggregate demand by the method of ordinary least squares. Lagged values of $AQDS_t$ are found to be insignificant. Thus, the best equation we obtain for the sample period 1970:1 to 1979:3 and 1980:11 to 1982:3 is

$$\ln AQDSA10_t = \underset{(-7.81)}{-\ 5.01} + \underset{(7.86)}{0.304} \ln \overline{QMDSD}_{t-1} + \underset{(11.86)}{1.50} \ln y_{t-1} \qquad \text{(EQ3)}$$

with (i) $R^2 = 0.578$;
(ii) $F(2,125) = 85.65$;
(iii) log likelihood = 160.9; and
(iv) standard error of regression = 0.0696;

where $\overline{QMDSDP}_{t-1} = (QMDSDP_{t-1} + QMDSDP_{t-2} + QMDSDP_{t-3})/3$ [20]

By assuming EQ3 as the generation process for producer's expected demand $\hat{\alpha}_t$, we proxy the variable in period 1970:1 to 1979:12 and 1980:8 to 1982:12 by the following formula:

$$\ln \hat{\alpha}_t = -\ 5.01 + 0.304 \ln \overline{QMDSDP}_{t-1} + 1.50 \ln y_{t-1}$$

4.5.2 Proxy for the unit labour costs w_t

As mentioned in section(4.4), there is no published data about wage rates and unit labour costs for steel beams. Thus, we can only proxy these by that of the metal

---

20: If we replace $\overline{QMDSDP}_{t-1}$ by $QMDSDP_{t-1}$, the estimated coefficient will be less significant and R^2 will be smaller than that reported in (EQ3). One possible reason is that QM_{t-1} and hence $QMDSDP_{t-1}$ is more sensitive to random shocks. As our aim is to obtain the "normal" demand, the average $\overline{QMDSDP}_{t-1}$ is preferred to $QMDSDP_{t-1}$.

industry. Yet, further approximations have to be made before one can generate the unit labour cost variable that is suitable for estimation. The procedures are as follows:

To see how we can construct our wage variable, consider the following sub-sample for the wage index in the metal industry between 1973:7 and 1974:8:

period	73:7	73:8	73:9	73:10	73:11	73:12	74:1	74:2	74:3	74:4	74:5	74:6	74:7	74:8
actual wage index	115	127	127	127	127	127	127	127	127	128	129	131	132	146

which illustrate that there will normally be major revisions of wages every 12 months. Between these major revisions, there is little variation in the wage index. It is therefore hard to believe that the actual wage index in 1974:7 (=132) is a good measure of the "wage pressure" in the producer's mind. Thus, as an approximation, we assume that the "wage pressure" is growing steadily between two major revisions. This implies that the above actual wage index will be replaced by the following adjusted wage index in our regression:

period	73:7	73:8	73:9	73:10	73:11	73:12	74:1	74:2	74:3	74:4	74:5	74:6	74:7	74:8
adjusted wage index	125.8	127	128.6	130.2	131.8	133.3	134.9	136.5	138.1	139.7	141.3	142.8	144.4	146

In the subsequent regressions, we also found that the results are better if we replace the adjusted wage index at 1978:1-1978:3 by that at 1978:4. The explanation for this is that the producers in 1978:1-1978:3 anticipated the failure of the incomes policy and hence a large rise in wage rate.

To generate w_t (the unit labour cost), we have to multiply the adjusted wage rate by the "normal" employment per unit output. We approximate the latter by $NM_t/\overline{QM_t}$, where (i) NM_t is employment in the metal industry at t; and (ii) $\overline{QM_t}$ is the average of metal output from t-6 to t+5 (ie $\overline{QM}_t = \sum_{i=-6}^{5} QM_t/12$) [21] so that

$$w_t = AWR_t * (NM_t / \overline{QM}_t)$$

where AWR_t is the adjusted wage rate.

4.5.3 Adjustments in the material cost index

By tabulating the price of steel beams and the material cost index of the iron and steel industry at the various dates when a revision in prices occurred:

time (day)	70:1 (27)	70:10 (16)	71:4 (11)	72:4 (2)	73:4 (30)	73:10 (16)	74:3 (28)	75:1 (2)	76:2 (1)	76:5 (30)	77:1 (2)	77:7 (10)	78:1 (1)	79:7 (1)	82:1 (17)
price of steel beam	50.11	53.35	56.80	59.90	65.15	71.55	96.35	121.7	137.4	157.3	174.5	184.9	203.5	219.0	246.5
cost index	43.22	45.76	48.84	50.29	54.32	58.63	75.36	92.30	107.9	126.4	134.2	141.3	140.5	168.1	223.2
ratio	1.159	1.166	1.159	1.191	1.202	1.220	1.279	1.319	1.273	1.244	1.300	1.309	1.448	1.303	1.104
$\ln P_1 - \ln \overline{P}_1$		6.3%	6.3%	5.3%	8.4%	9.4%	29.8%	23.4%	12.1%	14.5%	10.4%	5.8%	9.6%	7.4%	12.6%
$\ln C_1 - \ln C_{t-1}$		5.7%	6.5%	3.0%	7.5%	7.8%	25.1%	20.3%	15.6%	15.8%	6.0%	5.2%	-0.6%	17.9%	28.4%

Table (4.1)

We found that the cost index does move fairly closely with the price of steel beams in most of the dates except 1978:1 [22]. This suggests that the actual cost index of the iron and steel industry around 1978:1 may not reflect (even after the adjustment for inflation) the expected cost that will prevail until the next revision of price. As this possibility might

---

21: As we will see, the wage cost that based on $\overline{QM}_t$ is found to perform better than that based on QM_t. This is possibly because $\overline{QM}_t$ is less sensitive to random shocks than QM_t.

22: The ratio of the price of steel beams to the cost index is 1.448 which is exceptionally higher than that at any other date. Besides, despite the 9.6% rise in the price of steel beams, the cost index falls by 0.6%. If we calculate the percentage change of cost and price between 1977:7:10 and 1979:7:1, we get the value of 17.37% and 16.93% respectively which can be considered as a fairly close movement.

seriously distort the estimates, we consider the following two alternatives:

(a) estimating equation (4.26) without adjusting the material cost index between 1978:1 and 1979:6; and

(b) estimating equation (4.26) with the material cost index between 1978:1 and 1979:6 adjusted according to the following procedure:

(i) the cost index at 1978:1 is adjusted to a level so that the "price-adjusted cost ratio" at 1978:1 is equal to the average of 1.309 and 1.303 (i.e. the average of price-cost ratios at 1977:7 and 1979:7, where 1977:7 is the most recent regime 1 observation before 1978:1 while 1979:7 is the most recent regime 1 observation after 1978:1). That is

$$c_{78:1}' = P_{78:1}/1.306$$

where $c_{78:1}'$ is the adjusted cost price index at 1978:1; and
$P_{78:1}$ is the price of steel beam at 78:1.

(ii) the cost indices between 1978:2 and 1979:6 are adjusted so that

$$\frac{\%\text{ change between } c_t' \text{ and } c_{78:1}'}{\%\text{change between } c_{79:7} \text{ and } c_{78:1}'} = \frac{\%\text{ change between } c_t \text{ and } c_{78:1}}{\%\text{ change between } c_{79:7} \text{ and } c_{78:1}}$$

where c_t' is the adjusted cost index between 1978:2 and 1979:6;
c_t is the unadjusted cost index between 1978:2 and 1979:6;
$c_{78:1}$ and $c_{79:7}$ are respectively the unadjusted cost index at 1978:1 and 1979:7

(Note: As the cost index at 1979:7 is assumed to be correct, it follows that $c_{79:7}' = c_{79:7}$)

As it was found that estimation with alternative (b) gave a higher likelihood than alternative (a) (not reported in the next section), the proposed adjustment in the cost index between 1978:1 and 1979:6 is supported.

4.5.4 Maximum Likelihood Estimate of γ

(A) When there is no stochastic variations in the cost of changing price

First consider the simple case where (a) there is no stochastic variations in the cost of changing price; and (b) $\ln Z_t = 1$. According to the discussion in section 4.3.1, the maximum likelihood estimate of γ is the minimum non-zero percentage change of price in the sample. From table (4.1), this happened at 1971:4:11. Thus,

$$\hat{\gamma} = 5.31\%$$

which says that the cost of changing price is of such a magnitude that cumulative changes in cost and demand must make the "desired" percentage change of price greater than 5.31% before the price is actually raised to the "desired" level. Otherwise, the nominal price will remain fixed.

Nevertheless, the date 1971:4:11 is when metricfication is adopted in the quotation of the British Steel Corporation. There is thus a reason to suspect that the cost of changing price at this date is particularly low (i.e. without metricfication, the British Steel Corporation might have raised the price at a later date by a larger amount). If this is true, one should drop the observation at 1971:4:11 and look for the smallest non-zero ($\ln P_t$ - $\ln \overline{P}_t$) in the remaining sample. This implies the maximum likelihood estimate of γ is provided by the observation at 1977:7:10 which suggests that

$$\hat{\gamma} = 5.79\%$$

Whatever one's preference [23], the magnitude of these two percentages suggest that it is much higher than can be reasonably explained by the simple menu cost of changing price.

23: It is also possible, because of some reason we do not know, that the cost of changing price on 1977:7:10 is lower than usual. As we would not have sufficient information about this, the estimation procedures described in section 4.3.2 become more appealing.

This can be seen by calculating $\hat{A}_0$, the percentage of discounted profit above variable cost over the future, with the above estimates. From equation(4.11), we know that

$$\gamma = \bar{m}\sqrt{A_0/(1+\bar{m})}$$

$$\Rightarrow \quad \hat{A}_0 = \hat{\gamma}^2 (1+\bar{m})/\bar{m}^2$$

If $\bar{m}$ (the average profit margin above variable cost) is 50%, $\hat{A}_0$ will be 1.69% or 2.01% for the above two estimates. If $\bar{m}$ is 30%, $\hat{A}_0$ will be 4.07% or 4.84%. Given the huge tonnages of steel beams delivered, such shares of discounted profit would be equivalent to enormous amounts in sterling terms which could not reasonably be attributed to the presence of menu costs.

(B)<u>When there is stochastic variations in the cost of changing price</u>

As explained in section 4.3.2, the estimate of γ reported in the previous section may be biased downward if there are stochastic variations in the cost of changing price. For this reason, we propose to estimate also equation (4.26). [To allow for the possibility that the cost of changing price may rise with inflation, we also assume $\gamma \ln Z_t = \gamma_0 + \gamma_1 \Delta\ln(P_{wt}^e)$.] In table 4.2(a), we report results including the full sample. However, because decision lags in raising price might also unreasonably raise the estimate of γ_0 and γ_1, we also estimate equation (4.26) by excluding those observations when there is suspicion of a decision lag.[24] The result is reported in Table 4.2(b).

24: The following observations are excluded: 1974:2-1974:3; 1974:10-1974:12; and 1981:11-1982:1.

Table 4.2(a)

variable	parameter	estimated coefficient (t-stat)
constant	β_0	- 1.436 (-5.23)
$\ln c_t$	β_1	0.705 (9.59)
$\ln w_t$	β_2	0.296 (3.74)
$\ln \hat{\alpha}_t$	β_3	0.275 (4.92)
$\Delta\ln(P_{wt}^e)$	β_4	0.627 (7.93)
constant	γ_0	0.1308 (3.03)
$\Delta\ln(P_{wt}^e)$	γ_1	0.301 (1.59)
	σ_u	0.0186 (5.02)
	σ_v	0.0768 (3.24)

log likelihood = - 0.731276
n = 146

Table 4.2(b)

variable	parameter	estimated coefficient (t-stat)
constant	β_0	- 1.321 (-4.64)
$\ln c_t$	β_1	0.772 (10.47)
$\ln w_t$	β_2	0.216 (2.67)
$\ln \hat{\alpha}_t$	β_3	0.307 (5.94)
$\Delta\ln(P_{wt}^e)$	β_4	0.670 (7.73)
constant	γ_0	0.0814 (3.43)
$\Delta\ln(P_{wt}^e)$	γ_1	0.200 (1.67)
	σ_u	0.0214 (4.18)
	σ_v	0.0350 (2.66)

log likelihood = 5.52429
n = 138

As we can see from the two tables, all estimated coefficients are of the right sign and satisfy the theoretical restrictions. For example, the sum of the coefficients of $\ln c_t$ and $\ln w_t$ are respectively 1.001 and 0.988, which are close to unity, the expected value of the power index of b_t in the equation $P_t = (1+m_t)b_t$. The significant coefficients of expected inflation also suggest that there is preadjustment of price. Of great interest is the significant coefficient of the expected demand $\ln \hat{\alpha}_t$, which suggests that planned profit mark up is an increasing function of expected demand (i.e. there is evidence against the normal cost hypothesis). The size of the estimates also suggest that 1% change in the expected demand will respectively cause 0.275% and 0.307% rise in the "desired" price.

Next, the magnitudes of $(\hat{\sigma}_u, \hat{\sigma}_v)$ are respectively (0.0186, 0.0768) and (0.0214, 0.0350). Noting that our assumption of non-stochastic variations in the cost of changing price in section(A) is equivalent to the assumption that v_t is negligible when compared with u_t. The relative size of $\hat{\sigma}_u$ and $\hat{\sigma}_v$ reported in both tables reject such an assumption and hence cast doubt on the validity of the estimates reported in section (A).

Let us now turn to the estimates of γ_0 and γ_1. Although the t-statistic on γ_1 are strictly lower than the 95% critical values, they are still quite high at 1.59 and 1.67 and the coefficients are correctly signed. Thus, there is some weak evidence that the cost of changing price rises with expected inflation. Nevertheless, the existence of a relatively large (and significant) constant term $\hat{\gamma}_0$ suggests that doubling the inflation rate will not double the cost of changing price. This implies that price adjustments will be more frequent in the case of higher inflation [25]. Next, for an expected inflation of (say) 10%, table 4.2(b) suggests that the cost of changing price will be equivalent to a 10.14% [= 0.0814 + 0.200(10%)] rise in the "desired" price. This appears to be more reasonable than the 16.09% [= 0.1308 + 0.301(10%)] reported in Table 4.2(a), where the estimates are subjected to the problem of decision lags mentioned in section 4.4. Finally, the estimate of 10.14% [26] here is – as expected from previous discussions – higher than the 5.31% or 5.79% reported in section (A), indicating that the estimates in section (A) may be subject to downward biasedness.

As the t-statistics on γ_1 in tables 4.2(a) and (b) are less than the 95% significance level, we also extend the estimation of table 4.2(b) by dropping γ_1. This is reported in

---

[25]: As explained in the footnote of section 4.2.2, the frequency of price adjustment will be independent of inflation rate whenever γ_0 is negligible so that $\gamma \ln Z_t = \gamma_1 \Delta \ln(P_{wt}^e)$.

[26]: The estimate of 10.14% suggests that $\hat{A}_0$ (the reputation cost expressed as the percentage of discounted profit above variable cost over the future) are respectively 6.17% and 14.86% for $\bar{m}$ equal to 30% and 50%.

Table 4.3:

Table 4.3

variable	parameter	estimated coefficient	(t-statistic)
constant	β_0	- 1.341	(-4.75)
$\ln c_t$	β_1	0.722	(10.18)
$\ln w_t$	β_2	0.220	(2.65)
$\ln \hat{\alpha}_t$	β_3	0.308	(5.25)
$\Delta\ln(P_{wt}^e)$	β_4	0.607	(8.22)
constant	γ	0.1166	(4.83)
	σ_u	0.0205	(4.21)
	σ_v	0.0427	(2.47)

log likelihood = 4.20280
n = 138

The results suggest that the size of the cost of changing price is equivalent to an 11.66% rise in the desired price [27]. Such a size, is again far larger than can reasonably be accounted for by the "menu costs" emphasized in the recent literature.

4.6 Conclusions and Remarks

Unlike most empirical work on mark-up pricing, our estimation is based on disaggregate data. More importantly, because of the stronger theoretical foundation, we have shown that the mark-up equation will only hold for those observations in regime 1, and the planned profit margin was found to be an increasing function of expected demand. The latter result is an important challenge to the work of Godley and Nordhaus (1972).

The estimated cost of changing price was found to be at least equivalent to a

[27]: As the t-statistic of γ_1 of table 4.2(b) still exceeds the 90% significance level. The result reported in table 4.2(b) is preferred.

5.31% change in price (and possibly as high as 10,14% for an expected inflation rate of 10%). Since this implies a very large loss in discounted profits, it would seem unreasonable to assume that price stickiness is solely a result of menu cost.

Chapter 5

5.1 Introduction

In section 1.2.4, we briefly reviewed the following striking result by Caplin and Spulber (1987):

> If (i) monetary growth is monotonic; and (ii) the initial distribution of price deviations across firms is uniform, price stickiness will disappear on aggregation across firms. Thus, money will be neutral despite the presence of fixed cost of changing price and individual price stickiness.

As an attempt to justify their assumption on the initial distribution of price deviations across firms, Caplin and Spulber also showed that the distribution will survive with the specific monetary shocks they considered. Nevertheless, showing the survival of the distribution still does not explain where this initial distribution comes from. For example, if all price deviations in the Caplin and Spulber model are bunched to start with, they will remain bunched thereafter. In such case, price stickiness will not disappear on aggregation across firms, and Caplin and Spulber's explanation[1] of monetary neutrality will not apply. To extend the applicability of Caplin and Spulber's propositions to cases where the initial distribution of price deviations is not necessarily uniform, Blanchard and Fischer (1989) suggests that the addition of idiosyncratic shocks in the model will guarantee a uniform distribution of price deviations in the steady state. The idea was borrowed from Caplin (1985) which discusses the inventory decision, but the logic is similar:

> *"if firms face both idiosyncratic and aggregate shocks and use one sided (s,S) rules, price deviations will be independent across firms, even if the variance in the idiosyncratic shock is arbitrarily small (but not zero):*

---

[1]: Note, we only say that Caplin & Spulber's explanation does not apply here. There can still be some other reasons that might preserve the neutrality result. We will return to this point later.

knowing the price deviation of one firm will be of no help in predicting the price deviation of another. This is an important result as it implies that under one sided (s,S) rules, staggering (uniform distribution) is a natural outcome ."[Blanchard and Fischer (1989]

[Note, however, in stating the above result, Blanchard and Fischer have neglected an important assumption in Caplin(1985):

"demands are received in a well-defined order so that inventories cannot fall from s_k+1 to s_k-1 without passing through level s_k." [Caplin (1985)]

In analyzing the price decision, there may exist occasional and large common shocks (such as that in the oil crisis) whose occurrence will cause all the prices to be bunched at the optimal level. We will return to this point later.]

Despite the above extension, there were doubts in regards to the robustness of the Caplin and Spulber's result for more general monetary generation processes. The first one is the counter example given by Blanchard and Fischer (1989), where the steady state distribution is likely to have higher density at the return point. If their guess is right, one percentage change in money supply will cause less than one percentage change in price, then money will be non-neutral. The second doubt on the robustness of Caplin and Spulber's result is that an ergodic empirical distribution of price deviations may not exist for more general specifications. Bertola and Capallero (1990) have shown that this will happen whenever

(i) there exists ongoing large aggregate shocks

" In the aftermath of such a large shock, idiosyncratic shifts would spread the spike (of price deviations) and the cross section would tend towards the uniform, steady state distribution – but further large shocks would undo the gains in that direction and rebunch some agents anew. In this situation, there would be a continuous tension between the endogenous tendency towards uniformity ... and relatively infrequent structural

changes that prevent the cross section from reaching a steady state."

[Bertola and Caballero (1990)] ; or

(ii) there exists exogenous events that can make adjustments in both directions desirable. If ergodic distribution does not exist, it follows that the empirical distribution will not be uniform at most of the time, and hence the Caplin and Spulber's argument of neutrality will fail to apply in these cases.

It must, however, be emphasized that Caplin and Spulber's result (i.e. disappearance of price stickiness on aggregation across firms) is concerned about whether money is always neutral. Even without such kind of neutrality, money may still be "on average" neutral over time – with which the authorities cannot exploit unless they know the distribution of price deviations [2,3]. Therefore, from the policy point of view, we will be more interested in checking whether money or some particular types of monetary change (such as occasional reduction of money supply in an inflationary world) is neutral on average instead of checking whether Caplin and Spulber's result is robust in more general settings.

In what follows, we will first develop a small macro model suitable for stochastic simulation. The set of equations describing producer behaviour will be based on our discussion in Chapter 2. For a complete model, we need another set of equations

---

2: As an ergodic distribution does not appear to exist in the world, such an assumption seems to be quite reasonable. Indeed, if we ask whether the U.S. or U.K. Government know the empirical distribution of price deviations, the answer is likely to be no.

3: If the empirical distribution has a lower than average density at the margin of raising price, a small rise in money supply will raise the aggregate demand. On the other hand, if the distribution has a higher than average density at the margin of raising price, a small rise in money supply will cause a large reduction in aggregate demand (so that money is on average neutral). However, as the Government does not have perfect information about the empirical distribution, she would not be sure whether a rise in money supply will cause a rise or fall in aggregate demand.

describing consumers. These are borrowed from Blanchard and Kiyotaki (1987) [4]. Thus, the small macro model originates from utility and profit maximizing behaviour. In deriving the macro model, effort is also made to allow comparison and linkage with the model in Caplin and Spulber. Idiosyncratic technological shocks are then introduced to generate the dispersion of price deviations endogenously [5]. To concentrate on the effect of price stickiness, we assume that factor prices adjust instantaneously to clear the market [6]. Finally, we do various simulations with the model and check the correlation coefficient between changes in the money supply and changes in aggregate output. If the correlation coefficient is close to zero, the neutrality hypothesis is supported. Otherwise, the extension of Caplin and Spulber's result to the more general specifications is rejected.

4: It is also possible to use other sets of equations on consumer decisions. However, those in Blanchard and Kiyotaki are selected into the model because money will then be neutral whenever (i) the cost of changing price is zero; or (ii) monetary growth is monotonic. In other words, if non-neutrality result is ever obtained in the subsequent simulation, it must be due to the cost of changing price and the alternative monetary generation process.

5: Idiosyncratic technological shocks are use because it is the simplest way to produce the dispersion of price deviations in the subsequent model. It is true that many other idiosyncratic shocks (eg taste shocks) can generate the dispersion of price deviations, but the way they enter the model will be complicated.

6: As pointed out by Blanchard (1983), wage stickiness may be another reason for the non-neutrality of money. To make sure the non-neutrality result obtained is not due to wage strictness, we make the simplifying assumption that wage stickiness does not exist. If Caplin and Spulber's proposition is rejected, we can conclude that price stickiness alone can generate the non-neutrality of money. Of course, even if Caplin & Spulber's result is supported by the simulation result, it may still be true that wage stickiness can be another reason for the non-neutrality of money.

5.2 Derivation of the Model

5.2.1 The Demand for Goods and the Supply of Labour

Following Blanchard and Kiyotaki (1987), we assume that there are l households, n goods and real balances enter the utility function to avoid Say's Law [7]. Thus, the utility function of household j at any period t is specified as:

$$U_{jt} = C_{jt}^{\gamma} \left(\frac{M_{jt}}{P_{wt}}\right)^{1-\gamma} - N_{jt}^{\beta} \qquad j = 1,\ldots,l$$

$$\text{with} \quad C_{jt} = \left[n^{-1/\theta} \sum_{k=1}^{n} (C_{kjt})^{(\theta-1)/\theta} \right]^{\theta\gamma/(\theta-1)} \qquad (5.1)$$

$$\text{and} \quad P_{wt} = \left[\frac{1}{n} \sum_{k=1}^{n} (P_{kt})^{1-\theta} \right]^{1/(1-\theta)} \qquad (5.2)$$

where (i) C_{kjt} is the consumption of goods k by household j at time t;

(ii) C_{jt} is the index of household j's consumption basket which is assumed to be a CES function of C_{kjt};

(iii) P_{kt} is the price of goods k at time t;

(iv) P_{wt} is the aggregate price index;

(v) N_{jt} is the labour supply by household j at t;

(vi) M_{jt} is the desired holding of money by household j at t;

(vii) θ is the elasticity of substitution between goods in the utility function;

---

[7]: As pointed out by Blanchard and Kiyotaki(1987):
" A Clower constraint would lead to similar results. Developing an explicit intertemporal model to justify why money is positively valued did not seem to worth the additional complexity in this context."

(viii) γ is a parameter showing household's preference between goods and real balance; and

(ix) β is a parameter related to the marginal disutility of labour.

We assume each household takes P_{wt} and W_t as given. We also assume households maximize utility subjected to a budget constraint. By assuming that labour supply is homogeneous, we specify the budget constraint of household j as

$$\sum_{k=1}^{n} P_{kt}\, C_{kjt} + M_{jt} = W_t\, N_{jt} + \sum_{k=1}^{n} \pi_{kjt} + \bar{M}_{jt}$$

where $\bar{M}_{jt}$ is the initial endowment of money held by household j at the beginning of period t;

π_{kjt} is the share of profit from firm k to household j at t;

W_t is the wage rate at t.

Thus, household j's maximization problem at each time period t can be written as

$$\underset{\{C_{kjt},\, M_{jt},\, N_{jt}\}}{\text{Max.}}\ L = C_{jt}^{\gamma}\left(\frac{M_{jt}}{P_{wt}}\right)^{1-\gamma} - N_{jt}^{\beta} - \mu\left[M_{jt} + \sum_{k=1}^{n} P_{kt}\, C_{kjt} - W_t\, N_{jt} - \sum_{k=1}^{n} \pi_{kjt} - \bar{M}_{jt}\right]$$

where C_{jt} and P_{wt} are as defined in equations (5.1) and (5.2).

The first order condition for the maximization problem implies that

$$\gamma C_{jt}^{\gamma-(\theta-1)/\theta}\, n^{-1/\theta}\, C_{kjt}^{-1/\theta}\left(\frac{M_{jt}}{P_{wt}}\right)^{1-\gamma} = \mu P_{kt} \qquad (5.3)$$

$$C_{jt}^{\gamma}\,(1-\gamma)\left(\frac{M_{jt}}{P_{wt}}\right)^{-\gamma} = \mu P_{wt} \qquad (5.4)$$

$$\beta N_{jt}^{\beta-1} = \mu W_t \tag{5.5}$$

After some simplification, it can be shown that the demand for goods k and the supply of labour by household j at time t are given by

$$C_{kjt} = \frac{C_{jt}}{n} \left(\frac{P_{kt}}{P_{wt}} \right)^{-\theta}$$

$$N_{jt} = \frac{1}{\beta} C_{jt}^{\gamma} (1 - \gamma) \left(\frac{M_{jt}}{P_{wt}} \right)^{-\gamma} \left(\frac{W_t}{P_{wt}} \right)$$

Taking the summation over j and defining the aggregate output index Y_t as $Y_t = \sum_{k=1}^{n} (P_{kt} C_{kjt})/P_{wt}$, we have the following demand for goods k and total supply of labour:

$$C_{kt} = \frac{1}{n} (Y_t) \left(\frac{P_{kt}}{P_{wt}} \right)^{-\theta} \quad \text{where } Y_t = \left(\frac{\gamma}{1 - \gamma} \right) \left(\frac{M_t}{P_{wt}} \right) \tag{5.6}$$

$$N_t = \left[\frac{1}{\beta} \gamma^{\gamma} (1-\gamma)^{1-\gamma} \left(\frac{W_t}{P_{wt}} \right) \right]^{1/(\beta-1)} \tag{5.7}$$

We also assume that $\gamma \geq 0.5$ so that a shift of preference from money demand to consumption of goods (i.e. reduction in γ) will cause a rise or at least no change in labour supply.

5.2.2 The Price Equation and Demand for Labour

On the production side, we assume there are n firms, each producing one product. For simplicity, labour is assumed to be the only factor of production and efficiency units of labour are homogeneous. Each firm requires δ_{kt} unit(s) of labour to produce one unit of output. Thus, the demand for labour and the cost function of the firms are given by

$$N_{kt} = \delta_{kt} Y_{kt} \qquad \forall\ k,t \tag{5.8}$$

$$TC_{kt} = W_t N_{kt} \qquad \forall\ k,t \tag{5.9}$$

where N_{kt} is the demand for labour by firm k at time t;

Y_{kt} is the output of firm k at time t;

TC_{kt} is the total cost of firm k at time t; and

δ_{kt} is the technological coefficient of firm k at time t.

Combining equations (5.8) and (5.9). We have

$$TC_{kt} = (W_t \delta_{kt}) Y_{kt} \tag{5.10}$$

We also assume that there is no inventory and a high cost of unsatisfied demand. Such assumptions imply output is demand determined:

$$Y_{kt} = C_{kt}$$

Substituting equation (5.6) into the above equation implies

$$Y_{kt} = \frac{1}{n}\left(\frac{\gamma}{1-\gamma}\right)\left(\frac{M_t}{P_{wt}}\right)\left(\frac{P_{kt}}{P_{wt}}\right)^{-\theta} \tag{5.11}$$

We also assume each firm takes P_{wt}, W_t and aggregate demand as given. Thus, the producer's problem is to choose P_{kt} to maximize the expected sum of discounted profit over the future. Using our discussion in Chapter 2 for the three regimes case where the cost of raising price is assumed to be the same as the cost of reducing price, we have

$$f_{kt}(.) = \max \begin{cases} \sup\limits_{P_{kt} \neq \bar{P}_{kt}} \dfrac{P_{kt} Y_{kt} - TC_{kt}}{1-\rho} - A_{kt} \\ \dfrac{\bar{P}_{kt} \bar{Y}_{kt} - \overline{TC}_{kt}}{1-\rho} \end{cases}$$

where $\bar{Y}_{kt}$, $\overline{TC}_{kt}$ are the demand and total cost at the price $\bar{P}_{kt}$; and

A_{kt} is the cost of changing price.

Substituting equation (5.10) and (5.11) into the above equation implies

$$f_{kt}(.) = \max \begin{cases} \sup\limits_{P_{kt} \neq \bar{P}_{kt}} \dfrac{(P_{kt} - \delta_{kt} W_{kt}) \frac{1}{n} (\frac{\gamma}{1-\gamma}) (\frac{M_t}{P_{wt}}) (\frac{P_{kt}}{P_{wt}})^{-\theta}}{1-\rho} - A_{kt} \\ \dfrac{(\bar{P}_{kt} - \delta_{kt} W_{kt}) \frac{1}{n} (\frac{\gamma}{1-\gamma}) (\frac{M_t}{P_{wt}}) (\frac{\bar{P}_{kt}}{P_{wt}})^{-\theta}}{1-\rho} \end{cases}$$

Following the solution procedure described in Chapter 4, the following price equation is obtained

$$P_{kt} = \max \begin{cases} P_{kt}^{*} = \dfrac{\theta}{\theta - 1} \delta_{kt} W_{kt} & \text{if } \ln P_{kt}^{*} - \ln \bar{P}_{kt} > \psi_k \text{ or } \ln P_{kt}^{*} - \ln \bar{P}_{kt} < -\psi_k \\ \bar{P}_{kt} & \text{otherwise} \end{cases}$$

This equation is, however, only correct for the case of no expected inflation. For the case with constant expected inflation, the above equation should be replaced by [8]

[8]: For simplicity, we will assume all producers have the same cost of changing price so that $\psi'_k = \psi_o + \psi_1 g$, where $\psi_1 g$, reflect the effect of inflation on the cost of changing price. However, as g is assumed to be constant in the simulation exercise, we can simply write $\psi'_k = \psi_o + \psi_1 g = \psi$.

$$P_{kt} = \max \begin{cases} P_{kt}^{*} = \dfrac{\theta}{\theta - 1}\,\delta_{kt}\, W_t\,(1+\tau g) & \text{if } \ln P_{kt}^{*} - \ln \bar{P}_{kt} > \psi_k \text{ or } \ln P_{kt}^{*} - \ln \bar{P}_{kt} < -\psi_k \\ \bar{P}_{kt} & \text{otherwise} \end{cases} \quad (5.12)$$

where $(1+\tau g)$ is used to capture the preadjustment of price arising from the constant expected inflation g;

P_{kt}^{*} is now defined as the price chosen whenever the condition of changing price is satisfied.

Finally, equation (5.8) and (5.11) implies the demand for labour by firm k can be written as:

$$N_{kt} = \delta_{kt}\frac{1}{n}\left(\frac{\gamma}{1-\gamma}\right)\left(\frac{M_t}{P_{wt}}\right)\left(\frac{P_{kt}}{P_{wt}}\right)^{-\theta}$$

and the total demand for labour is given by:

$$N_t = \sum_{k=1}^{n} N_{kt} = \frac{1}{n}\left(\frac{\gamma}{1-\gamma}\right)\left(\frac{M_t}{P_{wt}}\right)\left[\sum_{k=1}^{n}\delta_{kt}\left(\frac{P_{kt}}{P_{wt}}\right)^{-\theta}\right] \quad (5.13)$$

5.2.3. Labour Market Equilibrium

As our aim is to check the effect of price stickiness on the neutrality of money, we assume that wage will adjust instantaneously to clear the labour market. Equating equation(5.7) and (5.13), we have the following wage equation:

$$\frac{W_t}{P_{wt}} = \beta\gamma^{-1}\left(\frac{\gamma}{1-\gamma}\right)^{\beta-\gamma}\left(\frac{M_t}{P_{wt}}\right)^{\beta-1}\left[\sum_{k=1}^{n}\delta_{kt}\left(\frac{P_{kt}}{P_{wt}}\right)^{-\theta}\right]^{\beta-1} \quad (5.14)$$

5.2.4 Aggregate Output

Aggregate output Y_t is defined by $Y_t \equiv \sum_{k=1}^{n} (P_{kt} C_{kjt})/P_{wt}$. Substituting equations (5.2) and (5.11) into the identity, we have

$$Y_t = \left(\frac{\gamma}{1-\gamma}\right)\left(\frac{M_t}{P_{wt}}\right) \tag{5.15}$$

Thus, given the assumptions made above, equation (5.15) implies that aggregate output will be determined by aggregate demand which is proportional to aggregate real balances. The equation also relates the sluggishness of price to the neutrality of money: if the change in the aggregate price index does not lag behind the change in money supply, money will be neutral; otherwise, money will be non-neutral. The equation also implies a shift in preferences from consumption to money demand (i.e. a fall in γ) will cause a reduction in aggregate demand and hence a "long-run" [9] reduction in aggregate output.

5.2.5 The Whole Model

Combining equations (5.2), (5.11), (5.12), (5.14) and (5.15), we now have the small macro model for the simulations:

$$Y_{kt} = \frac{1}{n}\left(\frac{\gamma}{1-\gamma}\right)\left(\frac{M_t}{P_{wt}}\right)\left(\frac{P_{kt}}{P_{wt}}\right)^{-\theta} \tag{5.16}$$

$$P_{kt} = \max \begin{cases} P_{kt}^{*} = \dfrac{\theta}{\theta - 1}\, \delta_{kt}\, W_t\, (1+\tau g) & \text{if } \ln P_{kt}^{*} - \ln \bar{P}_{kt} > \psi_k \text{ or } \ln P_{kt}^{*} - \ln \bar{P}_{kt} < -\psi_k \\ \bar{P}_{kt} & \text{otherwise} \end{cases} \tag{5.17}$$

9: "Long run" here is defined as the state where P_{kt} is adjusted to P_{kt}^{*}.

$$\frac{W_t}{P_{wt}} = \beta\gamma^{-1}\left(\frac{\gamma}{1-\gamma}\right)^{\beta-\gamma}\left(\frac{M_t}{P_{wt}}\right)^{\beta-1}\left[\sum_{k=1}^{n}\delta_{kt}\left(\frac{P_{kt}}{P_{wt}}\right)^{-\theta}\right]^{\beta-1} \quad (5.18)$$

$$P_{wt} = \left[\frac{1}{n}\sum_{k=1}^{n}(P_{kt})^{1-\theta}\right]^{1/(1-\theta)} \quad (5.19)$$

$$Y_t = \left(\frac{\gamma}{1-\gamma}\right)\left(\frac{M_t}{P_{wt}}\right) \quad (5.20)$$

The model will be similar to that of Caplin and Spulber if we assume (i) M_t is monotonically and continuously growing; (ii) δ_{kt} = constant $\forall k,t$; and (iii) the initial values of price deviations ($\ln P_{k0}^* - \ln \bar{P}_{k0}$) are uniformly distributed between the thresholds 0 and $-\psi$. However, to enable the dispersion of price deviations ($\ln P_{kt}^* - \ln \bar{P}_{kt}$) [10] to be generated endogenously, we differ from Caplin and Spulber by assuming the following generation process of technological shock [11]:

$$\ln \delta_{kt} = \ln \delta_{k,t-1} + \varepsilon_{kt} \quad (5.21)$$

where ε_{kt} is an intertemporally and cross-sectionally independent random variable. As Caplin (1985), and Blanchard and Fischer (1989) have suggested, the existence of an arbitrarily small idiosyncratic shock will, in their case of continuous monotonic monetary

10: With the presence of idiosyncratic shocks, it would be more appropriate to analyze in terms of the dispersion of price deviations $\ln (P_{kt}^*/\bar{P}_{kt})$ instead of the dispersion of relative price $\ln (P_{wt}/\bar{P}_{kt})$[c.f. that in Blanchard & Fischer (1989)].

11: A more convincing specification is to allow a drift parameter μ in the generation process (i.e. $\ln \delta_{kt} = \ln \delta_{k,t-1} - \mu + \varepsilon_{kt}$). Nevertheless, such assumption will, in the absence of a sufficient monetary growth, cause a continuously declining P_{kt}^*. To avoid further complications that would be arise from the drift parameter, we will rather work with the less realistic specification in equation(5.21) growth, be sufficient to ensure a uniform distribution at the steady state, we will follow their suggestion by assuming the following specific generation process of ε_{kt}:

$$\varepsilon_{kt} = \begin{cases} -\lambda & \text{with probability} = 5\% \\ 0 & \text{with probability} = 90\% \\ \lambda & \text{with probability} = 5\% \end{cases}$$

where λ is assumed to take a very small value (i.e. 0.0005)

Finally, to check the robustness of Caplin and Spulber's result, we will consider a few alternative stochastic processes for the money supply:

(a) symmetric random walk

example: $m_t = m_{t-1} + \omega_t$ with $\omega_t = \begin{cases} -\chi_\omega & \text{probability} = 0.5 \\ \chi_\omega & \text{probability} = 0.5 \end{cases}$

(b) monotonic monetary growth

example: $m_t = m_{t-1} + g + u_t$ with $u_t = \begin{cases} -\chi_u & \text{probability} = 0.5 \\ \chi_u & \text{probability} = 0.5 \end{cases}$ and $g \geq \chi_u$

(c) non-monotonic monetary growth

example: $m_t = m_{t-1} + g + v_t$ with $v_t = \begin{cases} -\chi_v & \text{probability} = 0.5 \\ \chi_v & \text{probability} = 0.5 \end{cases}$ and $g < \chi_v$

(d) occasional large shocks to monetary growth

example: $m_t = m_{t-1} + g + u_t + Z_t \eta_t$

with $u_t = \begin{cases} -\chi_u & \text{probability} = 0.5 \\ \chi_u & \text{probability} = 0.5 \end{cases}$ and $\eta_t \sim F(0,\psi)$

where (i) $m_t = \ln M_t$ so that g is the average growth rate of money supply for case (b) and (c);

(ii) $F(0,\psi)$ is a uniform distribution between 0 and ψ; and

(iii) Z_t is an indicator function with

$$\begin{cases} \text{Prob. } (Z_t = 1) = q & \text{(say, 0.01)} \\ \text{Prob. } (Z_t = 0) = 1 - q \end{cases}$$

5.3 A statistic for the checking of neutrality of money

To check the neutrality of money in the simulation exercise, it is found more convenient to set up a suitable criterion statistic first. This can be done with the help of equation (5.20) which implies

$$\frac{Y_t}{Y_{t-1}} = \frac{M_t / M_{t-1}}{P_{wt} / P_{w,t-1}}$$

for a constant γ. Taking the logarithm of the above equation we have

$$y_t - y_{t-1} = (m_t - m_{t-1}) - (p_{wt} - p_{w,t-1})$$

where y_t, m_t and p_{wt} are respectively the logarithm of Y_t, M_t and P_{wt}. Writing the equation in difference form, we have

$$\Delta y_t = \Delta m_t - \Delta p_{wt} \qquad (5.2.3a)$$

and $$\overline{\Delta y} = \overline{\Delta m} - \overline{\Delta p_w} \qquad (5.2.3b)$$

Consider the null hypothesis in Caplin and Spulber where price stickiness disappears on aggregation [12] so that the percentage change of M_t always equals the percentage change of P_{wt} (i.e. Δm_t - Δp_{wt}). According to equations (5.2.3a) and (5.2.3b), we have $\Delta y_t = \overline{\Delta y} = 0 \ \forall$ t, which implies that

$$\text{Cov}\ (\Delta m_t, \Delta y_t) = \frac{1}{T} \sum_{t=1}^{T} (\Delta m_t - \overline{\Delta m})\ (\Delta y_t - \overline{\Delta y}) = 0 \qquad (5.2.4)$$

and thus money is definitely neutral.

12: Even if price stickiness does not disappear on aggregation across firms, we might still have the neutrality result as long as the effect of money on output is "on the average" zero over time.

Next consider the case where price stickiness does not always disappear on aggregation across firms. This implies

$$\Delta m_t \neq \Delta p_{wt} \qquad \text{for some t}$$

and hence $\quad y_t \neq y_{t-1} \qquad$ for some t

In this case, money can still on average be neutral or non-neutral, depending on whether the mean of Δp_{wt} is equal to that of Δm_t or not. If it does, equation (5.23b) implies that Δy_t will have a mean zero. Thus, y_t will have a steady means even if money is growing continuously. In other words, money is still on average neutral and the covariance calculated according to formula (5.24) will have an expected value of zero [13]. On the other hand, if the mean of Δp_{wt} does not equal to that of Δm_t, money will be non-neutral and the covariance calculated according to formula (5.24) will be non-zero.

So far, we have been relying on the covariance defined in formula (5.24) to check the neutrality of money. The trouble with this is that the magnitude of the covariance will depend on the choice of the length of each time period, the growth rate of money supply and the size of parameters assumed in the model (5.16)-(5.20). To allow comparison for various specifications, we find it more convenient to rely on the correlation coefficient defined as follows:

$$\rho = \frac{\sum_{t=1}^{T} (\Delta m_t - \overline{\Delta m})(\Delta y_t - \overline{\Delta y})}{\left\{ \left[\sum_{t=1}^{T} (\Delta m_t - \overline{\Delta m})^2\right] \left[\sum_{t=1}^{T} (\Delta y_t - \overline{\Delta y})^2\right] \right\}^{1/2}} \qquad (5.25)$$

[Note: In the Caplin and Spulber's case where $y_t = y_{t-1}$ (or $\Delta y_t = \overline{\Delta y}$) $\forall t$, ρ will

[13]: Of course, in case of finite sample, the actual value may be slightly different from zero. Such deviation will however approach zero as T approaches infinity.

be taken as zero by convention.]

In general, ρ will be approximately zero if money is on average neutral. On the other hand, a significantly non-zero correlation coefficient will support the non-neutrality hypothesis. In the extreme case where price is perfectly sticky (i.e. $p_{wt} = p_{w,t-1}$ or $\Delta p_{wt} = 0\ \forall t$), equation (5.23a) and (5.23b) would imply

$$\Delta y_t = \Delta m_t \qquad \forall t$$

and

$$\overline{\Delta y} = \overline{\Delta m}$$

Substituting these into equation (5.25) would imply $\rho = 1$.

5.4. The Simulation Procedures and Results

5.4.1 The Procedures

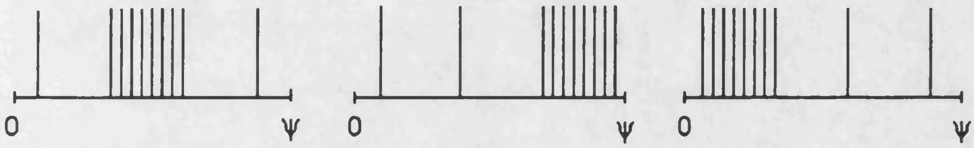
In the subsequent series of simulations, we will assume the following parameter values: $\psi = 0.1$; $\theta = 2.0$; $\beta = 2.0$; $\gamma = 0.75$; $n = 100$ [14]; $\chi_u = 0.0005$; $\chi_v = 0.0015$; $g = 0.0005$; $M_0 = 1.0$; and $\delta_{k0} = 1.0$.

With an appropriate assumption of the initial distribution of price deviations, we then (i) generate ε_{kt} and u_t; and (ii) calculate $\overline{P}_{kt}$, P_{kt}, y_{kt}, M_t, P_t and P_{wt} for t = 1 to 500,000. In those cases where an ergodic distribution exists, such a large number of iterations should be sufficient to ensure the steady state is reached. In many cases, however, we might have a steady set of asymptotically recurrent states instead of a single steady state. One simple example is where (i) producers are identical (ie no idiosyncratic shocks) so that the initial distribution of price deviations is bunched at some point; and

14: For the Blanchard and Fischer case, we will assume n = 200.

(ii) monetary growth is monotonic. In such case, the (degenerated) distribution of price deviations will repeat once whenever the cumulated rise in money supply has reached a multiple of the cost of changing price. For the case where there are idiosyncratic shocks, the reasoning will be slightly more complicated. To see this, we first noted that equation (5.17) implies that ($\ln P_{kt}^* - \ln \overline{P}_{kt}$) will lie between $-\psi$ and ψ. Indeed, if monetary growth is monotonic and the magnitude of ε_{kt} is of second order, we will very often [15] have most of the ($\ln P_{kt}^* - \ln \overline{P}_{kt}$) lying between 0 and ψ.

Thus, even if the dispersion of price deviations at one particular period has higher density at the middle of 0 and ψ, such as that shown in diagram (5.1a),



(a) (b) (c)

where: one vertical line represents one firm.

diagram (5.1)

subsequent rise in money supply will tend to raise $P_{k,t+i}^*$ (but not yet $P_{k,t+i}$) for some small i, making the dispersion of price deviations in the subsequent periods similar to that shown in diagrams (5.1b) and (5.1c). When the cumulated rise of money supply has reached 10% (the cost of changing price), the dispersion of price deviations should

[15] In a few cases where the starting price is close to the return point and there is some "significant" reduction in δ_{kt}, we might have some ($\ln P_{kt}^* - \ln \overline{P}_{kt}$) lying below 0. However, as long as the magnitude of is of ε_{kt} second order, these price deviations will still be very close to 0 within the range $-\psi$ and 0.

"asymptotically" [16] return to that of diagram (5.1a). Having established the steady state or the steady set of "asymptotically" recurrent states with the first 500,000 iterations [17], we repeat the iterations for another 500,000 times and start calculating the correlation coefficient for t = 500,001 to 1,000,000. If the correlation coefficient is close to zero, the neutrality hypothesis is supported. Otherwise, the extension of Caplin and Spulber's result to more general specifications is rejected.

All the procedures outlined above are contained in the computer program reported in the Appendix. In general, the program is rather straight forward. The only exception is due to the existence of two regimes in equation (5.17) which prevent an analytic solution to reduced form equations for P_{kt} and $\mathbf{P}_{wt}$ [18]. The reasoning is as follows. From equations (5.17) and (5.18), we can see that $\mathbf{P}_{wt}$ has to be known before an individual producer can decide whether to raise or maintain his price (i.e $P_{kt} = P_{kt}^*$ or $P_{kt} = \overline{P}_{kt}$). However, according to equation (5.19), $\mathbf{P}_{wt}$ will not be known unless P_{kt} is known before hand. Thus, even though we will know P_{kt} if $\mathbf{P}_{wt}$ is known or vice versa, we have the problem of not knowing P_{kt} or $\mathbf{P}_{wt}$ at the very beginning. Fortunately, we find the above problem can be circumvent by the following iteration procedure which will

---

16There might be a little "realization" difference due to the difference in realized δ_{kt}.

17: In the case where a steady state or a steady set of asymptotically recurrent steady state does not exist, the first 500,000 iterations would still allow the idiosyncratic shocks to reshuffle the distribution so that the correlation coefficient calculated in the next 500,000 iterations will be independent of the initial distribution assumed.

18: Such problem does not exist if there is zero cost of changing price so that there is only one regime in equation (5.17) (i.e. $P_{kt} = P_{kt}^*$). In such case, equation (5.17)-(5.19) can be solved simultaneously to give reduced form equations for P_{kt}, W_t and $\mathbf{P}_{wt}$.

search for P_{kt} and $\mathbf{P_{wt}}$ numerically:

Step 1:

We first set the initial search values $\tilde{\mathbf{P}}_{wt}^0$ and $\tilde{P}_{kt}^0$ to $\mathbf{P_{w,t-1}}$ and $P_{k,t-1}$ respectively.

Then for any iteration i = 1 to 1,000,000, the search values $\tilde{\mathbf{P}}_{wt}^i$ and $\tilde{P}_{kt}^i$ are calculated according to the following formula:

$$\tilde{W}_t^i = \beta\gamma^{-1}\left(\frac{\gamma}{1-\gamma}\right)^{\beta-\gamma}\left(\frac{M_t}{\tilde{\mathbf{P}}_{wt}^{i-1}}\right)^{\beta-1}\left[\sum_{k=1}^{n}\delta_{kt}\left(\frac{\tilde{P}_{kt}^{i-1}}{\tilde{\mathbf{P}}_{wt}^{i-1}}\right)^{-\theta}\right]^{\beta-1}\tilde{\mathbf{P}}_{wt}^{i-1}$$

$$\tilde{P}_{kt}^i = \max\begin{cases}\tilde{P}_{kt}^* = \dfrac{\theta}{\theta-1}\,\delta_{kt}\,\tilde{W}_t^i\,(1+\tau g) & \text{if } \ln\left(\dfrac{\tilde{P}_{kt}^*}{P_{kt}}\right) > \psi \text{ or } \ln\left(\dfrac{\tilde{P}_{kt}^*}{P_{kt}}\right) < -\psi \\ \bar{P}_{kt} & \text{otherwise}\end{cases}$$

$$\tilde{\mathbf{P}}_{wt}^i = \left[\frac{1}{n}\sum_{k=1}^{n}\left(\tilde{P}_{kt}^i\right)^{1-\theta}\right]^{1/(1-\theta)}$$

Step 2:

Then we check whether $\tilde{\mathbf{P}}_{wt}^i = \tilde{\mathbf{P}}_{wt}^{i-1}$. If yes, we have already search out the solution of $\mathbf{P_{wt}}$ and P_{kt} and we can set $\mathbf{P_{wt}} = \tilde{\mathbf{P}}_{wt}^{i-1}$ and $P_{kt} = \tilde{P}_{kt}^i$. If $\tilde{\mathbf{P}}_{wt}^i \neq \tilde{\mathbf{P}}_{wt}^{i-1}$, we proceed the iteration to i + 1 and so on until convergence. Indeed, for the model assumed in our simulation exercise, convergence is usually achieved within 3 iterations.

5.4.2 The simulation Result

(A) No idiosyncratic shocks

Our first simulation exercise is to show the correspondence between the special case of model (5.16)-(5.20) mentioned in section 5.25 and that of Caplin and Spulber. This is done by first examining the specific case where (i) money is growing monotonically [i.e. case(b) in section (5.25)] (ii) there is no idiosyncratic shock (i.e.ln δ_{kt} = 1.0 $\forall$k,t); and (iii) the initial distribution of price deviations is uniformly distributed between 0 and ψ. To allow our discrete time model to have an approximately continuous monetary growth [as required in Caplin and Spulber (1987)], we also define t as a very short period so that g, χ_u, χ_v take correspondingly small values such as those assumed in section 5.4.1. As reported in column 1 of Table 5.1, despite of the continuous and monotonic monetary growth, aggregate output always remains at the same level for all t (i.e. $y_t = y_{t-1}$ or $\Delta y_t = 0 \ \forall t$). As Δy_t is no longer stochastic here, ρ will be taken as zero here. Thus, the Caplin and Spulber's result is preserved here.

Next, let us see what happen if we drop assumption (iii) and assume the initial distribution of price deviations is bunched at 0. As reported in column 2 of Table 5.1, the correlation coefficient is close to zero (i.e. 0.5365 x 10^{-5}) indicating that money is still on average neutral. However, when we plot the distribution of price deviations, we found that – as predicted in section 5.4.1 – it enters a steady set of asymptotically recurrent states in which the bunch of price deviations will keep rising with the monetary growth until they reach the margin of raising price, ψ, where the bunch will jump back to 0 and everything proceed as before [19]. Thus, unlike that expected by Caplin and Spulber(1987)

[19]: Along with the movement of the bunch of price deviations, aggregate output is also rising with the monetary growth until the margin of raising price where aggregate output will jump back to the lower bound and everything will proceed as before.

Table 5.1: The Simulation Result

	Without idiosyncratic shocks (1)			With idiosyncratic shocks (2)			
	monotonic monetary growth (3)		money supply follows symmetric random walk (4)	monotonic monetary growth		money supply follows symmetric random walk	non-montonic monetary growth
	initial distribution of price deviations uniformly distributed between 0 and ψ	initial distribution of price deviations all bunched at the return point 0	initial distribution of price deviatioons all bunched at the return point 0	no large common shocks (3)	with large common shocks (5)	case of Blanchard & Fischer (1989) (6)	effect of reduction in money supply (7)
correlation coefficient	0.0 (8)	0.5366×10^{-5}	1.000	0.0390	−0.0388	0.9505	0.9767 (9)

Note: (1) $\ln \delta_{kt} = 1.0 \quad \forall\ k,t$

(2)

$$\ln \delta_{kt} = \ln \delta_{k,t-1} + \varepsilon_{kt} \qquad \text{where } \varepsilon_{kt} = \begin{cases} -0.0005 & \text{with probability} = 5\% \\ 0 & \text{with probability} = 90\% \\ 0.0005 & \text{with probability} = 5\% \end{cases}$$

As idiosyncratic shocks will tend to reshuffle the distribution, the result obtained in column (4) – (7) will be independent of the initial distribution assumed. However, to check the Blanchard and Fischer case (column 6), we will assume the initial distribution is uniformly distributed between $-\psi$ and ψ, and see whether the distribution after the 500,000 iterations will have a higher density at the return point. For all other cases (particularly in checking the neutrality result of column 4), we will assume the initial distribution has a higher density at the return point (i.e. ≈ triangularly distributed between 0 and ψ).

(3) $m_t = m_{t-1} + g + u_t$ with $u_t = \begin{cases} -0.0005 & \text{probability} = 0.5 \\ 0.0005 & \text{probability} = 0.5 \end{cases}$ and $g = 0.0005$

(4) $m_t = m_{t-1} + s_t$ with $s_t = \begin{cases} -0.00001 & \text{probability} = 0.5 \\ 0.00001 & \text{probability} = 0.5 \end{cases}$

(5) $m_t = m_{t-1} + g_1 + \omega_t + Z_t \eta_t$ where $\omega_t = \begin{cases} -0.0001 & \text{probability} = 0.5 \\ 0.0001 & \text{probability} = 0.5 \end{cases}$; $g_1 = 0.0001$; $\eta_t \sim F(0,\psi)$ and $\begin{cases} \text{Prob}(Z_t=1) = 1\% \\ \text{Prob}(Z_t=0) = 99\% \end{cases}$

(6) $m_t = m_{t-1} + \Omega_t$ where $\Omega_t = \begin{cases} -0.01 & \text{probability} = 0.5 \\ 0.01 & \text{probability} = 0.5 \end{cases}$
If the step size of the random walk is 0.001, the correlation coefficient will be 0.9927.

(7) $m_t = m_{t-1} + g + v_t$ where $v_t = \begin{cases} -0.0015 & \text{probability} = 0.5 \\ 0.0015 & \text{probability} = 0.5 \end{cases}$ and $g = 0.0005$

(8) As $\Delta y_t = 0$ t and $\text{Cov}(\Delta m_t, \Delta y_t) = 0$, the correlation here is set to zero by convention.

(9) The correlation coefficient in column 7 refers only to those observations with a reduction of money supply.

or Blanchard and Fischer (1989), the neutrality is not due to the disappearance of price stickiness on aggregation across firms. Instead, money is "on average" neutral over time (i.e. the cumulated small rise of output is cancelled by the occasionally large reduction in output when the bunch of firms raise their price).

To check the importance of monotonic monetary growth to the "on average" neutrality, we also report – in column 3 – the case where money supply is assumed to follow a symmetric random walk instead of growing monotonically. Such a case has a close resemblance with that presumed in the menu cost hypothesis [such as Akerlof and Yellen(1985a,b) and Mankiw(1985)], which implicitly or explicitly conclude the non-neutrality of money by showing the price stickiness of a representative firm. As expected, the correlation coefficient for the assumed size of random walk is 1.0 [20], indicating that the non-neutrality result presumed in the menu cost hypothesis will be right as long as there is no monetary growth. The result will however be weakened with the growing importance of monetary growth. When the relative importance of monetary growth reaches the point that money supply is growing monotonically, money will – as indicated by the result in column 2 – be on average neutral.

(B) With idiosyncratic technological shock

To generate the distribution of price deviations endogenously, we now introduce the idiosyncratic shock into Caplin and Spulber's variant by assuming that $\varepsilon_{\kappa t}$ is

---

[20]: If we assume the step size of the random walk is larger (say, 0.01 instead of 0.00001), cumulated changes in money supply might occasionally push the bunch of price deviations to the two threshold points where there will be a bunch revision of prices. If this ever happens, the associated correlation coefficient will be less than 1. Nevertheless, as long as the step size is not too large so that the touching of the thresholds remain occasional, the associated correlation coefficient should still be significantly different from zero. This is confirmed by our supplementary exercise which give a correlation coefficient of 0.6679 when the step size of the random walk is 0.01 instead of 0.00001

generated from the process described in section 5.2.5 [21]. Again, the close to zero (i.e. 0.0390) correlation coefficient indicates that money is approximately neutral. This is so because the asymptotic distribution of price deviations is uniform (with occasionally small, but "on average" zero, skewing in the realized distribution) [22].

What happens if there is something that stops the idiosyncratic shocks from reshuffling the distribution of price deviations towards the uniform distribution so that Caplin and Spulber's explanation of neutrality does not apply (i.e. price stickiness does not disappear on aggregation across firms)? Does it – as along the line of thinking of the usual criticism of Caplin and Spulber (1987) [23] – imply that money must be non-neutral? Our answer is that it all depends. Examples of these are reported in column 5 and column 6.

In the case of column 5 [i.e. one of the case in Bertola and Cabellero (1990)], the distribution of price deviations is prevented from converging to the uniform distribution by assuming the existence of ongoing (but occasional) large common shocks whose occurrence will bunch the price deviations to 0. Despite the fact that price stickiness does not necessarily disappear on aggregation across firms, the correlation coefficient reported in column 5 (= – 0.0388) indicates that monotonic monetary change is still "on average" neutral over time. This is so because, although a rise in the money supply at t may cause a change in output whenever the distribution of price deviations is not uniform, the change in the money supply also changes the distribution of price deviations. When the

---

21: The result will still be the same for some more general generation process of ε_{kt}.

22Monetary growth is still assumed to be monotonic and (approximately) continuous. With idiosyncratic shock, it does not matter whether the initial distribution of price deviation is bunched or uniform. The result is reported in Column 4 of Table 5.1.

23: Such as Blanchard & Fischer(1989).

cumulated rise of money supply has reached a multiple of the cost of changing price , the distribution of price deviations (and hence output) will "on average" be the same as those at t. The fact that the expected distribution of price deviations will return to the original one when the cumulated rise of the money supply reaches a multiple of the cost of changing price is important. It implies that money is neutral in the sense that one cannot keep raising output by indefinite increase in the money supply. Sooner or later, as long as money is growing monotonically, prices will be raised so that output can only fluctuate within some bounds allowed by the size of the cost of changing price.

In column 6, we have Blanchard and Fischer's counter example mentioned in section 1.2.4. Unlike the monotonic growth in Caplin and Spulber (1987), money supply here is assumed to follow a symmetric random walk. A plot of the price deviations for the few periods after the first 500,000 iterations suggest that the empirical distribution of price deviations – as expected by Blanchard and Fischer (1989) – has a higher density around the return point. Thus, as similar to column 5, neither do we have an ergodic uniform distribution here[24]. However, unlike that in column 5, the correlation coefficient here (0.9505) suggests that money is non-neutral. Why is there such a difference? This is so because, in the case of column 5, the monotonically growing money supply is "on average" neutral over time (i.e. the growing money supply will ensure that any skewing of the distribution of price deviations towards one threshold will be reversed in some subsequent period(s), thus the rise (fall) in output with the rise (fall) in money supply at t will be cancelled by the fall (rise) in output in some subsequent period(s). In the case of column 6, there is however no trend monetary growth. Instead, money supply follows a symmetric random walk so that the steady state distribution of price deviations has a higher density around the return point. With such kind of

---

[24]: Note also the difference between column 5 and 6: in column 5, no ergodic distribution exists; while in column 6, an ergodic distribution exists although it is not uniform.

distribution, a small rise in the money supply will, of course, cause a less than proportionate rise [25] in the aggregate price index, (and hence a rise in aggregate output), giving the result that the small rise in the money supply is non-neutral. As it is assumed to follow a symmetric random walk, the money supply – unlike the case with monetary growth – may fall back in some future period. Suppose this happens in the next period, the small reduction in the money supply will also cause a less than proportionate reduction in price and hence a reduction in aggregate output, giving the result that a small reduction in money supply is non-neutral.

From the cases in column 5 and column 6, we can see that monotonic monetary growth is an important reason for the "on average" neutrality of money over time. As a cross check of this, we can add an underlying trend growth of money supply to the case of column 6 (i.e. Blanchard and Fischer's counter example). This would imply two different sub-cases. The first is where the trend growth of the money supply exceeds the variations arising from the symmetric random walk. In such a case, monetary growth is monotonic. Indeed, one of these is already analyzed in column 4

i.e. $$m_t = m_{t-1} + g + u_t \quad \text{with } \omega_t = \begin{cases} -\chi_u & \text{probability} = 0.5 \\ \chi_u & \text{probability} = 0.5 \end{cases} \text{ and } g \geq \chi_u$$

Thus, the close to zero correlation coefficient reported in column 4 supports this hypothesis. The second case is where trend growth is less than the variations arising from the symmetric random walk so that monetary growth is non-monotonic. Here, we suggest that any occasional reduction of the money stock in an "inflationary world" will be non-neutral. We can do so by selecting only those observations with a reduction in money stock and calculate the associated correlation coefficient. The result is reported in column 7. The close to unity correlation coefficient (= 0.9767) suggests that a reduction

25: In case the steady state distribution has very high density around the return point, the rise in aggregate price index will be close to zero.

of the money supply in an " inflationary world" will be non-neutral. The reasoning is as follows. With underlying monetary growth, most of the price deviations – after the first 500,000 iterations – will lie between 0 and ψ, with only a few that might lie slightly below 0 (and well above $-\psi$, if we assume the variations in ε_{kt} is of second order). Thus prices will remain unchanged so that aggregate output falls with a small reduction in money supply.

As a supplement, we repeat the above exercise for a "deflationary world" by writing the money supply equation as

$$\text{example: } m_t = m_{t-1} - g + v_t \quad \text{with } \omega_t = \begin{cases} -\chi_v & \text{probability} = 0.5 \\ \chi_v & \text{probability} = 0.5 \end{cases} \text{and } g < \chi_u$$

and calculate the correlation coefficient for only those observations with a rise in money supply. Again, the correlation coefficient is close to unity. Thus, a rise in money supply in a deflationary world will also be non-neutral.

5.5 Conclusions and Remarks

Having gone through the simulation exercises, we summarize our view about the neutrality of money as follows:

(A) Money is neutral in the sense that one cannot keep raising (reducing) aggregate output by indefinite increases (reduction) of the money supply. The most simple case is where (i) all producers are identical so that price deviations are always bunched within - ψ and ψ such as that in diagram (5.2); and (ii) there is no idiosyncratic technological shock.

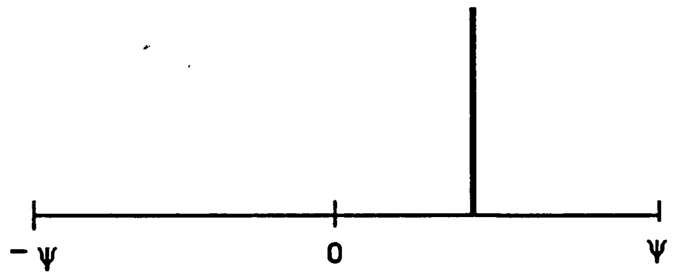


diagram (5.2)

If the bunch of producers are not at the margin of raising price, rise in money supply will, at the very beginning, cause some rise in output. However, the dispersion of price deviations will also shift towards the upper threshold ψ with the rise of money supply [26]. By the time the cumulated rise in the money supply has shifted the bunch of price deviations to ψ, prices will be raised by ψx100%. Thus, the bunch of price deviations jump to 0 and aggregate output falls. With further rises in money supply, prices will remain sticky and aggregate output will rise until the price deviations reach ψ again. Thus aggregate output will be fluctuating between a lower bound Y_L and an upper bound Y_H such as that shown in diagram (5.3).

---

[26]: This is so because a rise in money supply will cause a rise in P_{kt}^* and hence a rise of $(\ln P_{kt}^* - \ln \bar{P}_{kt})$

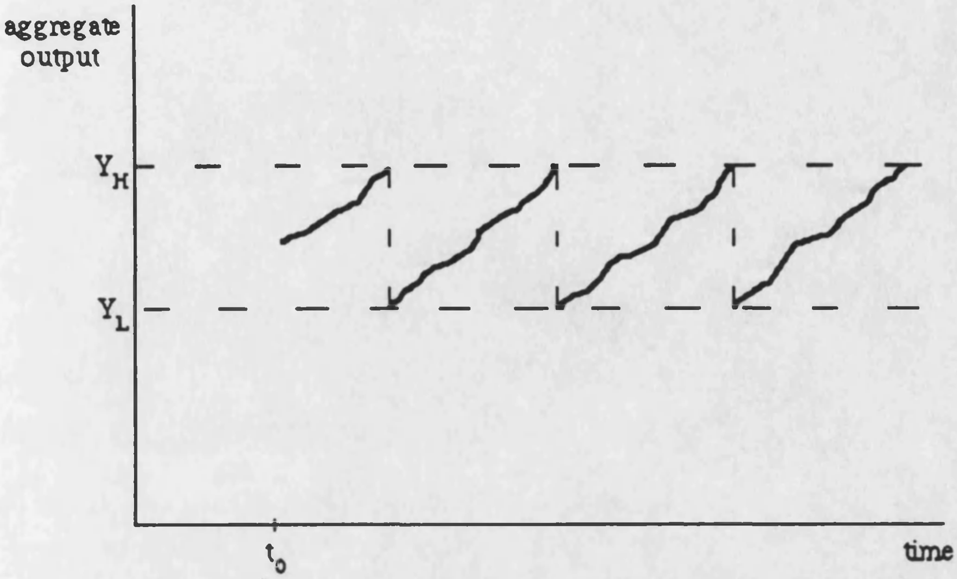


diagram (5.3)

For the case where price deviations are somewhat more dispersed, the logic will be similar and the only difference is a smaller gap between Y_L and Y_H . In the limiting case where the price deviations are evenly distributed between 1 and $1/(1-\psi)$, Y_L and Y_H will be the same and output will always stay at that level. Introduction of idiosyncratic technological shocks does not change the conclusion as long as money is growing. This is so because, although the rise of the money supply might cause a rise in output at this moment, it also shifts the dispersion of price deviations towards the upper threshold, which will soon reverse the rise in output with further monetary increases. Indeed, as discussed in section 5.4.1, as long as money is growing (which is a reasonable assumption), the dispersion of price deviations will enter a steady set of asymptotically recurrent states. As a result, aggregate output will still fluctuate between two bounds [27] and the continuous rise in the money supply cannot

[27]: Or more correctly, two variable bounds which will rise or fall, depending on the realized δ_{kt}. Nevertheless, such variations in the bounds are small because of the assumption of small variations in δ_{kt}.

cause a continuous rise in aggregate output.

(B) Money is non-neutral in the sense that a moderate reduction in the money supply in an inflationary world will cause reduction in aggregate output. To see this, assume that the money stock has been rising before t+1 and then comes a reduction in money supply. As explained in the simulation exercise, most of the price deviations ($\ln P_{kt}^* - \ln \bar{P}_{kt}$) will lie between 0 and ψ, with at most a few slightly below 0. This implies that most $\ln \bar{P}_{kt}$ will lie between ($\ln P_{kt}^* - \psi$) and $\ln P_{kt}^*$, with only a few slightly above $\ln P_{kt}^*$:

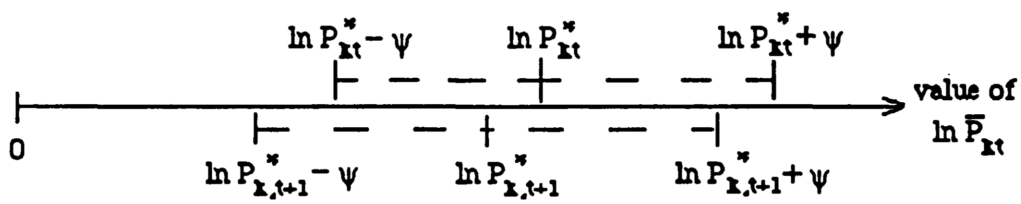


diagram (5.4)

The moderate [28] reduction of the money supply at t+1 will cause a moderate reduction in thresholds such as that shown in diagram (5.4). As all $\ln \bar{P}_{k,t+1}$ ($=\ln P_{kt}$) still lie within the new thresholds ($\ln P_{k,t+1}^* - \psi$) and ($\ln P_{k,t+1}^* + \psi$), all producers will keep their price unchanged. According to equation (5.20), this implies aggregate output will fall with the moderate reduction in money supply.

---

[28]: In case of large reduction in money supply, producers may reduce their price and aggregate output will only fall by a small amount. Such reduction of output arises because most producers have the starting price $\bar{P}_{k,t+1}$ less than P_{kt}^* which implies aggregate price index will fall by a smaller percentage than the reduction in money supply.

(C) The reduction of aggregate output mentioned in (B) will vanish with any subsequent rise of the money stock to the level prevailing at t. After that, everything will be the same as if there were no reduction (and the subsequent recovery) of money supply.

(D) Suppose at t+1, instead of having a reduction in the money supply, we have a moderate reduction in γ (i.e. a shift of preferences from the consumption of goods to money balances which causes a "long run" reduction in the bounds within which output is fluctuating [29]. Although changing the money supply has no effect on the expected values of the bounds of output, it is usually [30] possible to alleviate the initial reduction of output by increasing the money supply so as to push output at t+1 towards the upper threshold. The policy will be particularly useful when the

29: This happens when $\gamma \geq 0.5$ which is what we have assumed. The proof can be done by first finding $\mathbf{P}_{wt}^*$ and Y_t^*, which are respectively defined as the aggregate price and output when $P_{kt} = P_{kt}^*$ $\forall k$, and then showing the thresholds do change with γ. Thus, replacing equation (5.17) by the following equation for P_{kt}^*

$$P_{kt}^* = \frac{\theta}{\theta - 1} \delta_{kt} W_{kt}$$

and solving the small macro model, it can be shown that

$$\mathbf{P}_{wt}^* = K_t^{-1} \left(\frac{\gamma}{1-\gamma}\right)^{(\beta-1)/(\beta-\gamma)} \gamma^{-1/(\beta-1)} M_t$$

and
$$Y_t^* = K_t \, \gamma^{\gamma/(\beta-1)} (1-\gamma)^{(1-\gamma)/(\beta-1)}$$

where
$$K_t = \frac{1}{\beta}\left[\frac{\theta}{\theta-1}(1+\tau g)\right]^{\theta-\beta/(\beta-1)} \left\{\left[\frac{1}{n}\Sigma\left(\frac{\delta_{kt}}{a_{kt}}\right)^{1-\theta}\right]\right\}^{\theta+1/(\beta-1)} \left\{\left[\frac{1}{n}\Sigma\left(\frac{a_{kt}}{\delta_{kt}}\right)^{\theta-1}\right]\right\}^{-1}$$

It can also be shown that Y_t^* will rise and fall with γ when $\gamma \geq 0.5$. As the expected thresholds of $\mathbf{P}_{wt}$ and Y_t will be lying to both sides of $\mathbf{P}_{wt}^*$ and Y_t^*, and the range between these thresholds is independent of γ, we conclude that the thresholds of output will rise and fall with γ, for the case $\gamma \geq 0.5$

30: The only impossible case is when output at t + 1 is at the upper threshold.

reduction in γ or aggregate demand is temporary. Thus, even if monetary policy cannot affect the expected bounds of output or average output in the "long run", it can still be use to fine-tune exogenous variations in aggregate demand.

```
      PROGRAM SIMULATION
      PARAMETER (MI=1,N=100)
      INTEGER I,K,IJ,JJ,KK
      REAL COSTDP1,COSTDP2,THETA,LAMDA,BETA,SEGTA,GAMMA,FI,THESEGFI
     1      ,RMW,RM1(N),G05DAF,G05DDF
      DOUBLE PRECISION E(MI),SLM(MI),SM(MI),WAGE(MI),PSTAR(N,MI),
     1 SSSDM,SSSDY,YBAR,SLMLAG(MI),PBAR(N,MI),COV(MI),P(N,MI),
     2 MPR,PPHAT,PKHAT(N),PKSHAT(N),DPKSHAT(N),PPHATL,SPKHAT,
     3 DPPHAT,THETHE,GAMGAM,SSEGPK,SEGTAK(N),FIIJ,COSTU,COSTL,
     4 NEGTAK(N),YLAG,VARM,VARY,CORRE,SLOPE,COVAR,LRPKSHAT(N),
     5 LPKSHAT(N),SEK(N),SSEK1,SSEK2,RSEK12,EXP04,EXN04,RPKSHAT(N),
     6 SDM(MI),SDY(MI),SDMSQ(MI),SDYSQ(MI),SDMDY(MI),POSTAR,
     7 LPOSTAR,LP(N,MI),LSEK(N),
     8 DPSTAR(N,MI),PI(MI),WPI(MI),Q(N,MI),Y(MI),DM(MI),DY(MI)
C
CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
C     ASSIGN PARAMETER VALUE
CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
C
      THETA=2.0
      COSTDP1=0.1
      LAMDA=0.0005
      BETA=2.0
      DO 368 K=1,N
  368    SEGTAK(K)=EXP(1.0)
      SEGTA=EXP(1.0)
      GAMMA=0.75
C
CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
C     ASSIGN INITIAL VALUES
CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
C
      SLM(1)=0.0
      SM(1)=1.0
      THETHE=THETA/(THETA-1.0)
      GAMGAM=(GAMMA/(1.0-GAMMA))**(BETA-GAMMA)/GAMMA
      COSTU=EXP(COSTDP1)
      COSTL=EXP(-COSTDP1)
      DO 191 K=1,5
  191    SEK(K)=EXP(-0.0*COSTDP1/8.0-((K*1.0-1.0)/5.0)*COSTDP1/8.0)
      DO 192 K=6,15
  192    SEK(K)=EXP(-1.0*COSTDP1/8.0-((K*1.0-6.0)/10.0)*COSTDP1/8.0)
      DO 193 K=16,30
  193    SEK(K)=EXP(-2.0*COSTDP1/8.0-((K*1.0-16.0)/15.0)*COSTDP1/8.0)
      DO 194 K=31,50
  194    SEK(K)=EXP(-3.0*COSTDP1/8.0-((K*1.0-31.0)/20.0)*COSTDP1/8.0)
      DO 195 K=51,70
  195    SEK(K)=EXP(-4.0*COSTDP1/8.0-((K*1.0-51.0)/20.0)*COSTDP1/8.0)
      DO 196 K=71,85
  196    SEK(K)=EXP(-5.0*COSTDP1/8.0-((K*1.0-71.0)/15.0)*COSTDP1/8.0)
      DO 197 K=86,95
  197    SEK(K)=EXP(-6.0*COSTDP1/8.0-((K*1.0-86.0)/10.0)*COSTDP1/8.0)
      DO 198 K=96,100
  198    SEK(K)=EXP(-7.0*COSTDP1/8.0-((K*1.0-96.0)/5.0)*COSTDP1/8.0)
      DO 9877 K=1,N
         SSEK1=SSEK1+SEGTAK(K)*SEK(K)**(-THETA)/(N*1.0)
 9877    SSEK2=SSEK2+SEK(K)**(1.0-THETA)/(N*1.0)
      RSEK12=SSEK1/(SSEK2**(-THETA/(1.0-THETA)))
      FI=BETA*(RSEK12**(BETA-1.0))*GAMGAM
      THESEGFI=THETHE*SEGTA*FI
      POSTAR=THESEGFI*SM(1)*EXP(0.5*COSTDP1)
      DO 95 K=1,N
```

```
   95  P(K,1)=POSTAR*SEK(K)
       PI(1)=0.0
       DO 96 K=1,N
   96     PI(1)=PI(1)+(P(K,1)**(1.0-THETA))/N
       WPI(1)=PI(1)**(1.0/(1.0-THETA))
C
       LPOSTAR=LOG(POSTAR)
       WRITE(3,4560)'LPOSTAR:',LPOSTAR
       WRITE(3,4560)'WPI(1)0:',WPI(1)
 4560  FORMAT(/3X,A8,2G20.8)
       DO 4559 K=1,N
          LP(K,1)=LOG(P(K,1))
 4559     LSEK(K)=LOG(SEK(K))
       WRITE(3,4561)(LP(K,1),K=1,N)
       WRITE(3,4561)(LSEK(K),K=1,N)
 4561  FORMAT(/(10G12.5))
C
C      DO 97 K=1,N
C  97     Q(K,1)=(GAMMA/(1-GAMMA))*(SM(1)/WPI(1))
C    1           /((P(K,1)/WPI(1))**THETA*N)
C      YBAR=0.0
C      DO 98 K=1,N
C  98     YBAR=YBAR+Q(K,1)*P(K,1)/WPI(1)
C
       YBAR=(GAMMA/(1.0-GAMMA))*(SM(1)/WPI(1))
C
       Y(1)=YBAR
CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
C      START THE ITERATIONS
CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
C
       DO 100 J=1,1000000
          YLAG=Y(1)
          SLMLAG(1)=SLM(1)
          DO 369 K=1,N
             RM1(K)=G05DAF(0.0,0.1)
             IF(RM1(K).LT.0.005)NEGTAK(K)=-0.0005
             IF(RM1(K).GT.0.095)NEGTAK(K)=0.0005
             IF(RM1(K).GE.0.005.AND.RM1(K).LE.0.095)NEGTAK(K)=0.0
  369        SEGTAK(K)=SEGTAK(K)*EXP(NEGTAK(K))
          E(1)=G05DAF(-0.0015,0.0015)
          SLM(1)=SLMLAG(1)+0.0005+E(1)
          SM(1)=EXP(SLM(1))
          PPHAT=WPI(1)
          DO 330 K=1,N
  330        PKHAT(K)=P(K,1)
C
          DO 331 IJ=1,10000
             PPHATL=PPHAT
             SSEGPK=0.0
             DO 332 K=1,N
  332           SSEGPK=SSEGPK+SEGTAK(K)*(PKHAT(K)/PPHAT)**(-THETA)
             FIIJ=BETA*(SSEGPK/(N*1.0))**(BETA-1)*GAMGAM
             DO 333 K=1,N
             PKSHAT(K)=THETHE*SEGTAK(K)*FIIJ*SM(1)*EXP(0.5*COSTDP1)
             RPKSHAT(K)=PKSHAT(K)/P(K,1)
                IF(RPKSHAT(K).GE.COSTU.OR.RPKSHAT(K).LE.COSTL)THEN
                   PKHAT(K)=PKSHAT(K)
                ELSE
                   PKHAT(K)=P(K,1)
                ENDIF
  333        CONTINUE
             SPKHAT=0.0
             DO 334 K=1,N
```

```fortran
  334             SPKHAT=SPKHAT+(PKHAT(K)**(1.0-THETA))/(N*1.0)
              PPHAT=SPKHAT**(1.0/(1.0-THETA))
              DPPHAT=(PPHAT-PPHATL)/PPHATL
              IF(DPPHAT.LE.0.00000000)GO TO 9999
              IF(IJ.EQ.10000.AND.DPPHAT.GT.0.00000000)GO TO 999
  331 CONTINUE
 9999 CONTINUE
C
      IF(J.GE.500000.AND.J.LE.500010)THEN
      DO 1332 K=1,N
 1332    LRPKSHAT(K)=LOG(RPKSHAT(K))
      WRITE(3,1333)(LRPKSHAT(K),K=1,N)
 1333 FORMAT(/(10G12.5))
      ENDIF
C
         DO 340 K=1,N
  340       P(K,1)=PKHAT(K)
         WPI(1)=PPHAT
C
C
         Y(1)=(GAMMA/(1.0-GAMMA))*(SM(1)/WPI(1))
         DM(1)=SLM(1)-SLMLAG(1)
         DY(1)=LOG(Y(1))-LOG(YLAG)
         IF(J.GT.500000.AND.SLM(1).LT.SLMLAG(1))THEN
            JJ=JJ+1
            SDMSQ(1)=SDMSQ(1)+DM(1)**2.0
            SDYSQ(1)=SDYSQ(1)+DY(1)**2.0
            SDMDY(1)=SDMDY(1)+DM(1)*DY(1)
            SDM(1)=SDM(1)+DM(1)
            SDY(1)=SDY(1)+DY(1)
         ENDIF
      IF(J.LE.105.OR.J.GT.999890)THEN
      WRITE(3,124)J,SLM(1),SM(1),PPHAT,FIIJ,Y(1),DY(1),
     1            DM(1),E(1),IJ,SEGTAK(1),NEGTAK(1),NEGTAK(2)
  124 FORMAT(/I7,3G12.5,2G12.5,3F8.5,I5,3F9.5)
      ENDIF
  100 CONTINUE
C
C     WRITING OUTPUT
C
      SDY(1)=SDY(1)/(JJ*1.0)
      SDM(1)=SDM(1)/(JJ*1.0)
      COVAR=SDMDY(1)/(JJ*1.0)-SDM(1)*SDY(1)
      WRITE(3,110)'COVAR:',COVAR
      WRITE(3,110)'SDY(1)',SDY(1)
      WRITE(3,110)'SDM(1)',SDM(1)
      SSSDM=SDMSQ(1)/(JJ*1.0)-SDM(1)**2.0
      SSSDY=SDYSQ(1)/(JJ*1.0)-SDY(1)**2.0
      WRITE(3,110)'SSSDY:',SSSDY
      WRITE(3,110)'SSSDM:',SSSDM
      CORRE=COVAR/((SDMSQ(1)/(JJ*1.0)-SDM(1)**2.0)*
     1            (SDYSQ(1)/(JJ*1.0)-SDY(1)**2.0))**0.5
      WRITE(3,110)'CORRE:',CORRE
  110 FORMAT(/X,A6,G20.9)
C
  999 CONTINUE
      IF(IJ.EQ.10000.AND.DPPHAT.GT.0.00000000)THEN
      WRITE(3,567)'IJ=10000 AND DPPHAT>0.00000000','IJ:',IJ,'J:',J
  567 FORMAT(/X,A30,3X,A3,I5,3X,A2,I5)
      ELSE
      WRITE(3,568)'CONVERGENT PPHAT'
  568 FORMAT(/X,A16)
      ENDIF
      STOP
      END
```

Chapter 6

With the recent development of microfoundations, it has become apparent that neither the Keynesian nor the various Classical Schools are entirely right in their description of behaviour in the product and labour markets. The limitation of the Classical Schools is that: because of customer-supplier and worker-employer relationships, prices and wages are unlikely to be fully flexible in the short run so that any assumption of instantaneous market equilibrium in the product and labour market would be unrealistic. However, as the recommendation of a non-activist policy in the various Classical Schools (particularly those in the New Classical School such as Sargent and Wallace (1975,1976)) is based on a model assuming instantaneous market-clearing, a rejection of this assumption implies that the policy ineffectiveness proposition of these schools should be reconsidered more carefully [1]. On the other hand, the assumption of an institutionally fixed wage (and possibly fixed price) in some of the Keynesian models has seriously limited their relevance to long run analysis where market forces do appear to act.

Indeed, both product and labour markets behave in a manner in some ways similar but in other ways different from that depicted by the Keynesian and Classical Schools. Perhaps the following statements in Schultz (1985) [2] will give a good idea of

---

[1]This is not to say that all the works by the New Classical School are without insight. For one contribution that should never be overlooked is their emphasis on the importance of expectation formulation (i.e. rational expectation instead of adaptive expectation) in macroeconomic modelling. Indeed, rational expectations has already been included in many types of Keynesian models. Taylor(1977), Fischer(1977), and Buiter and Miller(1981,1982) are a few examples of these. Nevertheless, it must be emphasized that the New Classical contribution to expectation formulation should be separated from those results related to the assumption of instantaneous market equilibrium.

[2]While Schultz's discussion is on the wage decision, a similar type of logic also applies to the price decision.

what a satisfactory model should be capable of explaining:

> *"Within the constraints imposed by implicit contracts, wages (and prices) in individual firms have to be adjusted to deal with changing condition." [P.12]*
>
> *" Since wage (and price) changes are difficult and impose strains on long-term relationships, the wage (and price) once set has to last for a while, typically at least a year..."[P.12]*
>
> *" ...expectations...will...exert an important influence over the current wage (and price) decision...But what is central to my message is that the relevant forecast does not assume prompt adjustment to a new equilibrium wage (and price) but rather the more hesitant and gradual process described above." [P.12]*

Here, the first statement refers to the Classical requirement that prices and wages should be flexible and market forces would exert their influence in the long run. The second requirement is Keynesian in the sense that there should be some kind of price and wage rigidity in the short run. The third statement is related to the signal extraction problems and the cost of changing price discussed in Chapter 2. Indeed, our model is capable of explaining that, before the cumulated change in costs and demand exceeds the thresholds, prices will remain sticky in the short run. The effect of these changes in costs and demand will however be revealed once the threshold is exceeded. This is why we believe that the Classical Schools are right in claiming that market forces will exert their influence in the long run.

The models built in Chapter 2 and 3 also illustrated that both the Keynesian and Classical Schools – after appropriate refinement – can be integrated under a general framework. Indeed, we regard this as a more important message than the mere identification of the necessary refinements in the two schools of thought. We believe that one of the most important items for future research is to explain how planned excess capacity, cost oriented pricing, sticky price with respect to demand shocks, long run Neoclassical maximization, customer-supplier relationships, worker-employer relationships, wage rigidity, involuntary unemployment, cyclical variations of output and employments, implicit guarantee of wage and employment, labour hoarding, procyclical productivity and so on can be made consistent with each other.

Before returning to the philosophy or perception behind the thesis, let us first summarize the results obtained in Chapters 2-5 and the potential areas of further research.

6.1 Pricing Decision

6.1.1 Summary of Results

Well before any formal development of the microfoundations of sticky price, proponents of mark-up pricing had long suspected that the mark-up formula P=(1+m)AC – albeit being an average cost pricing rule – could indeed be superior to the instantaneously marginal cost pricing rule of Neoclassical theory. Their suspicions concerning the stickiness/sluggishness of price with respect to demand shocks were then partially supported by the development of the B-M-R models (remember the limitations of these models discussed in Chapter 1). The aim of Chapter 2 was to build a more satisfactory model so as to explain the extensive degree of observed price stickiness (with respect to demand shocks) and to check the other hypotheses implicit in the usual average cost pricing rule P=(1+m)AC. Unlike the Menu Cost Hypothesis, we believe that the observed degree of price stickiness is far more extensive than can be explained by a small menu cost. Instead, we emphasize the role of a large reputation cost of changing price, the significance of signal extraction problems and their interaction in generating the observed degree of price stickiness. This view of the world is supported by the empirical work in Chapter 4. Our model also demonstrated that some of the assumptions implicit in the Menu Cost Hypothesis are unjustified. On the whole, we believe our model does help to account for the observed degree of price stickiness something which cannot be satisfactorily explained within the existing literature.

We then argued that the three sources of price stickiness will be much weaker for the case of cost shocks, therefore justifying the asymmetric treatment of cost shocks and demand shocks by proponents of mark-up pricing. The argument here also provides some hints about the evolution of cost-oriented pricing: the fact that most of the cost shocks are general and persistent (especially in an inflationary environment) is likely to be one of the main reasons for producers developed the practice of raising price with cost. Once such a practice or rule of thumb is developed, customers will find it "normal" and "fair" for price to rise with cost [ie they still find their rule of inter-temporal comparison of price reliable], and each producer's expectation of an eventual rise in competitors' prices in the face of a general and persistent cost shocks will be justified by experience. These further reduce each producer's cost of changing price with respect to cost shocks, and hence make price more responsive to cost shocks than demand shocks of the same size [3]. Nevertheless, the asymmetry between cost shocks and demand shocks should not be carried too far. This is so because:

(a) as long as $A_c > 0$, prices will not be adjusted with every small change in cost [Indeed, the fact that most suppliers have chosen to pledge the constancy of price for some reasonably long period (eg one year) suggests that A_c (the fixed cost of changing price with respect to cost shocks) would not be too small, albeit smaller than that with respect to demand shocks.]; and

(b) when prices are changed, the change will take into account the permanent change in demand as well as the change in cost.

With regard to the third hypothesis of P=(1+m)AC, we have also explained that

[3] "Same size" here refers to shocks that will give the same % change in the desired level of price P*.

the unitary elasticity of price with respect to AC will be justified as long as general cost shocks (instead of specific shocks) dominate. This was supported by the empirical work in Chapter 4 [4] where the sum of the coefficients on wage and material costs is unity.

6.1.2 Further Remarks

A. Debates within the theory of mark-up pricing

Our model here also sheds light on the debates within the theory of mark-up pricing. As discussed in Chapter 1, even within the theory, there are disagreements on whether (i) the mark-up is fixed [5] or variable [6]; and (ii) the mark-up is based on full cost [7] or variable cost [8].

The model in Chapter 2 and the empirical work in Chapter 4 suggest that the planned profit mark-up will be a complicated (stepwise) but increasing function of expected demand. Thus, strictly speaking, neither of the two sides in the debate is right. We will come back to this in section 6.3. With regard to the second issue, the model in Chapter 2 suggests that, in the absence of inflation, the mark-up will be based on variable cost but subject to the condition that the full cost is covered by revenue. However, with a moderate inflation rate, both the fixed cost and the variable cost will be moving closely with each other, and the distinction between the full cost principle and the "variable cost principle" will be blurred.

---

[4] Note that there had been some kind of approximations in the derivation of the pricing decision.

[5] Such as Nordhaus & Godley (1972).

[6] Such as Gordon(1975).

[7] Such as Hall & Hitch (1939).

[8] Such as Kalecki (1939).

B. Importance to Macroeconomics

The microfoundations of price stickiness are important because different definitions of stickiness may give different macroeconomic predictions. As explained in Chapter 1, while the Lucas-Sargent proposition would still be preserved with McCallum's definition of price stickiness (defined as the difference between the anticipated and the market clearing prices), there is no sufficient foundation for McCallum's definition and a replacement by the correct definition (defined as the difference between the current and the previous price level) will give the Keynesian non-neutrality result.

6.1.3 Areas of further research

While it yields some interesting results, our model in Chapter 2 is also subject to many limitations, and further research in this area is necessary. Firstly, our specification of the cost of changing price is only a short cut and a more satisfactory way of modelling is to include an explicit specification of the customer's response to price changes. Two further extensions therefore are: the formal modelling of such customer response; and the integration of the customer's decision and supplier's decision into a general model [9].

Next, even within our "short-cut" analysis with its "ad hoc" specification of the cost of changing price, solution for a few more complex cases might provide a better understanding of pricing behaviour. For example, while the discussion in section 6.1.1 tends to suggest that the existence of a moderate inflation rate will ensure that mark-up pricing is a good approximation to profit maximizing behaviour in a customer market, a formal proof is not yet available and a solution procedure for the case with expected

[9]That is, how the shopping/search cost provide incentives for the suppliers to pledge some kind of constancy of price on one hand and the customers to rely on intertemporal comparison of prices on the other; and how a price change affect a customer's decisions between sticking with the original supplier and making new searches.

inflation [10] should be developed. Moreover, instead of having the cost of changing price fixed for the whole year, the cost might well exhibit a seasonal pattern. For example, it may be less costly to raise price at the beginning of the year, or when new models are introduced, than any other period. In that case, price adjustment might also exhibit a seasonal pattern. A formal analysis (with an expected sequence of the cost of changing price) of this might also be worthwhile. Besides, most of our discussion in Chapter 2 focused on the special cases of either a pure demand shock or a pure cost shock. Even in checking the normal cost hypothesis , we only considered the special case where the combined effect of the cost and demand shocks is large enough so that the threshold is exceeded and any independent effect of demand on price will be revealed in the current change of price. We have not analyzed the more general (and possibly more realistic) case where demand and cost shocks coexist. Analysis of this general case is clearly worthwhile.

Finally, although the model in Chapter 2 is particularly helpful in explaining the extensive degree of price stickiness with respect to positive demand shocks, this may not be the case for negative demand shocks. If one were to believe a single reduction of price would only cause a gain (or negligible cost of changing price), neither the simple model in Chapter 2 nor those in Barro (1972), Mussa (1981), Rotemberg (1982a,b), Akerlof and Yellen (1985a,b), and Blanchard and Kiyotaki (1987) would be able to explain the extensive degree of price stickiness with respect to negative demand shocks. More satisfactory explanation is to use Okun's argument for a kink demand curve at the existing price P_0 (see section 3.2.1(A)(a)). The argument will be stronger if we extend the analysis with the fact that demand shocks in recession are usually temporary, as contrast to the permanent shocks assumed in the model in Chapter 2. Thus, an extension to the case of Okun's kink demand curve and temporary demand shocks are necessary

[10] See section 2.6.

for a satisfactory explanation of price stickiness with respect to temporary negative demand shocks.

6.2 Employment Decision

6.2.1 Summary of Results

The work of Akerlof and Miyazaki (1980) is useful because it explains the widely observed phenomenon of insured employment to those within the firms. It also implies that employment cannot be an entirely free variable so that the Keynesian version of Quantity Adjustment is unlikely to apply to the labour market. Nevertheless, Akerlof and Miyazaki were a bit hasty in going towards the conclusion of full employment equilibrium. In Chapter 1, we have seen how the efficiency wages, shirking, and turnover cost models can explain the existence of a pool of involuntary unemployment in the economy despite the insured employment within the firms. Our model in Chapter 3 provides another explanation of unemployment. It was shown that:

(a) Massive layoffs can occur in the face of a very adverse demand shock
[This result appears to be in accordance with what happens in periods of great recessions, such as that in U.K. in 1980-82 [11].];

(b) In the case of a moderate demand shock, there will be labour hoarding. Production effort – instead of wages and employment – will be the variable of adjustment. Our model also explains procyclical productivity, and implies that the Keynesian and efficiency wage models can be refined to incorporate this. The model also implies that previous empirical work on productivity should be refined to distinguish the regimes of layoff and no layoff; and

(c) In the case of a very adverse demand shock, it is always better – from the point of view of avoiding unemployment – to stimulate the economy <u>before</u> rather than <u>after</u>

---

[11] Note, we have only shown the result for the case of a permanent reduction in demand. Strictly speaking, this is different from that of recession where the reduction is temporary (no matter how large and prolonged). While we have not formally shown this, we believe the logic will apply as long as the reduction is large and long enough.

employers start laying off workers. Mild stimulation policies after the retrenchment will have no effect on employment.

With these results in mind, the Remarks in section 2.7.1 have briefly discussed how cyclical unemployment can evolve even in the cases of mild demand shocks.

Our results also point out the limitation of relationships such as Okun's Law or the (Short Run) Phillip Curve in which a stable one-to-one relationship between unemployment rate and the level of aggregate demand is explicitly or implicitly assumed. According to our analysis, the employment response for the case where cumulated demand changes exceeds the threshold will be very different from the case where the threshold is not exceeded. Even with a mild demand shock, the effect will still depend on the initial cross-sectional distribution of employment between the hiring point and layoff point. Thus, unless the cross distribution always stays the same, there will not be any stable relationship between unemployment and aggregate demand. However, Bertola and Caballero (1990) have shown that such an ergodic distribution will not exist whenever there are occasional but ongoing large demand shocks. Since the latter is quite a reasonable assumption, it will be misleading to assume that the effect of aggregate demand policies on the unemployment rate will be the same at all times [12].

Before leaving for the potential areas of further research, it would be interesting to compare the type of unemployment analyzed in our model and that in the efficiency wage, shirking and labour turnover models. In the latter, the pool of unemployment is static. Also, a small amount of such unemployment has the desirable effect of discouraging shirking, reducing turnover and raising efficiency; and the Government is incapable of eliminating all such unemployment by means of macroeconomic policies. On

[12] In other words, the Phillip Curve, being a function of the initial cross distribution of employment, will be shifting instead of being a single non-shifting well defined curve at all time.

the other hand, the unemployment arising in our model is strictly undesirable and it pays for the Government to use stabilization policies to eliminate such unemployment.

6.2.2 Areas of Further Research

For the sake of simplicity, the model in Chapter 3 assumed that producers were only allowed to change either (i) production effort; or (ii) employment in the face of demand shocks. In reality, there are more options available. For example, in view of the cost of layoff or the cost of hoarding excessive labour in the case of adverse demand shocks, employers (particularly those with erratic demand) may find it worthwhile to establish the practice of

(i) employing casual labour with higher wages but no guarantee of employment; or/and

(ii) subcontracting part of the production out.

Although these practices may involve a higher cost at periods of high demand, they have the advantage of reducing the burden on employers (in terms of the cost of layoff or the cost of hoarding excessive labour) in the case of an adverse demand shock. It would be interesting to formulate this idea explicitly and show that those employers with erratic demand may choose to follow such practices.

Besides, if the negative demand shock is expected to be temporary, employers may also have the choice of

(i) building up inventory;

(ii) asking the workers to do maintenance work such as painting, repairing etc.; or

(iii) slowing down new recruitment (for the replacement of retirements or quits).

We conjecture that the addition of these options could raise the stickiness of employment (and possibly the stickiness of price) with respect to demand shocks. We expect employers to establish the above practices up to the level where no layoff is required for normal variations in demand. Only in the case of very adverse demand shocks will employers consider breaking the implicit guarantee of employment.

Yet, even if it is necessary to reduce employment, there may still be ways to reduce the associated cost. For example, instead of laying off the workers indiscriminately, the employers can adopt the practice of

(i) laying off the new or the least productive workers; and/or

(ii) closing a whole plant, with the plants in other geographic areas being unaffected.

Alternatively, some employers might find it less costly to introduce part-time working schedule (a model similar to that of Chapter 3 can be easily formulated to give the bang-bang decision characteristic of a part-time working schedule). Thus, despite the insightful result we have obtained with the simplified model in Chapter 3, the presence of other options to the firm implies that further research on this area is needed.

Another potentially fruitful line of research is the wage decision. As mentioned in Chapter 1, it appears that wages are usually

(i) fixed in nominal terms for some periods; and

(ii) revised periodically with the real wage being guaranteed within a narrow range.

This kind of phenomenon can be explained by the following intuition. In addition to the reputation cost of having a real wage that deviates from the norm, there is another important cost: the cost of bargaining between the employers and workers. To reduce such a cost, both parties may find it worthwhile to establish the practice of reviewing the wage periodically. When such a period comes, the employer will revise the wage. If the employer decides to raise the wage below the norm, efficiency and worker's attachment may be reduced (the second is part of the reputation cost). If the employer decides to raise the wage above the norm, efficiency and worker's attachment are raised but the payroll is also raised. An explicit model on this analysis would be fruitful to clarify the relationship between nominal and real wage rigidity.

Bearing in mind the result of insured employment in the employment decision, the above intuition of wage decisions suggests that both the wage and employment will be sticky in the short run. This suggests that the traditional debate on Price versus Quantity Adjustment is not applicable to the labour market. To formalize this idea we need to integrate the above intuition regarding the wage decision into the model of employment decisions discussed in Chapter 3.

6.3 Empirical Work

In the empirical work reported in Chapter 4, the cost of changing price is found to be at least equivalent to 5.31% of the change in the "desired" level of price. For an average profit margin of (say) 30%, this is equivalent to a 4.07% loss of discounted profit (above the variable cost) over the future. Given the huge tonnage of steel beams delivered, this is equivalent to enormous amounts in sterling which could not be reasonably attributed to the mere presence of menu costs. On the other hand, the estimate is not incompatible with the presence of important reputation costs. Moreover, the menu cost interpretation has the disadvantage of predicting more frequent price change for products with a large market, since the fixed menu cost is relatively less important. On the contrary, our emphasis of a significant reputation and a negligible menu cost will imply that the total cost of changing price will be roughly proportional to the size of the market and hence predict that the frequency of price change is independent of the size of firm.

While our lowest estimate of the cost of changing price is found to be equivalent to a 5.31% change in the "desired" level of price, we also explained in Chapter 4 that this estimate may be subject to downward biasedness. By allowing stochastic variations in the threshold, our preferred equation suggests that, for a 10% inflation rate, the cost of changing price could be as much as a 10.14% change in the "desired" level of price. Our estimates also suggest that there is weak evidence that the cost of changing price will rise with expected inflation. Nevertheless, the significant constant term (and the weak inflation term) in the threshold ($\gamma_0 + \gamma_1 \ \Delta\ln(P_{wt}^e)$) implies that price adjustments will be more frequent in the case of higher inflation.

In addition to the aim of estimating the cost of changing price, we also attempted to test the Normal Cost Hypothesis (ie whether the planned profit mark-up is a fixed or an increasing function of demand) in Chapter 4. Unlike previous empirical work, the theoretical foundation here is stronger and the data set is more suitable. First, the estimation is based on individual data instead of aggregate data. With the help of a more elaborate theoretical foundation in Chapter 2, we also derive a two-regime price equation where the mark-up equation [P=(1+m)AC] will only hold in the raised price regime (i.e. $P = \overline{P}$ for the sticky price regime). Moreover, expected demand instead of current demand is used as one of the explanatory variables. A significant demand effect was found. Thus, unlike Godley and Nordhaus (1972), our estimation implies a rejection of the simple Normal Cost Hypothesis. Moreover, the sum of the wage cost and material cost is found to be close to the value of unity. This is also more satisfactory than the coefficient of 0.6 reported by Godley and Nordhaus [c.f. our review of Laidler and Parkin's criticism on Godley and Nordhaus (1972) in section 1.2.4]. While rejecting the Normal Cost Hypothesis, our results are also at odds with the previous empirical work on the determination of the mark-up. While this earlier work generally assumes that price is a <u>continuously</u> increasing function of demand, our two-regime equation suggests that price will only be a <u>stepwise</u> increasing function of demand (ie demand will only exert its effect on price when the conditions for raising price are satisfied).

In addition to the above, our results also suggest that:

(i) with expected inflation, there will be some pre-adjustment of price (if the condition for raising price is satisfied); and

(ii) a 1% change in expected demand will cause an approximately 0.3% change in the "desired" level of price.

6.4 The Simulation Result

Although the simulation exercises in Chapter 5 were only carried out for specific values of the parameters, they help to place some of the apparently contradictory results on the neutrality of money, such as those in Caplin and Spulber (1985) and Blanchard and Fischer (1989), into context.

In the basic Menu Cost model, it was implicitly assumed that producers are bunched at the neighbourhood of the optimal price so that a small change in the money supply, and hence also nominal demand, will leave all prices unchanged and therefore impinge entirely on output. With such an implicit assumption, we are effectively ignoring the possibility that some producers are just at the margin of raising price so that a small rise in the money supply will cause them to make a large change in price. If there are enough producers in this position then the previous effect of monetary increment on output could be reversed. The case that producers are bunched at the margin of raising price is not as unlikely as Akerlof and Yellen (1985a,b) suggest. Indeed, as long as money is growing monotonically, our simulation exercise suggests that they are sufficiently frequent to guarantee that money is "on average" neutral.

The case of neutrality considered by Caplin and Spulber is also a very special one where, at any point of time, the rise of output in some firms arising from a monetary expansion is just cancelled by the reduction of others. While Caplin and Spulber's case is certainly an interesting example where money is perfectly neutral, we do not require such strong conditions to produce a result that is strikingly different from the menu cost hypothesis. Indeed, as long as money is growing monotonically, our simulation exercises illustrate that money will be on average neutral over time. This is true even if there were occasional but ongoing large common shocks which prevent the idiosyncratic shocks from reshuffling the distribution of price deviations towards a uniform distribution at the steady state.

Following Caplin and Spulber, Blanchard and Fischer (1989) tried to re-establish the non-neutrality result by considering the case where the ergodic distribution may not be uniform. Nevertheless, their case of a symmetric random walk in the money supply is not very realistic in view of the observed upward trend of the money supply in most countries. With the addition of an underlying trend, their conclusions are changed significantly (e.g. a very strong underlying trend will imply monotonic monetary growth and hence neutrality of money over time).

On the whole, we think that the discussion of neutrality versus non-neutrality in the above papers is not general enough. Instead, with the result from the simulation exercises in mind, we propose the following hypotheses:

(a) As long as money is growing (falling) in one direction, any change of money supply in the same direction will be on average neutral over time.

The reasoning is that: even if the distribution of price deviations is skewed away from the return points at this moment, subsequent change in money supply will reverse the skewing at a later point of time. As a result, the cumulative small changes in output associated with the monetary change around this moment will be cancelled by a large but opposite change in output at a later point of time. Thus, output is approximately fluctuating between an upper and a lower bound. A corollary of this is that one cannot keep raising demand (and output) by means of indefinite increases in money supply. Of course, this does not mean that every type of monetary change will be neutral. Indeed,

(b) if the money supply has been going in one direction, a change in the money supply in the opposite direction will be non-neutral.

Thus, if there is an underlying growing trend in money supply, a reduction in money supply will cause a reduction in output. The policy implication of this is very Keynesian: In a generally inflationary economy, the government should avoid any reduction in money supply. Moreover, if there were any exogenous reduction in aggregate demand, a rise in money would usually alleviate the initial reduction of demand (provided that the price deviations are not skewed at the upper thresholds). Nevertheless, it must be emphasized that the rise in money supply will have no "long run" effect on demand (i.e. no effect on the bounds of output) so that the policy will be more effective for temporary instead of permanent reductions in demand.

6.5 Final Remark

We end the thesis with an important message for further research: our belief that many of the practices or rules of thumb established in the economy may not be as suboptimal as commonly supposed. On the contrary, they are actually cost saving devices, with at least one party of the market recognizing the possibility of a substantial gain and hence offering such a scheme that will induce the other party's participation for a share of the possible gain.

In this thesis, we have encountered a few examples of these cost saving devices. The first example with significant macroeconomic implications is the suppliers' tendency to pledge a stable pricing policy. Such practice is welcomed by both parties because it reduces the customers' shopping cost on one hand and ensure the suppliers more stable demand on the other. The implicit guarantee of employment is another important example which has the effects of (i) reducing the risk of the worker being unemployed and enables the employer to offer a lower wage; and (ii) allowing the non-transferable surplus arising from worker's acquisition of firm specific skill/knowledge through on-the-job training to be shared by both parties. Parallel to the second example is the implicit guarantee of a wage which allows the exploitation of a surplus arising from the difference in risk aversion between the employer and the worker. Moreover, the agreement to review the wages periodically; and the implicit understanding of charging a price on a mark-up basis of cost (i.e. cost-oriented pricing) are also devices to reduce the bargaining cost between the two parties.

Yet, the list of these practices, even just in the case of product and labour markets, is much longer than that listed above. The announcement of sales as a package of a fall and then a rise of price (instead of two separate announcements of price changes); and the inclusion of all seasonal price variations in one single price menu are devices to reduce uncertainty to customers and the reputation cost of changing price to suppliers arising from an otherwise non-preannounced irregular price change. In the labour market, the practices of hoarding labour; building up inventories; asking the workers to undertake maintenance; and stopping recruitment are also cost saving devices in the face of temporary negative demand shocks.

The implication of the above message is that Neoclassical economists should be prepared to look more closely at what is actually happening in the economy; and that Keynesian economists should perhaps spend more effort in explaining how the observed phenomena originate. The fact that many of the Keynesian results appeared to be in better accordance with reality may just be a reflection of a second best approach: they have spent more effort in observing the phenomena/characteristics in the economy and are therefore more often capable of producing a realistic result. Nevertheless, they do not have a monopoly in truth. As we have seen in this thesis: employment is not as flexible as their theory of Quantity Adjustment supposes, and mark-up pricing is only an approximation to true profit maximization. This is why good microfoundations for the

Keynesian approach are so necessary. Indeed, the common criticism of the Keynesian approach as an essentially short run analysis is a reflection of this very weakness.

Bibliography

Akerlof, G.A. and Miyazaki (1980). The Implicit Contract Theory of Unemployment Meets the Wage Bill Argument. Review of Economic Studies (January 1980) XLVII, pp. 321-338.

Akerlof, G.A. and J.L. Yellen (1985a). Can Small Deviations from Rationality Make Significant Differences to Economic Equilibria. American Economic Review, vol.75, No.4, pp. 708-720.

Akerlof, G.A. and J.L. Yellen (1985b). A Near-Rational Model of the Business Cycle, with Wage and Price Inertia. Quarterly Journal of Economics, vol.100, Supplement 1985, pp. 823-838.

Azariadis, C. (1975). Implicit Contract and Underemployment Equilibria. Journal Of Political Economy, vol.83, No.6, pp.1183-1202.

Bailey, M.N. (1974). Wages and Employment Under Uncertain Demand. Review of Economic Studies (January 1974), vol.41, pp. 37-50.

Baldwin, R. and P. Krugman (1986). Persistent Trade Effect of Large Exchange Rate Shocks. NBER Working Paper No. 2017, September 1986.

Baldwin, R. (1988). Hysteresis in Import Prices: The Beachead Effect. American Economic Review, September 1988, vol.78, no.4, pp.773-785.

Barro, R.J. (1972). A Theory of Monopolistic Price Adjustment. Review of Economic Studies, 1972, pp. 17-26.

Bean, C. (1987). Sterling Misalignment and British Trade Performance. Centre for Labour Economics, LSE, Discussion Paper No. 288, July 1987.

Berle, A.A. and G.C. Means (1932). The Modern corporation and Private Property. Commerce Clearing House, 1932.

Bertola, G. and R.J. Capallero (1990). Kinky Adjustment Costs and Aggregate Dynamics. NBER Macroeconomic Annual Conference, March 9-10, 1990.

Blair, J.M. (1974). Market Power and Inflation: A Short Run Target Return Model. Journal of Economic Issues, vol. 8, June 1974, pp.453-478.

Blanchard, O.J. (1983). Price Asynchronization and Price Level Inertia, in R. Dornbusch and M.H. Simonsen (eds), Inflation, Debt, and Indexation. Cambridge, MA: MIT Press.

Blanchard, O.J. and N. Kiyotaki (1987). Monopolistic Competition and the Effects of Aggregate Demand. American Economic Review, September 1987, pp. 647-666.

Buiter, W.H. and M. Miller (1981). Monetary Policy and International Competitiveness. Oxford Economic Papers, 33, July Supplement, pp. 143-175. Also in W.A. Eltis and P.J.B. Sinclair eds. The Money Supply and the Exchange Rate, Clarendon Press, Oxford 1981.

Buiter, W.H. and M. Miller (1982). Real Exchange Rate Overshooting and the Output Cost of Bringing Down Inflation. European Economic Review, 18.

Caplin, A.S. (1985). The Variability of Aggregate Demand with (S,s) Inventory Policies. Econometrica, vol.53, No.6, November 1985.

Caplin, A.S. and D.F. Spulber (1987). Menu Costs and The Neutrality of Money. Quarterly Journal of Economics, vol.CII, Issue 4, November 1987, pp.703-725.

Fischer, S. (1977). Long Term Contracts, Rational Expectations and The Optimal Money Supply Rule. Journal of Political Economy, 85, No.1, pp. 191-205.

Gordon, D.F. (1974). A Neoclassical Theory of Keynesian Unemployment. Economic Inquiry, December 1974, 12, 431-459. Also reprinted in the Phillips Curve and Labour Market, Brunner and Meltzer eds., Amsterdam: North-Holland, 1976.

Gordon, R.J. (1975). The Impact of Aggregate Demand on Prices. Brookings Papers on Economic Activity, 3:1975, pp.631-655.

Gordon, R.J. (1981). Output Fluctuations and Gradual Price Adjustment. Journal of Economic Literature, vol.XIX, June 1981, pp.493-530.

Godley, W. and W.D. Nordhaus (1972). Pricing in the Trade Cycle. Economic Journal, September 1972, pp.853-882.

Gray, J.A. (1976). Wage Indexation: A Macroeconomic Approach. Journal of Monetary Economics 2, 1976, pp.221-235.

Gray, J.A. (1978). On Indexation and Contract Length. Journal of Political Economy, 86 (1).

Hall, R.L. and C.J. Hitch (1939). Price Theory and Business Behaviour. Oxford Economic Papers, 1939, pp. 12-45.

Kalecki, M. (1939). Essays In the theory of Economic Fluctuations. London: Allen and Unwin, pp.13-41.

Koutsoyiannis, A. (1979). Modern Microeconomics. London: MacMillan.

Laidler, D. and M. Parkin (1975). Inflation: A Survey. Economic Journal, 85 (December 1985) pp. 741-809.

Mankiw, N.G. (1985). Small Menu Costs and Large Business Cycles: A Macroeconomic Model of Monopoly. Quarterly Journal of Economics (May 1985) pp.529-537.

McCallum, B.T. (1977). Price-Level Stickiness and the Feasibility of Monetary Stabilization Policy with Rational Expectations. Journal of Political Economy, 1977, vol.85, no.3, pp.627-634.

Means, G.C. (1935). Industrial Prices and their Relative Inflexibility. S.Doc. 13, 74, Congl Sess (Government Printing Office, 1935).

Means, G.C. (1939). The Structure of the American Economy, pt. 1 : Basic Characteristics. A report prepared by the Industrial Section under the direction of Gardiner C. Means (Government Printing Office, 1939) pp.126-145.

Miller, M.T. and D. Orr (1966). A Model of the Demand for Money by Firms. Quarterly Journal of Economics, 1966, 80, pp 413-435.

Morishima, M. (1984). The Economics of Industrial Society. Cambridge University Press.

Mussa, M. (1981). Sticky Prices and Disequilibrium Adjustment in a Rational Model of Inflationary Process. American Economic Review, vol.71, No.5, pp. 1020-1027.

Okun, A.M. (1981). Prices and Quantities: A Macroeconomic Analysis. Washington D.C.: The Brookings Institution, 1981.

Poole, W. (1976). Rational Expectations in the Macro Model. Brookings Papers on Economic Activity, 2, pp. 463-514.

Rotemberg, J.J. (1982a). Monopolistic Price Adjustment and Aggregate Output. Review of Economic Studies, 1982, vol.XLIX, pp. 517-531.

Rotemberg, J.J. (1982b). Sticky Prices in the United States. Journal od Political Economy (1982), vol.90, No.6, pp.1187-1211.

Rotemberg, J.J. (1983). Aggregate Consequences of Fixed Costs of Price Adjustment. American Economic Review, June 1983, vol.73, No.3, pp.433-436.

Salop, S.C. (1979). A Model of the Natural Rate of Unemployment. American Economic Review, March 1979, vol.69, pp.117-125.

Sargent, T.J. and N. Wallace (1975). Rational Expectations, the Optimal Monetary Instrument and the Optimal Money Supply Rule. Journal of Political Economy, 83, No.2, pp. 241-254.

Sargent, T.J. and N. Wallace (1976). Rational Expectations and the Theory of Economic Policy. Journal of Monetary Economics, 1976, vol.2.

Schlicht, E. (1978). Labour Turnover, Wage Structure and Natural Unemployment. Zeitschrif fur die Gesamte Staatswissen-Schaft, June 1978, 134, pp. 337-346.

Schultz, C. (1985). Microeconomic Efficiency and Nominal Wage Stickiness. American Economic Review, vol.1975, No.1, pp. 1-15.

Shapiro, C. and J.E. Stiglitz (1984). Equilibrium Unemployment as a Worker Discipline Device. American Economic Review, June 1984, vol.74, No.3, pp. 433-444.

Solow, R. (1979). Another Possible Source of Wage Stickiness. Journal of Macroeconomics, Winter 1979, 1, pp.79-82.

Solow, R. (1980). On theories of Unemployment. American Economic Review, March 1980, vol.70, pp. 1-11.

Stiglitz, J.E. (1974). Alternative Theories of Wage Determination and Unemployment in LDC's: The Labour Turnover Model. Quarterly Journal of Economics, 1972, pp. 194-227.

Stiglitz, J.E. (1976). The Efficiency Wage Hypothesis, Surplus Labour and the Distribution of Income in LDCs. Oxford Economic Papers, July 1976, 28, pp.185-207.

Taylor, J.B. (1977). Aggregate Dynamics and Staggered Contracts. Journal of Political Economy, 1980, vol.88, No.1, pp.1-23.

Yellen, J.L. (1984). Efficiency Wage Models of Unemployment. American Economic Review, Papers and Proceedings, May 1984, pp.200-205.

Yip, S.L. and T.G. Yung (1985). Signal Extraction Problems, Cost of Changing Price and Price Sluggishness. University of Hong Kong, Department of Economics, discussion paper no.60, 1985.