THE PERSISTENCE OF INFLATION

Thomas M. Humphrey

Among the most exasperating and puzzling of recent economic phenomena is the apparent intractability of the inflation rate. Once started, an inflation becomes difficult to subdue. It seems to develop a momentum of its own, independent of other basic economic conditions. It resists or at best responds only sluggishly to traditional restrictive policies. Its persistence in the face of high unemployment and excess capacity has resulted in the addition of the term *stagflation* to the economist's lexicon.

What accounts for the stickiness of the inflation rate? Why is it so policy-resistant and difficult to control? Why is it so slow to decelerate even when demand is slack? Many economists believe that the answers lie in the mechanism through which inflationary impulses are transmitted through the economy.[1] Embedded in this mechanism are certain delays or lags that may slow the spread of inflation over the total price structure and may also prolong its duration. Particular prices that lag behind general inflationary movements have to catch up later to reestablish their relative position in the price structure. This lag/catch-up characteristic of the inflationary transmission mechanism is offered by some as an explanation of why strong upward pressures on prices persist long after demand slackens.

The most important lags in the process by which inflation is diffused through the economy are the *price-adjustment lag* and the *expectations-formation lag*. The first refers to the delayed response of the rate of price increase to shifts in aggregate demand. Demand pressure is transmitted to prices via a complicated and circuitous channel that runs from output to inputs to input prices to costs and finally to product prices. The price-adjustment lag accounts for the time it takes for demand to affect prices through the channel of costs. It should be noted that many of these costs themselves adjust slowly, partly because they are influenced by sticky price anticipations. These sticky anticipations are described by the second lag, which refers to the slowness with which expectations of future inflation are revised when individuals realize that actual inflation has turned out to be different than was expected. For example, immediately following periods of rapidly accelerating inflation, expectations about the future behavior of prices continue to reflect the preceding price experience even though the current rate of inflation may be decelerating. Some analysts point to these lags as the reason that inflation is so persistent and hard to subdue, even in the face of high unemployment and excess capacity.

The purpose of this article is to examine the price-adjustment and expectations-formation lags and to indicate how they may affect the speed, pattern, and duration of inflation. The article proceeds in the following manner. First, it identifies the location and describes the operation of the lags in the inflationary transmission mechanism. Second, it provides an explanation of the existence of the lags. Third, and most important, it analyzes the policy implications of the lags. Finally, the appendix contains a brief description of how the lags are treated in simple analytical models of the inflationary process.

Sketch of the Inflationary Process The first objective is to describe the operation of the lags in the inflationary mechanism. As a necessary preliminary, a brief description of the inflation process is offered, with emphasis on the time sequence or chronological order in which key economic variables (spending, output, costs, prices, expectations, etc.) adjust to inflationary pressures. The hypothetical example presented below may not conform to all of the inflationary episodes experienced in the U. S. although it probably typifies most of them.

As a first approximation, the sequence of events in a typical inflation may be described as beginning with an increase in aggregate demand to a level in excess of the economy's capacity to produce.[2] Such

---

[1] See, for example, the studies by Cagan [1, 2, 3], and Laidler [5, 6]. See also the articles by Friedman [4] and Selden [8]. The present article draws heavily from Cagan and Laidler.

[2] In starting with demand, the example does not deny that supply shocks—crop failures, strikes, natural disasters and the like—may also play a role in inflation. Given the random, transient, and often reversible nature of these shocks, however, it is unlikely that they could produce a continuing (sustained) inflation. What they can do is to temporarily intensify an inflation generated by demand forces. This is the problem of price blips illustrated by the transient double-digit episode of 1974 when supply constraints caused inflation to deviate temporarily from its basic trend rate determined by excess demand. When the supply shocks abated, inflation returned to its basic path. Note also that supply shocks can be treated as an excess demand phenomenon, albeit one in which the excess arises from falling supply rather than rising demand.

a step-up in the rate of spending might stem from any of a variety of causes including: (1) an increase in the government's budgetary deficits; (2) an episodic expansion in private spending associated, say, with a major innovation; and (3) an acceleration in the rate of monetary growth either actively engineered or passively permitted by the monetary authorities. Whatever the cause, the increased rate of spending, if it is to continue, must be supported by a sustained higher rate of monetary growth. Without this concomitant monetary growth, rising prices would simply reduce the real value of existing cash balances below the level that people desire to hold. In an effort to rebuild their price-eroded real balances, cashholders would cut expenditures, thereby bringing the inflation to a halt. An increased rate of spending cannot be maintained for long without the monetary expansion necessary to finance it. Assuming such expansion occurs, however, the corresponding increased rate of spending leads to a buildup of excess demands on the economy. Businessmen initially respond to this rising demand not by raising prices but rather by reducing inventories and expanding output. That is, their initial adjustments are to quantities; price adjustments come later—hence the so-called price-adjustment lag.

The stimulus to final output resulting from these quantity adjustments is transmitted back to earlier stages of production via increases in the demand for labor and material inputs, as well as for intermediate (semi-finished) products. Demand pressure passes downward to lower stages of production where prices —especially those of raw materials and other basic commodities—are particularly sensitive to increases in demand. These prices begin to rise almost immediately as do certain competitive (non-union) wages whose adjustment is not delayed by long-term union contracts or collective bargaining agreements. Since the output of earlier stages of the productive process constitutes the input of later stages, these rising wages and prices pass back up through the interindustry structure in the form of increased costs. In short, demand pressure is transmitted downward thus invoking cost increases that ripple upward until they finally reach the finished product stage. Here businessmen, operating with fixed percentage markups over costs, pass these cost increases on in the form of higher prices. These price increases induce workers operating under collective bargaining agreements to bargain for cost of living increases in the next wage contract. The resulting increases in labor costs are passed through into still higher prices. Such cost-induced price increases, of course, constitute the delayed price response to the prior shift in

demand. These cost and price increases provoke further rounds of wage and price increases that add impetus to the inflation through the channel of costs.

It should be noted that the expectations lag is in operation throughout the early and intermediate stages of the inflation. Price anticipations are based on perceived trends and so do not change quickly. Because these trends are estimated largely from past experience, it takes time to adjust expectations to higher current rates of inflation. Thus when the actual rate of inflation begins to rise, the expected rate is not immediately affected. Only after the actual rate has exceeded the expected rate for a time will the latter start to rise. But it will rise slowly at first because it continues to reflect the lower past rates of inflation. Eventually, however, the expected rate rises faster and faster as people become acclimated to the higher actual rate and adapt their expectations to it. Once aroused, inflationary expectations feed back into the current rate of inflation as firms and unions seek to raise prices and wages at the same rate as they expect prices in general to rise. At this stage of the inflationary process, where cost increases are passing through into price increases and the expected rate of inflation is rising to catch up with the actual rate, the main impact of the inflationary stimulus shifts from quantities to prices. The temporary output and employment effects diminish, and purely inflationary price effects dominate.[3] The lags, however, may be long; and the price system may take several years—years that may even encompass a business recession—to complete all adjustments. Particular

---

[3] Note that the expectations lag temporarily influences real economic activity by causing the nominal and real (i.e., inflation-corrected) prices of factor inputs to move in opposite directions. Put differently, the expected inflation component of nominal wage and interest rates may not adjust sufficiently to offset actual inflation thus causing changes in the corresponding real (price deflated) values of those variables. For example, when an inflationary stimulus begins, nominal wage and interest rates adjust upward by the amount of the expected inflation. Because the expected rate of inflation lags behind the actual rate, however, the adjustment is incomplete and consequently real wage and interest rates fall. Businessmen take advantage of the falling real wage and interest rates by hiring more labor and borrowing from banks in order to expand production. These effects, together with decreases in the demand for money (and hence rises in spending) induced by higher nominal interest rates, result in a temporary rise in real output and employment. The stimulus ends, however, when expected inflation catches up with actual inflation. When this happens, nominal wage and interest rates fully reflect the actual rate of inflation and real wage and interest rates return to their initial levels. Similar real effects operate on the downside when actual inflation falls below expected inflation. In this case the expectations lag results in a temporary rise in real wage and interest rates and these rising real costs act to temporarily depress real economic activity below its equilibrium level.

costs and prices that have lagged behind the others will have to catch up in order to restore the equilibrium price relationships disturbed by the inflation. Also, at the end of a serious inflation, inflationary psychology may continue to exert upward pressure on wages and prices even after other basic pressures have begun to diminish. Wages and prices continue to mount as inflationary expectations adjust to catch up with the accelerating price movements of the recent past. This catch-up process may extend well into an ensuing recession, thus resulting in the anomaly of rising prices despite slackening demand.
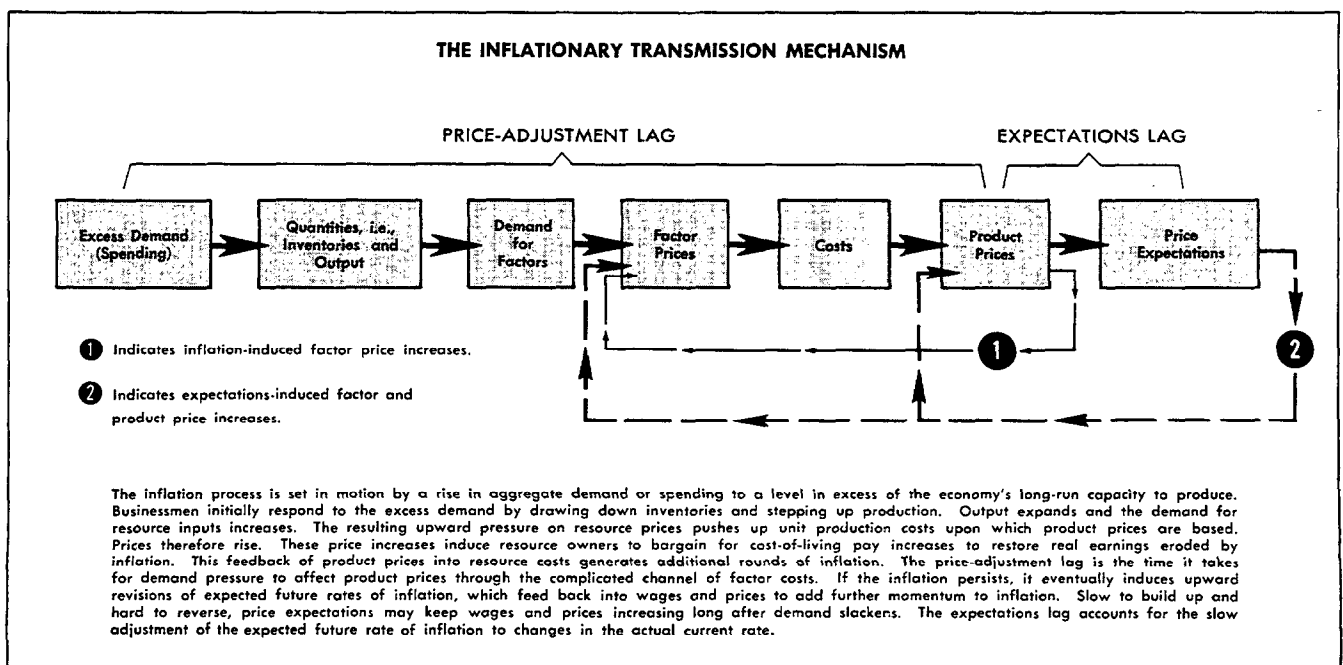
In the process just described, the chain of causation or pattern of adjustment runs from spending to output to costs to prices to expectations and back again to input and product prices. This sequential process is represented schematically in the chart, which also shows the location of the price-adjustment and expectations lags. It should be emphasized, however, that an important constraint exists to the continuous operation of this process. Specifically, the process cannot continue unless it is constantly refueled with additional supplies of money. A determined stand by the monetary authorities to deny this fuel will eventually bring inflation to a halt.

The Winding-Down Process The lags described in the preceding section continue to operate even after an effective program of monetary restraint is implemented. Spending slackens, output falls, and input demands decline; but prices nevertheless continue upward because recent cost increases are still working their way through the system and because

inflationary expectations continue to mount. But as demand pressures slacken throughout successive stages of production, the resulting weakening of costs eventually leads to an easing of prices of final goods and services. As the actual rate of price increase falls and remains for a time below the expected rate, expectations of further rises are gradually revised downward. Since the expected rate is a determinant of the actual rate, the decrease of the former will produce a further deceleration of the latter. But the price-adjustment and expectations lags draw out the winding-down process and the actual rate of inflation will converge on its new equilibrium level only slowly.

Factors Retarding Price Adjustment What characteristics of a modern economy account for the lags described in the foregoing sections? In the case of the price-adjustment lag, the answer is fairly straightforward. The lag arises from institutional arrangements and behavioral practices that operate to make many costs and prices relatively unresponsive to short-run shifts in demand.[4] For example, long-term contracts fix some wages and prices for substantial intervals of time. In addition, regulated rates of

---

[4] Note that the rigidities or inflexibilities described in this section refer not to **levels** of particular prices and costs but rather to **rates of change** of those variables. The concept of rigid price levels, of course, implies the complete absence of price change and is thus irrelevant to the analysis of inflation, which deals with continually rising, not constant, prices. The problem of sluggish inflation is not one of downward inflexibility of price levels, but rather the resistance of rates of price increase to departures from the established trend.

THE INFLATIONARY TRANSMISSION MECHANISM

PRICE-ADJUSTMENT LAG   EXPECTATIONS LAG

❶ Indicates inflation-induced factor price increases.

❷ Indicates expectations-induced factor and product price increases.

The inflation process is set in motion by a rise in aggregate demand or spending to a level in excess of the economy's long-run capacity to produce. Businessmen initially respond to the excess demand by drawing down inventories and stepping up production. Output expands and the demand for resource inputs increases. The resulting upward pressure on resource prices pushes up unit production costs upon which product prices are based. Prices therefore rise. These price increases induce resource owners to bargain for cost-of-living pay increases to restore real earnings eroded by inflation. This feedback of product prices into resource costs generates additional rounds of inflation. The price-adjustment lag is the time it takes for demand pressure to affect product prices through the complicated channel of factor costs. If the inflation persists, it eventually induces upward revisions of expected future rates of inflation, which feed back into wages and prices to add further momentum to inflation. Slow to build up and hard to reverse, price expectations may keep wages and prices increasing long after demand slackens. The expectations lag accounts for the slow adjustment of the expected future rate of inflation to changes in the actual current rate.

public utilities can be changed only after lengthy administrative and judicial proceedings. Moreover, there exists a variety of legal restrictions on wage and price flexibility, including minimum wage laws, rent controls, oil price ceilings, interest-rate ceilings, and the like. There are also numerous government-sanctioned trade-restriction and price-fixing arrangements, such as production and marketing quotas for certain agricultural products, import quotas, licensing and other restrictions to entry of professions, and resale price maintenance agreements. All these contribute to the stickiness of the prices affected.

An important factor retarding price adjustment is the inflexibility of many wages. Owing to long-term labor contracts, delays in the adjustment of price expectations, and perhaps also workers' money illusion (i.e., failure to distinguish between nominal and real wage increases), some wages tend to lag behind inflation during certain phases of the business cycle. When inflation accelerates in the upswing, these wages are often slow to respond. In later stages of a sharp inflation, however, wage increases may outstrip price increases. Wages may then tend to rise ahead of prices to restore real earnings eroded by past price increases and to protect real earnings from anticipated future price increases. But the lag becomes noticeable again when inflation is subsiding as the rate of wage increase shows little tendency to decelerate even though the rate of rise of other prices is diminishing.

Another important factor contributing to price rigidity is the price-setting behavior of large-scale firms operating in manufacturing industries. In the typical imperfectly competitive market, the price mechanism works in the long run as prices adjust to clear the market. In the short run, however, prices are relatively inflexible. Firms do not typically alter their prices in response to short-run shifts in demand. At least three explanations of this behavior have been offered.

According to the first, firms have difficulty distinguishing between real demand shifts specific to a particular industry and shifts in nominal aggregate demand for the output of all industries. Because they alter only the composition and not the overall level of aggregate demand, specific demand shifts may occur without inducing a corresponding rise in input prices. By contrast, an economy-wide demand increase leads to an equiproportional bidding-up of input prices. In the former case, industry cost and supply curves remain unchanged. In the latter case, however, cost and supply curves shift sharply upward, just matching the rise in demand. The rational entrepreneurial reaction to the former is a change in some quantity whereas the rational reaction to the latter is a change in prices. But when aggregate demand alters there is a good likelihood that each producer will tend to regard the shift in demand for his product as special to him and so adjust quantity rather than price. Only later, when cost changes become widespread, will producers correctly perceive the demand shift as an economy-wide phenomenon. Then and only then will they start to change their prices. A second explanation stresses the administrative inconvenience and costs of frequent price changes—for example, the expense of printing and disseminating new price lists—as a reason for sticky oligopolistic prices.

Still a third explanation of price inflexibility in imperfectly competitive markets begins with the observation that in the complex and dynamic modern industrial economy there is always much uncertainty about the equilibrium or market-clearing price. Firms operating in this kind of environment try to avoid the market disruption, confusion, and perhaps even outright price warfare that could result if each sought individually to determine the equilibrium price. In order to prevent such confusion from developing, firms seek ways to coordinate price changes. Such coordination, if successful, will assure that firms raise prices in unison and that price changes will not occur when demand shifts are thought to be temporary and reversible. Firms have devised several techniques to facilitate price coordination. Infrequent price changes are perhaps the simplest of these. Price leadership constitutes another such technique. In this case, one firm—often the largest in the industry—initiates price changes, and the rest more or less automatically follow that price change. Perhaps the most widely used mechanism for coordinating price behavior, however, is to base selling prices on unit labor and material costs that are the same for all firms in the industry. This practice is accomplished by the use of so-called unit cost or mark-up pricing formulas.

Unit-cost pricing is thought to be characteristic of many of the large oligopolistic firms that operate in American industry. The mark-up technique of pricing involves setting prices on the basis of a constant percentage markup applied to production costs per unit of output at some standard level of plant operation or capacity utilization. The chief cost components are unit labor and unit material costs. Included in the markup is the firm's profit margin per unit of output. This profit margin is usually set to provide a fixed target rate of return on equity. Since percentage markups or profit margins are treated as fixed constants in unit-cost pricing formulas, it follows

that formula-based price changes are strictly cost-determined, that is, they result solely from changes in unit labor and material costs. Prices respond to costs, not to demand. Moreover, since standard unit labor and material costs are roughly the same for all firms in the industry, unit-cost formulas insure that price changes will be uniform throughout the industry, thereby minimizing the risks of competitive price undercutting. Thus unit-cost pricing is consistent with the slow response of prices to shifts in demand, the dependence of prices on costs, and the coordination of prices in concentrated industries.

A Source of Confusion The long delay in the adjustment of many prices to demand pressure makes it difficult to distinguish cause from effect in the inflationary sequence and contributes to confusion in popular understanding of the source of inflation. The sequence of cost-price responses observed in concentrated industries, for example, may suggest that inflation is initiated by autonomous increases in costs. The use in such industries of mark-up or unit-cost pricing formulas, and the resulting dependence of prices on costs, means that firms do not raise prices unless there occurs a prior increase in costs. This cost-price sequence, with costs rising first and prices later, makes it appear that inflation is *caused* by rising costs when in fact excess demand is usually the culprit. For, as pointed out earlier, both the cost increase and the ability of firms to pass on this increase are due to a prior expansion in aggregate demand for final goods and services.

Similarly, the long lag in the response of oligopolistic prices to prior inflationary pressures may make it appear that large firms in concentrated industries play an important independent role in generating inflation. Due to the price-adjustment lag, catch-up price increases in such industries are often delayed until well after restrictive policies start to turn the inflation process around. The apparent perverse behavior of prices when markets are slack and unemployment is rising fosters the notion of giant firms arbitrarily exercising monopoly power by effecting autonomous increases in prices totally independent of economic conditions.[5] But it should be

noted that the price increases in such situations can be interpreted as constituting a delayed reaction to prior economic conditions of expanding demand and demand-induced cost increases and a near perfect example of the price-adjustment lag. They might well be interpreted as a delayed manifestation of the effects of a general inflation rather than of inflation-initiating price behavior.

From this interpretation it follows that big firms are not inflation starters. Nevertheless they may play an important role in the working out of the inflationary process. Specifically, they may be inflation prolongers. It follows logically from the mere existence of the price-adjustment lag that administered prices do not usually initiate inflation. Instead they tend to slow it down. By doing so, however, they extend the duration of inflation and protract the period necessary for its reversal. Administered prices are a problem because they act to prolong inflation once it gets started and because they delay the success and exacerbate the adverse side effects of anti-inflationary stabilization policy. Just as in an upward price spiral sticky administered prices retard the spread of inflation, so when the spiral is unwinding they delay its deceleration and impede the return to price stability.

The Expectations-Formation Lag As noted earlier, the complicating effects of the so-called price-adjustment lag are reinforced by a comparable lag in price expectations, which tend to adjust slowly to the actual rate of inflation. Why does this expectations lag occur and what determines its length?

The expectations lag occurs because people have imperfect foresight and cannot predict the future with certainty. If the future were known with absolute certainty, there would be no forecasting errors and no expectations lag. Anticipations would adapt themselves instantaneously to realized outcomes, and the expected rate of inflation would always be the one that actually occurred. Although the future is emphatically uncertain, people nevertheless try to make informed guesses about it, often on the basis of analyses of the past. Thus one longstanding explanation of the lag holds that price predictions are based on perceived trends as estimated from past price experience perhaps modified by current information.

---

[5] This simplistic cost-push view is implausible and at odds with the orthodox theory of monopoly behavior. According to the latter view, a monopolist sets a relative price for his product that maximizes profits in real terms and maintains that real price by adjusting his nominal price to allow for inflation. The logical implication is that given the degree of monopoly power monopolists would have no incentive to raise prices other than to keep pace or catch up with inflation. With real prices already established at profit-maximizing levels, any further upward adjustment would reduce profits. On the other

hand, if prices are being raised to exploit hitherto unexploited monopoly potential, the question arises as to why those gains were foregone in the past. In either case, rising real prices imply non-rational behavior. It is true that rising real prices would be consistent with profit-maximizing behavior if the degree of monopoly power were increasing. But there is little empirical evidence that monopoly power is on the rise.

Because they reflect the past, these trends and the predictions derived from them change slowly over time. Transitory deviations from the trend have little impact on expectations. Thus if the current rate of inflation departs from the recent trend, the expected rate will remain unchanged for a time. It requires the cumulative influence of a sustained change, or at least a very large change, in the actual rate to produce a change in the expected rate. And even then the adjustment will not be instantaneous or complete. The expected rate will continue to lag behind the current rate. As time passes, however, and the new rate persists, it will eventually begin to dominate the trend. The price experience of the more distant past gradually will be forgotten and the expected rate will finally converge on the actual rate.

The length of time required for this convergence cannot be specified with any degree of precision. It depends on what interval or intervals of the past people consider in formulating their expectations and, in particular, on what relative weight they assign to the price experience of the more distant past. In general, the greater the weight assigned to the distant past, the longer will be the lag. The lag may indeed be quite long because people, in formulating their expectations, may look not merely at a chronological succession of prior years but rather at the relevant phases of a succession of past inflation-recession cycles and at the public policy response to these episodes. Looking back at a succession of what might be regarded as stop-go policies, for example, people may expect the recession phase to be brief and to be followed immediately by an expansionary phase that experience suggests may bring new inflationary pressures. To use a popular expression, they may "look across the valley."

A second factor that may affect the length of the expectations lag is the *variability* of the rate of inflation about the trend. The greater the variability of the actual rate of inflation the greater is the financial incentive to forecast it more accurately, i.e., to predict the variations, not just the trend. Accordingly, forecasts of the inflation rate will be revised more frequently and will rely more heavily on current and recent information. The whole time frame employed in observing past price behavior and past forecasting errors will contract and shift toward the present. Correspondingly, the expectations lag will shorten.

Time Lags and Demand Management Policies

This article has focused on the role of time lags in the inflationary transmission mechanism. These lags raise several potential problems for demand manage-

ment (i.e., monetary and fiscal) policy. First, if the lags are variable and hard to predict, policymakers may experience difficulty in accurately forecasting when their policy actions will take effect on the rate of inflation. In this case discretionary policymaking becomes a potentially risky undertaking. That is, unpredictability of the lag effects of policy may render the latter destabilizing rather than stabilizing. For this reason, some analysts have advocated the abandonment of discretionary policy in favor of fixed policy rules. Other observers, however, argue that the lags are not so variable and unpredictable as to defeat effective policy forecasting and control. Actually, little is known about the variability of the lags. The evidence is simply not sufficient to settle the issues. Whether rules would be superior to discretion in the conduct of stabilization policy remains an open question.

Additional policy problems arise from the *length* rather than the variability and unpredictability of the lags. As previously mentioned, the price-adjustment and expectations lags extend the time it takes for a policy-induced change in spending to work its way through to the rate of inflation. Moreover, they influence the pattern of response of quantities and prices to inflationary stimuli. Specifically, they cause the quantity response to precede the price response. When spending changes, output and employment adjust first, prices only later. What are the policy implications of such consequences?

First, owing to the slow response of inflation to a spending change, anti-inflationary monetary-fiscal policy can be expected to produce observed results only after a corresponding lag. Quick monetary remedies for inflation are not likely to be found. Moreover, since the initial impact of a change in spending is on output and employment rather than on prices, a move to monetary restraint would almost surely entail a recession or at least a marked retardation in the rate of expansion of the economy. In short, a temporary but protracted period of unemployment, idle capacity, and sluggish growth might have to be endured if restrictive policy were to be successful in permanently lowering the rate of inflation.

Second, due to the difference in timing of the response of output and prices to a spending change, anti-inflationary policy may appear impotent or even perverse. Because inflationary movements tend to subside so slowly, prices may continue to rise long after output and employment have turned down. Thus inflation can persist even in slack markets, a condition variously known as inflationary recession, stagflation, or slumpflation. During such periods,

monetary restraint may be wrongly blamed for causing both the slump and the accompanying inflation. At such times, monetary authorities, anxious to achieve quick and relatively painless results, may be tempted to abandon the policy of restraint as ineffective at best and harmful at worst.

Third, the same asymmetrical pattern of response —output first, prices only much later—may create the dangerous illusion that expansionary policy in the upswing can achieve permanent gains in output and employment at the cost of little additional inflation. If inflation proceeds as described in the foregoing sections, this view may have unfortunate consequences. For it is virtually impossible to peg output and employment above their natural or equilibrium levels without continuously accelerating the rate of inflation. In any case, time lags may well compound the problem of curbing inflation by leading to the undue prolongation of expansionary policy, thus increasing the momentum behind inflation when it finally occurs.

Indeed, these time lags, together with society's commitment to full employment, may bias monetary-fiscal policy toward inflation over the entire policy cycle. The lags cause output and employment to adjust before prices. Consequently, on the upside the desired output results of demand expansion occur before the undesired inflationary results thus encouraging the prolongation of policy stimulus. On the downside, however, the painful effects come first as restrictive policy produces undesired and costly rises in unemployment with little initial effect on the rate of inflation. Not surprisingly, the appearance of painful rather than desired results leads to impatience with restrictive policies and often to their premature reversal. The policies are abandoned before they have a chance to eliminate the inflation generated in the preceding upswing. The net result is that restrictive policies reduce the inflation rate less than expansionary policies raise it and inflation ratchets upwards.

Direct Controls If orthodox demand-management policies cannot curb inflation without causing painful rises in unemployment, then what other means can be utilized? One suggestion is to reimpose direct wage-price controls. Several arguments have been advanced in behalf of controls. The most naive is that controls can constrain the rate of inflation while the authorities pursue demand-expansion policies necessary to insure full employment. This argument assumes that the rate of inflation can be permanently pegged in the face of persistent excess demand. In fact, however, the rate of inflation tends to gravitate to an equilibrium level where excess demand is zero. Controls can delay the adjustment to that level but they cannot stop it. The equilibrium rate of inflation will inevitably be reached either via evasion, i.e., through black markets, or after the controls are lifted. In the final analysis, controls will have no effect on the equilibrium rate of inflation, which is determined by aggregate demand. It logically follows, therefore, that a decrease in aggregate demand, not the application of controls, is the *sine qua non* for the reduction of the inflation rate.

Controls and the Expectations Lag A more sophisticated argument for controls calls for using them as a supplement to monetary and fiscal policy. The prescription here calls for first employing a monetary-fiscal policy sufficient to eliminate excess aggregate demand and then using controls to speed the actual rate of inflation to its new equilibrium level. In this view, a controls program properly coordinated with restrictive demand policy could help short-circuit the painful process of winding-down a stubborn inflation.

The foregoing argument rests on the belief that controls can exert an independent influence on otherwise sticky price expectations. Eradication of inflationary expectations is of course a prerequisite to the elimination of inflation, since the former is a determinant of the latter. The problem is how to dampen expectations. The orthodox method is to create slack (excess supply) in the economy, thus causing the actual rate of inflation to fall below the expected rate, inducing a downward revision of the latter. Owing to the expectations lag, however, this adjustment may be a slow process and a prime example of how sluggish price anticipations can impede the return to price stability. Direct controls are viewed as a means of speeding the expectations adjustment and thus reducing the duration and severity of the economic slowdown necessary to bring inflation to its equilibrium level determined by aggregate demand.

How could controls influence expectations? Two mechanisms have been suggested. First, the mere announcement of controls might alter expectations in the desired direction. Second, by freezing prices or at least severely constraining their rate of rise, controls would cause the gap between the expected and actual rates of inflation to be greater than it would otherwise be. Assuming that the rate at which expectations are revised varies directly with the size of the gap, controls would thus accelerate the downward revision of expectations.

This view, however, probably overestimates the influence of controls. To have anything more than a temporary impact on price expectations, the controls must convince the public that the trend of prices when controls are in force is a reliable indicator of the likely future trend of prices after they are lifted. The public may be hard to convince, especially if the controls have failed to stop inflation in the past. Aside from this point, it is hard to understand why controls should have a stronger impact on expectations than, say, an announced policy of a permanent reduction in the growth rate of the money stock. What counts is not the means by which the government announces its determination to permanently reduce inflation but that those intentions be believed.

Controls and the Price-Adjustment Lag It should be noted, moreover, that even if controls *could* reduce the expectations lag, they would tend to lengthen the price-adjustment lag, and this would impede the return to equilibrium. The attainment of economic equilibrium requires that two conditions be satisfied. First, inflation must be correctly anticipated, i.e., the expected rate of inflation must equal the actual rate. Second, the equilibrium structure of relative prices must be restored. Lagging prices and costs must be allowed to catch up to and reestablish their equilibrium relationships with other prices that have already adjusted to inflation. Controls interrupt this catch-up process by arbitrarily freezing all prices and costs. As a result, catch-up adjustments are postponed until the controls are lifted and account for the observed tendency for prices to rise sharply when controls are terminated. Note that this problem would not arise if controls could distinguish between legitimate catch-up and unwarranted anticipatory price increases. In practice, however, the two are virtually indistinguishable and controls prohibit both. By forcing the postponement of catch-up price increases, controls probably protract the inflationary process and lengthen the interval required for the rate of inflation to reach its equilibrium level. In the interim, the controls-distorted price structure generates inefficiency, resource misallocation, and income redistribution. Clearly, in the context of the explanation presented here, controls offer no solution to the problem of persistent inflation.

Other Proposed Solutions One of the more promising, and yet untried, solutions to the lag problem is indexation, i.e., the inclusion of purchasing power clauses in all contracts. With all contractual prices tied to automatic cost of living escalators, inflation would be transmitted more quickly and evenly throughout all markets, thereby permitting faster restoration of relative price equilibrium. On the downside, indexation could help reduce expectations more quickly, thus shortening the time required to remove inflation. Critics argue that indexation would intensify inflation. This criticism confuses the *level* of inflation with the *speed* with which inflationary impulses are propagated through the economy. There is no reason to believe that escalator clauses by themselves would have any effect on the rate of inflation. The speed of inflation would be increased, to be sure, but the level of inflation would not necessarily be any higher. Another partial solution to the lag problem would be to shorten the length of contracts. For example, trade-union collective bargaining agreements could be renegotiated more frequently as could long-term contracts for raw materials and energy supplies.

A distinction should be made here between reforms designed to increase the speed of response of inflation to changes in aggregate demand and reforms aimed at reducing the equilibrium rate of unemployment, i.e., the unemployment rate that, given the inevitable frictions, rigidities, and imperfections in the economy, is just consistent with zero excess demand and stable (nonaccelerating, nondecelerating) rates of inflation. Indexation falls in the former category while so-called structural reforms fall in the latter. Structural reforms refer to microeconomic policies directed at improving the efficiency of labor and product markets. True, these policies may increase the responsiveness of inflation by eradicating market imperfections that inhibit price flexibility. But their chief purpose is to reduce the equilibrium rate of unemployment at which the demand for and supply of labor are in balance and at which any stable rate of inflation (including a zero rate) is possible. By so doing, structural policies may make it easier in at least two ways to bring inflation under control. First, they may render the equilibrium rate of unemployment and the corresponding zero or other desired steady-state rate of inflation a more acceptable policy option. By bringing equilibrium unemployment down to a socially tolerable level, structural policies reduce the risk that political pressure will be put on the authorities to peg unemployment at even lower levels that can only be maintained by a constantly accelerating rate of inflation. Second, they reduce the severity of the recession necessary to achieve price stability. Since a dampening of inflation requires a temporary rise in unemployment above its equilibrium level, it follows that any policy that reduces the latter also reduces the transitional level of unemployment.

Conclusion The policy implications stemming from the preceding analysis are straightforward. Admit that there are no quick and easy solutions to the inflation problem. Realize that any serious anti-inflation program faces the formidable obstacles of entrenched inflationary expectations and eroded public confidence in the government's ability and determination to fight inflation and to tolerate the resulting economic slack. Comprehend that, obstacles notwithstanding, the costs of removing inflation may be far less than the costs of accepting it as a permanent way of life.[6] On the basis of this cost-benefit analysis establish a permanent target rate of inflation. Rely on demand-management (i.e., monetary and fiscal) policies to bring inflation down to the desired target level. Assuming that level is zero, reduce the rate of monetary expansion and trim budget deficits until the growth rate of total spending (aggregate demand) is equal to the long-term trend growth rate

of real output (aggregate supply). Eschew controls. Recognize that a reduction of inflation necessitates a transitional rise in unemployment above its equilibrium level with the extent and duration of the rise depending upon the speed with which inflation is to be removed. In other words, recognize that quick eradication of inflation requires high excess unemployment for a relatively short period whereas slower eradication requires a lower level of excess unemployment for a prolonged period. Choose the desired path to price stability always realizing that the choice is between higher excess unemployment for a short time or lower excess unemployment for a long time. Acknowledge also that the severity and duration of the recession necessary to eliminate inflation depends upon the responsiveness of wages and prices to deflationary pressure and upon the adjustment speed of price expectations. Use indexation, if possible, to increase the downward flexibility of wages and prices. Similarly, seek to influence expectations by preannouncing the inflation target and by adhering to a path consistent with achieving it. Once inflation is removed, maintain price stability by avoiding expansionary policies that generate excess demand. Balance the budget each year and maintain a money stock growth rate roughly equal to the long-term growth rate of real output. If society is unwilling to tolerate the equilibrium unemployment rate associated with price stability, design structural policies to lower that rate by improving the efficiency of labor and product markets. If society is still unwilling to accept that rate, face up to the fact that price stability may be impossible to achieve.

---

[6] A policy of accommodating permanent inflation is likely to lead to a pattern of accelerating and highly erratic (variable) inflation. Such inflation, by virtue of its variability and unpredictability, imposes substantial costs on the economy. Not only will there be repeated falls in output and employment whenever the inflation rate drops, but the very unpredictability of volatile inflation increases business uncertainty, makes capital investment decisions riskier, diverts energies and skills from industry and production into speculative activities designed to beat the inflation, and reduces the information content of market prices thereby making the price system a less efficient mechanism for coordinating economic activity. All this makes for a wasteful and inefficient use of resources that results in a lower level of output than the economy is capable of producing. The costs of inflation can be measured in terms of this lost output. Similarly, the benefits of removing inflation can be reckoned in terms of the increased output that would result.

APPENDIX

THE TREATMENT OF LAGS IN SIMPLE ANALYTICAL MODELS
OF THE INFLATIONARY TRANSMISSION MECHANISM

The inflationary mechanism and its constituent lags can be summarized in the form of simple analytical models.[1] Economists have long used such models to study how lags affect the speed and duration of inflation. More recently, such models have been employed to estimate the impact of lags on the effectiveness of anti-inflationary policy. These

models specify the chief determinants of the current and expected future rates of inflation. They also specify the lags linking the variables and determining the pattern of their interaction over time. Comprising these models are a *price-adjustment equation* and an *expectations-formation equation*.[2] The price-adjustment equation explains how the current rate of inflation responds to inflationary expectations and to

---

[1] For a thorough review of inflation models see the article by Laidler and Parkin [7] especially pp. 774-81 that describe models similar to the one presented in this appendix.

[2] Some models of the inflationary process contain a third equation that explains how inflation-generating excess aggregate demand is determined.

lagged excess demand—the lag on the latter variable representing the price-adjustment delay. The expectations-formation equation explains how price anticipations are generated and revised in the light of past price experience. As shown below, the latter equation expresses the expected future rate of inflation as an exponentially-declining weighted average of past rates of inflation.

A crude version of this two-equation model is presented in the following paragraphs.[3] It should be strongly emphasized, however, that the model constitutes a severe oversimplification of a complex process and thus should be interpreted with some skepticism. Presented solely as an illustration, the model purposely omits many of the variables and behavioral relationships that a more complex, sophisticated, and realistic model would contain.

The Price-Adjustment Equation Most models of the inflationary mechanism contain an equation that explains how the current rate of inflation is determined, i.e., the rate at which businessmen mark up their prices. One such equation shows the rate of price inflation p varying directly with lagged excess demand x_{-1} and with the expected rate of price increase p^e. The equation is written as follows:

$$(1) \quad p = ax_{-1} + p^e_{-1}$$

where p is the current rate of inflation (expressed as a percentage rate), x_{-1} is excess demand lagged one period, p^e_{-1} is the present period's expected rate of inflation forecast at the end of the preceding period, and a is a coefficient specifying how much each unit of lagged excess demand contributes to the rate of price increase. The excess demand variable x is measured in terms of real output since businessmen initially respond to changes in demand by altering quantity produced. More specifically, excess demand x is represented by the difference between actual and capacity real output. Actual output can exceed capacity output because the latter is defined not as the absolute physical limit or maximum ceiling level of output but rather as the output associated with the economy's normal or standard level of operation.

Equation (1) states that if aggregate demand and supply are equal so that there exists no excess demand (x = zero), then actual price inflation p will just equal expected inflation p^e_{-1}, i.e., businessmen will be raising their prices at the rate at which they expect other businessmen to be raising theirs. If, however, an expansion in demand raises x above zero, businessmen will eventually react to the excess demand by raising prices at a rate in excess of the expected rate of inflation. This price response, however, is not instantaneous. For a while, quantities rather than prices tend to absorb the impact of excess demand as businessmen temporarily expand output and perhaps allow their inventories to be depleted. These quantity changes affect demands for and prices of factor resources and ultimately invoke cost increases that signal the desirability of raising the rate at which prices are marked up. Later, therefore, businessmen respond to the excess demand by raising prices. The same price-adjustment lag operates on the downside. Thus if a subsequent slackening of spending causes excess demand x to become negative (i.e., a situation of excess supply) the actual rate of price increase p will eventually fall below the expected rate p^e_{-1}. The key word here is *eventually* because the lag prevents prices from responding immediately to shifts in demand.

The Price-Response Lag The one-period delay on the excess demand variable symbolizes the tendency for price adjustments to lag behind shifts in demand.[4] This price-adjustment lag is meant to account or the time it takes for demand pressure to work backward through the interindustry structure and for costs to work forward. To summarize, the association of the price-adjustment lag with the excess demand variable x implies that the impact of a shift in demand is initially registered on x. That impact is not immediately transmitted to prices, however. Instead it is transmitted first to quantities and subsequently to costs. Prices do not respond until rising costs induce them to do so.

The Expectations-Formation Equation The second equation of the model is the expectations-formation equation. It is written as follows:

$$(2) \quad p^e = bp + (1{-}b)p^e_{-1}$$

or, alternatively, as

$$(2a) \quad p^e - p^e_{-1} = b(p{-}p^e_{-1}).$$

[3] The model presented here is adapted from similar models developed by Phillip Cagan and David Laidler. See Cagan [2; pp. 94-6] and Laidler [5, 6]. For an elementary description of Laidler's model, together with a diagrammatic illustration of its dynamic properties see Laidler and Parkin [7; pp. 776-8].

[4] More sophisticated models would express the delayed price adjustment as a distributed lag, i.e., a lag spread over a number of time periods.

Equation (2a) states that the change in the expected rate of inflation $p^e - p^e_{-1}$ is proportional to the amount by which the period's actual inflation p deviated from expected inflation as forecast at the end of the preceding period p^e_{-1} with the factor of proportionality b having a value between zero and unity.

Embodied in the equation is a particular theory—the so-called adaptive-expectations or error-learning hypothesis—of how inflationary expectations are formed. According to the error-learning hypothesis, people formulate expectations about the inflation rate, observe the discrepancy between the actual and anticipated rates, and then revise the anticipated rate by some fraction of the error between the actual and anticipated rates. Expectations are revised in proportion to the error associated with the previous level of expectations.

It can also be shown that the adaptive-expectations hypothesis is equivalent to the theory that people formulate price expectations from prior price experience by looking at a geometrically-weighted average of past rates of inflation with the weights diminishing exponentially as time recedes. This alternative interpretation of the adaptive expectations hypothesis is written as follows:

$$(2b) \quad p^e = b \sum_{i=0}^{\infty} (1-b)^i \, p_{-i}.$$

Here Σ is the *summation operator* indicating the mathematical operation of adding a succession or series of terms, in this case the weighted past rates of inflation. The *summation index* i represents each past time period starting with the most recent (i=0) and extending backward to the most distant (i = ∞). The variables p_{-i} are the past rates of inflation, one for each of the i periods stretching backward into time. Attached to each past rate of inflation p_{-i} is a corresponding *weight* that measures the degree of influence that each p_{-i} has on the formation of price expectations p^e. The weights are expressed as $(1-b)^i$, one for each of the i time periods. Since, as mentioned previously, the coefficient of expectations b is a fraction whose magnitude lies between zero and one, it follows that the term (1—b) will also be a fraction. And since any given fraction raised to progressively higher integral powers yields successively smaller numbers, it follows that the weights $(1-b)^i$ must decrease as the exponent i increases, i.e., the weights must diminish the further back in time one looks.

Graphically, the weights are distributed along an exponentially declining curve whose slope reflects the speed of adjustment of expectations. A steep slope represents a short weighting scheme, implying swift adjustment, and conversely for a relatively flat slope. The slope itself is determined by the magnitude of the fraction (1—b). A value of (1—b) close to zero implies that the weights decline rapidly as time recedes, and so future price expectations depend primarily on recent experience. On the other hand, if (1—b) is closer to 1 in value, rates of inflation from the more distant past enter the equation with higher weights, and recent price information is discounted more heavily. Econometricians who have attempted to fit equation (2b) to the statistical data have found the fraction (1—b) to be both significantly greater than zero and less than 1. These findings imply that while people generally assign higher weights to more recent phenomena, these weights do not dominate the cumulative weight of all past price experience. In short, price anticipations continue to reflect past price experience, which explains why the expected future rate of inflation does not adjust instantaneously to the current rate.

To summarize, equation (2b) states that the expected future rate of change of prices is based on a geometrically declining weighted average of past rates of change of prices. The equation therefore constitutes a precise specification of the common-sense notion that expectations are based on past experience with more emphasis given to recent, rather than distant, experience.

The Expectations Lag The preceding discussion clearly implies that the length of the expectations lag can be defined in terms of the coefficient of expectations b. The coefficient b itself measures the speed of adjustment of expectations to experience, i.e., the quickness of response of p^e to realized actual rates of inflation p.

The average length of the expectations lag is the counterpart of the speed of adjustment. This lag is expressed as (1—b)/b. The closer b is to 1, the shorter the lag. In the extreme case where b equals 1, the lag is nonexistent, and the expected rate of inflation adjusts instantaneously to the current rate. Thus when b is set equal to 1 in the expectations-formation equation

$$(2) \quad p^e = bp + (1-b)p^e_{-1},$$

the equation collapses to $p^e = p$, i.e., anticipated and actual inflation are always identical.

On the other hand, the lag will be longer the closer b is to zero. In the extreme case where b

equals zero, the lag is infinitely long, i.e., the expected rate of inflation never changes. This result can be demonstrated by setting b at zero in equation (2), which yields $p^e = p^e_{-1}$, showing that the expected rate of inflation always remains unaltered from the preceding period. In short, if b is zero, the lag is of infinite length and expectations never change regardless of what is happening to the actual rate of inflation.

One possible shortcoming of the adaptive expectations or error-learning model is that it regards the speed of adjustment or coefficient of expectations b as a fixed constant. And since b determines both the length of the expectations lag and the slope of the weighting pattern used to distribute the lag, it follows that the model, by implication, also treats these phenomena as given constants. This treatment is surely too restrictive. Some analysts think, contrary to the model, that the coefficient b is capable of being influenced by outside information and by the behavior of the rate of inflation itself. For example, it has been suggested that the sensitivity of price anticipations is greater and the corresponding adjustment lag shorter for high and volatile rates of inflation than for low and steady rates. Finally, some observers believe that the expectations coefficient can be influenced by government policy. In fact, this idea constitutes one rationale for direct wage and price controls.[5]

The Complete System   Taken together, the price-adjustment and expectations-formation equations summarize the operation of the inflationary transmission mechanism. These two equations explain the mutual determination of actual and expected rates of price increase. They also indicate the iterative interaction process whereby the expected inflation rate influences the actual current rate, which in turn becomes a determinant of next period's expected rate, which feeds back into next period's actual rate, etc. Moreover, the model demonstrates how inflationary expectations operate to lengthen the lagged adjustment of prices to short-run shifts in demand. This latter result is obtained by substituting the expectations-formation equation into the price-

[5] The argument here is that, by directly altering the expectations coefficient, controls could speed the downward adjustment of price expectations necessary for the removal of inflation. Of course controls might also speed the adjustment process simply by constraining the actual rate of inflation below the level that would otherwise occur at a given level of unemployment. In the latter case, controls would take the expectations coefficient as a given constant.

adjustment equation and then solving recursively for p. The resulting "reduced form" expression is:

$$(3) \quad p = ax_{-1} + ab \sum_{i=2}^{\infty} x_{-i}$$

where the second term on the right-hand side of the equation represents the delayed price impact of excess demand attributable to the operation of the expectations lag. Equation (3) states that once price anticipations enter into price-setting behavior, they tend to prolong the inflationary process. They cause current prices to respond not only to last period's excess demand but also to excess demand in the more distant past.

Finally, the model identifies excess demand as the proximate source of inflation. Specifically, the model implies the following causal chain.

(1) Inflation is determined by excess demand and inflationary expectations.

(2) Inflationary expectations are generated by previous inflationary experience and hence by previous excess demand.

(3) Therefore, excess demand—past and present—is the proximate cause of inflation.

The inflation-generating role of excess demand is made explicit in the reduced-form equation (3) where past levels of x constitute the sole independent variables. The model does not explain how excess demand itself is generated. Such an explanation would require an additional equation expressing the relation between excess demand and the independent variables that determine it. At a very minimum, the list of independent variables would include the money stock since excess demand cannot be long sustained without the monetary growth necessary to support it. Note, however, that there is at least one situation in which excess demand would properly be treated as an independent variable and the money stock as a dependent variable. Such would be the case if society were committed to a full-employment objective in excess of the natural or equilibrium level of employment. Here the policymakers would be expected to pursue the target employment rate (or level of excess demand), passively permitting the money stock and the rate of inflation to adjust so as not to inhibit attainment of the full-employment goal. In this case the level of excess demand would enter the system as a datum to determine the size of the money stock. Thus, depending upon the policy regime, excess demand may appear either as an endogenous or an exogenous variable.