A STICKY-PRICE MANIFESTO

Laurence Ball
N. Gregory Mankiw

Working Paper No. 4677

We are grateful to John Leahy, Robert Lucas, Allan Meltzer, David Romer, and Julio Rotemberg for comments on an earlier draft. This paper is part of NBER’s research programs in Economic Fluctuations and Monetary Economics. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.
A STICKY-PRICE MANIFESTO

ABSTRACT

Macroeconomists are divided on the best way to explain short-run economic fluctuations. This paper presents the case for traditional theories based on short-run price stickiness. It discusses the fundamental basis for believing in this class of macroeconomic models. It also discusses recent research on the microeconomic foundations of sticky prices.

Laurence Ball  
Department of Economics  
Johns Hopkins University  
Baltimore, MD 21218  
and NBER

N. Gregory Mankiw  
Department of Economics  
Harvard University  
Cambridge, MA 02138  
and NBER
Introduction

There are two kinds of macroeconomists. One kind believes that price stickiness plays a central role in short-run economic fluctuations. The other kind doesn’t.

Those who believe in sticky prices are part of a long tradition in macroeconomics. This tradition includes prominent twentieth-century economists such as John Maynard Keynes, Milton Friedman, Franco Modigliani, and James Tobin, and it dates back at least to David Hume. The assumption of sticky prices is an essential underpinning of the IS-LM model, which is taught almost universally to undergraduates as the theory of short-run fluctuations. Traditionalists believe that this model contains a large element of truth.

By contrast, those who deny the importance of sticky prices depart radically from traditional macroeconomics. These heretics hold disparate views: some argue that fluctuations arise from technological shocks to competitive economies, while others emphasize non-Walrasian phenomena such as increasing returns and sunspot equilibria. Nonetheless, heretics are united by their rejection of propositions that were considered well-established a generation or more ago. They believe that we mislead our undergraduates when we teach them models with sticky prices and monetary non-neutrality.

A macroeconomist faces no greater decision than whether to be a traditionalist or a heretic. This paper explains why we choose to be traditionalists. We discuss the reasons, both theoretical and empirical, that we believe in models with sticky prices. And we survey recent research in this paradigm, highlighting both the new discoveries and the questions that remain open for future work. Recent research both strengthens the foundations of traditional models
and extends the range of phenomena that these models can explain.

Our dictionary defines "manifesto" as "a public declaration of principles or intentions." This word describes perfectly what we try to do in this paper. Rather than present new theoretical or empirical results, we attempt to set out what we believe and why we believe it. Of course, our goal is to persuade others. We realize that what follows will not convert a confirmed heretic. But we hope that it will persuade readers who have not yet made up their minds.

Why We Believe What We Believe

We are led to be traditionalists by three convictions. First, we believe that shifts in monetary policy often have important effects on real economic activity. Second, based on microeconomic evidence, we believe that sluggish price adjustment is the best explanation for monetary non-neutrality. Finally, we give weight to the long tradition in macroeconomics in which monetary non-neutrality and price stickiness have central roles.

Recent research on sticky prices, which we discuss below, has reinforced our belief in traditional macroeconomics. Yet our most basic traditionalist convictions predate this work. Indeed, these convictions have been the motivation for us and others to pursue research on sticky prices. We therefore begin by discussing each of these convictions in turn.

Money Matters

We believe that monetary policy affects real economic activity. The main reason for our belief is the evidence of history, especially the numerous episodes in which monetary
contractions appear to cause recessions.

In his course on monetary economics given over a decade ago, Stanley Fischer posed the question, "How do we know that money matters?" His answer was, "Friedman and Schwartz, and Paul Volcker." Friedman and Schwartz's 1963 treatise, *A Monetary History of the United States*, identified a number of episodes in which the money supply contracted sharply. Economic activity declined after each of these shocks, and so it is natural to conclude that money has real effects.

Paul Volcker's disinflation, coming almost two decades after Friedman and Schwartz's treatise, is another episode of this sort. Monetary policy tightened in 1979 because Volcker was more committed to the goal of low inflation than was his predecessor, William Miller. It is easy to explain the deep recession that accompanied the disinflation of the early 1980s if one believes that monetary policy affects output.

Today we can add "Romer and Romer" to Fischer's list of reasons to believe in monetary non-neutrality. In important and controversial studies (1989, 1992), Christina Romer and David Romer have extended the work of Friedman and Schwartz. The Romers read through the minutes of the meetings of the Federal Reserve's Open Market Committee and identified seven dates since World War II when the Fed shifted its policy towards reducing the rate of inflation. They show that shortly after each of these dates, the economy experienced a decline in output and employment. Indeed, the seven policy tightenings account for most postwar recessions. The Romers' results suggest not just that money is non-neutral, but that monetary contractions are a major source of U.S. business cycles.

If monetary contractions are followed systematically by contractions in the real
economy, how can heretics maintain that money is neutral? A common argument is that causality runs from output to money, rather than the other way around. We agree that it is difficult to establish the direction of causation. Indeed, the identification problem leads us to put little weight on the many studies that test for monetary non-neutrality through statistical correlations, such as tests of Granger causality. Changes in monetary policy, whether measured by the money stock or interest rates, are most often endogenous: policymakers are responding to past or expected changes in the economy. Thus money-output correlations cannot establish true causality.

The crucial advantage of the "narrative approach" used by Friedman and Schwartz and by Romer and Romer is that a careful reading of history can yield evidence about the direction of causality. In many of the tight-money episodes that these authors identify, it appears that policy is shifting in ways not determined by events in the real economy. For example, the Fed’s unusually passive stance as the money supply and the economy collapsed in the early 1930s is often attributed to Benjamin Strong’s death in October 1928. Similarly, policy changed in 1979 because William Miller chose to resign in that year and his replacement had a more fervent distaste for inflation. This historical fact makes it plausible to interpret the Volcker disinflation as an exogenous event that caused the 1981-82 recession. It seems less likely that causality ran in the other direction -- that Volcker looked ahead to the upcoming recession (over which he had no control) and decided, for some reason, that it was a good time to pursue contractionary policy.

Historical analysis is, of course, intrinsically open to dispute. Different historians can tell different stories about what happened in a given episode. Yet it is noteworthy that there
is no treatise entitled *A Technology-shock History of the United States*. The leading interpretation of U.S. macroeconomic history remains monetary.

There is another kind of historical episode that provides evidence of monetary non-neutrality: changes in exchange-rate regimes. One can unambiguously identify episodes of sharp shifts between fixed and floating exchange rates, such as the breakdown of Bretton Woods and the entry of various countries into the EMS. If money were neutral, such shifts in policy towards nominal exchange rates would not affect the behavior of real exchange rates. In practice, as Mussa (1986) documents, real exchange rates become vastly more volatile when there is a shift from fixed to floating rates, and vastly less volatile when policy shifts in the other direction. Eichengreen shows (1993) that these changes are too large and too sudden to attribute to changes in the economic environment that might have triggered the shifts in policy. Krugman (1993) summarizes the evidence this way: "I personally think that the effort to explain away the apparent real effects of nominal shocks is silly, even if one restricts oneself to domestic evidence. Once one confronts international evidence, however, it becomes an act of almost pathological denial."

We have focussed on the effects of money because they provide a clean test for the existence of nominal imperfections. Nominal variables such as the money stock have no role in the standard theory of general equilibrium, where the unit of account is indeterminate and irrelevant. The evidence that money matters implies that the economy contains an important nominal imperfection, and (as we argue below) sticky prices are the most realistic candidate for such an imperfection.

Once one assumes sticky prices, however, the implications go well beyond the effects
of money. In traditional macroeconomics, this assumption generates an upward-sloping aggregate supply curve, and thus explains the output effects of any shift in aggregate demand. Our traditionalist beliefs are reinforced not only by the recession that accompanied the Volcker disinflation, but also by the boom that accompanied high government spending during the Vietnam War. As we discuss below, sticky prices can also play a central role in explaining the effects of aggregate-supply shocks, such as large changes in oil prices. Purely real models can in principle explain the effects of government spending or oil shocks, but they cannot match the unified theory of real and monetary shocks that follows from sticky prices.

Microeconomic Evidence on Sticky Prices

We believe that price stickiness is the best explanation for monetary non-neutrality. It is natural to consider the macroeconomic effects of sticky prices, because we observe many prices that change infrequently. Both of us go to barbers who keep haircut prices fixed for several years.

There are now many microeconomic studies of the behavior of prices, and the finding of substantial stickiness is universal. In an early study, Stephen Cecchetti (1986) examined the newsstand prices of magazines. He found that the typical magazine allows inflation to erode its real price by about 25 percent before it raises its nominal price. When inflation is 4 percent per year, the typical magazine changes its price about every six years.

Dennis Carlton (1986) examined a very different data set: the Stigler-Kindahl data on transactions prices among firms buying from other firms. He concludes: "The degree of price rigidity in many industries is significant. It is not unusual in some industries for prices to
individual buyers to remain unchanged for several years."

The most comprehensive study is the recent one by Alan Blinder (1991), who has interviewed managers in a large, representative sample of U.S. firms. One of his questions is how often the firms change their prices. He finds that 37.7 percent of firms change their prices once a year, and another 17.4 percent change their prices less than once a year. The median firm in the economy changes its prices about once a year.

It is of course true that many prices in the economy are quite flexible. Blinder finds that 10.1 percent of prices are adjusted more than once a month. The most extreme cases are the prices of commodities traded on organized exchanges, which change every few minutes. We live in a world in which some prices are sticky, and some are flexible. Such a hybrid world, however, is more likely to be described accurately by a fixed-price model than by a flexible-price model. Both empirical evidence and theories of "real rigidities" suggest that firms' desired relative prices are not very sensitive to economic fluctuations (Blanchard and Fischer, 1989, Ch. 10). Since flexible-price firms desire fairly constant relative prices, they do not adjust their nominal prices substantially when others do not adjust. Thus flexible-price firms inherit sluggish adjustment from fixed-price firms (Haltiwanger and Waldman, 1989; Bonomo, 1992).

In addition, for the purpose of explaining monetary non-neutrality, not all prices are equally important. According to traditional theory, money has real effects because the price level does not adjust to equilibrate the supply and demand for money. For this theory, the most important prices are for those goods bought with money, since the prices of goods bought with credit do not directly affect the demand for money. Goods bought with money tend to
be small retail items, such as newspapers and haircuts. Experience suggests that these are the goods for which prices are most sticky.

**The Long Tradition**

The view that monetary policy has potent effects is not questioned by the general population. Policymakers and the press clearly believe that policy can speed up or slow down real economic activity. The chairman of the Federal Reserve is sometimes called the second most powerful person in the United States. In 1987 a book made the best seller's list with the title, "Secrets of the Temple: How the Federal Reserve Runs the Country." The heretical belief in short-run monetary neutrality has never been taken seriously outside of the ivory tower.

Moreover, it is only recently that academics have taken this view seriously. Heretics often go under the label "new classical," suggesting that their views harken back to a more enlightened era before the Keynesian revolution. Yet the term is a misnomer. The classical economists themselves never suggested that money was neutral in the short run. Here is what David Hume said in his 1752 essay "Of Money":

> In my opinion, it is only in the interval or intermediate situation, between the acquisition of money and the rise in prices, that the increasing quantity of gold or silver is favourable to industry.... The farmer or gardener, finding that their commodities are taken off, apply themselves with alacrity to the raising of more.... It is easy to trace the money in its progress through the whole commonwealth; where we shall find that it must first quicken the diligence of every individual, before it increases the price of labour.
A key insight of the classical economists, which today we take for granted, is that money is neutral in the long run. In claiming that money is also neutral in the short run, today's heretics take classical economics more seriously than did the classical economists themselves.

In the 1960s, economists engaged in heated debate over the best way to view economic fluctuations. See, for example, the exchanges in Milton Friedman's Monetary Framework: A Debate with His Critics (1974). Yet no one in these debates questioned that money affects output because prices adjust gradually. Indeed, it is ironic that Milton Friedman is sometimes viewed as the intellectual grandfather of today's heretics. In fact, he was the economist who argued most forcefully that monetary policy is a frequent cause of economic fluctuations, and he never doubted that wages and prices adjust gradually. Although traditionalists are often called "new Keynesians," this label is also a misnomer. They could just as easily be called "new monetarists." (We regret our contributions to this terminological confusion.)

Why does it matter that monetary non-neutrality and price stickiness are part of a long tradition in economics? Scientific truths, unlike legal decisions, are not determined by appeals to authority. There was once a long tradition asserting that the sun revolves around the earth, but that tradition did not stop the heliocentric solar system from becoming the reigning paradigm.

The answer, we believe, is Occam's razor. Occam's razor is the philosophical premise that if several competing theories are consistent with the facts, the simplest one is probably right. The heretics would have us believe that the Fed is an institution with little power. Yet, somehow, the entire economics profession before 1980, and the world outside the ivory tower still today, have been misled into thinking the Fed is a powerful force controlling the economy.
Occam's razor suggests that such a contorted theory should be viewed with skepticism. It is noteworthy that Occam's razor is sometimes used implicitly to argue in favor of heretical theories, such as real business cycle theory. Real business cycle theory takes a standard model of economic growth—the Ramsey model—and applies it to study economic fluctuations. Indeed, the primary appeal of early real business cycle theory was its parsimony. Yet, over time, real business cycle theory has found itself deviating more and more from standard growth models. Real business cycle theorists now emphasize non-convexities and home production, for example. As more epicycles are added, the theory loses the virtue of parsimony.

In our view, Occam's razor dictates that one should not dismiss a long tradition of thought without compelling reasons. Certainly, it is possible that the tradition is wrong, and that the heretical view of monetary neutrality and price flexibility will turn out to be true. But, just as certainly, Occam's razor gives the heretics the burden of proof.

Challenges to Our Basic Convictions

In the previous section we described our most fundamental beliefs about macroeconomics. We recognize that recent work has challenged some of these beliefs. Here we evaluate three of the most important challenges.

The Cyclical Behavior of Prices

Traditional macroeconomists believe that shifts in aggregate demand generate procyclical movements in prices. That is, booms tend to raise prices, and recessions tend to
lower them. Several recent empirical studies call this prediction into question (Kydland and Prescott, 1990; Cooley and Ohanian, 1991). These studies argue that the price level is countercyclical in the postwar United States. They reach this conclusion by detrending the price level and real output using the Hodrick-Prescott filter, and then showing that the correlation between the detrended series is negative. Economists such as Barro (1993) interpret these results as evidence against traditional models and in favor of real business cycle models, in which productivity shocks generate countercyclical price movements.

We are not persuaded by this evidence for two reasons. First, even traditional macroeconomists believe that supply shocks, such as the oil shocks of the 1970s, are important in some historical periods and can move prices and output in opposite directions. Second, we believe that the statistical methodology in recent studies is misleading. This point is made by Chadha and Prasad (1993), who perform stochastic simulations of a traditional model. The shocks in the model are shifts in aggregate demand, and they affect output because nominal prices adjust slowly. Nonetheless, when the simulated data are detrended, the model yields a negative correlation between the price level and output. Thus recent empirical results are consistent with traditional models.

To gain some intuition for these issues, consider an economy in which inflation falls from ten percent per year to zero because of a large, sudden change in monetary policy. In this economy, the price level rises quickly until the new policy is implemented and then remains constant. A fitted trend smooths out the kink in the price level, and therefore lies below the price level at the time of stabilization; the detrended price level is highest at this point. According to traditional theories, the disinflation will temporarily decrease output. Thus
the detrended price level will appear countercyclical. This reasoning explains the finding that
the HP-filtered price level rises during recessions that are accompanied by disinflation, such
as the Volcker episode (Barro, 1993).

We are more persuaded by other approaches to examining the comovement of output
and prices. Jo Anna Gray and David Spencer (1990) estimate a structural aggregate supply
equation, carefully controlling for supply shocks and changes in the natural rate of
unemployment. They find that unexpected changes in the price level are positively associated
with output.

Ball (1994b) follows another approach, which imposes less theoretical structure. He
identifies 28 episodes in OECD countries in which an economy experienced a large, sustained
reduction in inflation. In 27 of the 28 cases, output fell below trend during the disinflation.
Traditional macroeconomics predicts exactly this pattern: the tight monetary policy needed to
reduce inflation reduces output as well.

Other Explanations of Monetary Non-neutrality

We believe that sticky prices provide the most natural explanation of monetary non-
neutrality since so many prices are, in fact, sticky. Other economists, however, accept
monetary non-neutrality but resist the assumption of sticky prices. They have been led to
develop models of non-neutrality with flexible prices.

As a matter of logic, any model of monetary non-neutrality must include some nominal
imperfection. The developers of flexible-price models replace nominal price rigidity with some
other nominal imperfection. In our view, these alternative imperfections are less plausible
empirically than sticky prices.

The most famous alternative to sticky-price models is the Lucas model (1972, 1973). In Lucas's framework, the nominal imperfection is informational: agents do not know the price level and thus cannot distinguish movements in nominal and relative prices. We have mixed feelings about this model. As many authors point out, the assumption of an unobservable price level appears implausible, at least for economies such as the United States where reliable price data are released monthly. Thus the Lucas model is not a convincing substitute for sticky-price models. On the other hand, Lucas's broad theme -- that imperfections in information help explain monetary non-neutrality -- is appealing. As we discuss below, incomplete information can explain why firms set prices for fixed intervals of time, as often assumed in traditional models. Interpreted broadly, Lucas's approach is complementary with theories of monetary non-neutrality based on sticky prices.

A popular current alternative to sticky-price models for explaining non-neutrality is models of "liquidity effects" developed by Lucas (1990), Fuerst (1992), and Christiano and Eichenbaum (1991, 1992). In these models, the key imperfection is that some agents set their nominal money holdings in advance and cannot adjust immediately if the price level changes. By itself, such an assumption is reasonable: empirical studies of money demand find that nominal balances adjust slowly towards the optimal level. We do not believe, however, that the assumption provides a convincing explanation for the effects of money.

Most macroeconomic models with flexible prices dichotomize. Real output and the real interest rate are determined by the goods market and the supply side of the economy, and the money market determines only the price level. Thus, under flexible prices, imperfections in
the money market are irrelevant to the behavior of output. To break this classical dichotomy, liquidity-effect models introduce a channel through which money directly affects aggregate spending: a cash-in-advance constraint that sets consumption equal to real balances. With this constraint, slow adjustment of money holdings to a change in prices implies a change in real balances and hence a change in consumption. In IS-LM terms, a monetary shock influences output because money enters the IS curve.

We doubt that this channel of monetary transmission is empirically important. In reality, consumers do not face a period-by-period cash-in-advance constraint: the ratio of money balances to consumption need not be constant. If a person's real balances are low because of sticky nominal balances, he can adjust by increasing the velocity of money -- by visiting the bank more often -- rather than reducing his consumption. We suspect that this is the more relevant margin of adjustment. Certainly most empirical researchers on consumption do not take seriously the idea that a person's money holdings are an important determinant of his consumption, given his overall levels of income and wealth.

Credit Policy vs. Monetary Policy

At this point we admit a nagging doubt about our basic convictions. In our discussion of monetary non-neutrality, we assumed that monetary authorities directly control only nominal variables. In reality, the Federal Reserve engages in activities in addition to controlling the stock of money. As emphasized by Plosser (1991) and Romer and Romer (1993), the Fed sometimes intervenes in credit markets by changing bank regulations, applying informal pressure for reduced lending, or restricting consumer credit. There is evidence that such credit
actions influence output; for example, the 1980 credit controls appear to have contributed to the recession of that year.

By itself, such evidence does not cast doubt on traditional models. A restriction in credit is likely to cause a shift in aggregate demand. Thus, one can explain the effects of credit actions through traditional channels involving sticky prices.

Yet the existence of these credit actions casts doubt on some of the evidence of monetary non-neutrality. Credit actions often coincide with shifts in conventional monetary policy as the Fed uses various approaches to shift aggregate demand. And, by disrupting financial intermediation, credit actions can reduce aggregate supply as well as aggregate demand. In our view, the best case for monetary neutrality centers on the supply-side effects of credit actions. It is logically possible that real disruptions of intermediation account for the output losses in episodes of "tight policy," giving the illusion of monetary non-neutrality.

We are skeptical that direct credit actions can explain the apparent effects of money in all episodes. At this point, however, there is little hard evidence on the relative importance of credit actions and traditional monetary policy, or on whether credit actions mainly affect aggregate demand or aggregate supply. We hope that future research will fill these gaps.

**New Theories of Sticky Prices**

Despite the merits of the traditional approach to economic fluctuations, it went through a rough period during the 1970s and 1980s. Everyone knows the story of the "new classical" revolution, in which Robert Lucas and his followers convinced economists that there were irreparable flaws in traditional macroeconomics. The most compelling argument was that
traditional models were incompatible with microeconomics. Traditional models simply assumed the crucial feature of nominal rigidity, even though agents could gain by eliminating it. For example, the models of Fischer (1977) and Taylor (1980) assumed that firms and workers sign long-term nominal contracts even though both sides would benefit from indexation. Similarly, Brunner, Cukierman, and Meltzer's (1983) model of persistence in output fluctuations assumed that prices are fixed in advance, even though firms could raise profits by adjusting to current shocks. In Lucas's famous quip, traditional models assumed that people left $500 bills on the sidewalk.

The new classical attack convinced many researchers that they should abandon traditional macroeconomics and start the field afresh. But die-hard traditionalists such as ourselves were not convinced. The incompatibility of traditional models with optimizing behavior was a serious problem, but not a fatal one. Over the past decade, many economists have sought to support this view by developing sticky-price models based on firm microeconomic foundations. Often this research goes under the label "new Keynesian." Here we summarize the outcome of this effort. (For more detailed surveys, see Ball, Mankiw, and Romer [1988], Rotemberg [1987], and Romer [1993]).

Static Models of Nominal Rigidities

The first step was to build simple, static models to explain nominal rigidity. In our view, this work is largely complete. The explanation for nominal rigidity rests on three foundations: imperfect competition, small "menu" costs of nominal price adjustment, and "real rigidities."

16
In retrospect, it is obvious that imperfect competition must be part of any coherent theory of price stickiness. Under perfect competition, firms are price takers, not price setters. Only under imperfect competition can we ask whether a firm will choose to keep its price fixed or to set a new price in response to a shock.

It is also essential to posit a menu cost or some similar friction in nominal price adjustment. As discussed below, many kinds of price rigidities arising from efficiency wages, customer markets, implicit contracts, and so on can help explain economic fluctuations. But these are not enough, because they are rigidities in real wages and prices: absent irrational money illusion, workers and customers care only about real variables. Adjustment to a monetary shock requires changes in nominal variables but not in real variables, and so real rigidities do not imply monetary non-neutrality. As a matter of logic, nominal stickiness requires a cost of nominal adjustment.

Recent research on sticky prices began with an insight of Mankiw (1985) and Akerlof and Yellen (1985): imperfect competition and menu costs are not only separate ingredients of a sticky-price model, but are highly complementary. Critics of sticky-price models point out that menu costs in the real world are small: it costs something for firms to print menus and replace price tags, but not very much. How then can menu costs generate price rigidity that in turn generates recessions with large social costs? The answer is that imperfect competition creates a wedge between the social and private gains from price adjustment. If a firm fails to reduce its nominal price when the money stock falls, its loss of profits may be small -- too small to warrant paying the menu cost. Nonetheless, with imperfect competition, the social costs of rigidity may be large.
More concretely, the social costs of price rigidity are likely to exceed the private costs because imperfect competition creates aggregate demand externalities. This point is formalized by Blanchard and Kiyotaki (1987) and Ball and Romer (1989), who use the Dixit-Stiglitz model of monopolistic competition. If the money stock falls and prices do not adjust, then, in these models, the lower real money stock reduces total spending in the economy. When aggregate spending falls, the demand curve facing each firm shifts inward -- a firm sells less at any given price. Consequently, the firm's profits fall.

In this setting, the private and social gains from price adjustment are very different. If a single firm adjusts its price, it does not change the position of its demand curve; it simply moves to a new point on the curve. This adjustment raises profits, but the gain is second order. In contrast, if all firms adjusted to the monetary shock, the aggregate price level would fall, real balances would return to their original level, and each firm's demand curve would shift back out. The gains in profits would be large: a firm gains more from an outward shift of its demand curve than from a movement along the curve. Unfortunately, an individual firm does not take this effect into account because, as a small part of the economy, it takes aggregate spending and hence the position of its demand curve as given. Thus firms may not bother to make price adjustments that, taken together, would end a recession.

Blanchard and Kiyotaki and Ball and Romer show that, because of aggregate demand externalities, costly economic fluctuations can in principle arise from arbitrarily small menu costs. In their models, however, this result arises only for extreme parameter values -- in particular, both labor demand and labor supply curves must be very flat. With plausible amounts of curvature in production and utility functions, small menu costs cannot generate
substantial nominal rigidity. The reason is that the private costs of rigidity, although smaller than the social costs, are large for non-negligible changes in money. Firms and their workers would gain substantially from adjusting prices and thereby dampening fluctuations in employment and output, even if other firms' prices are rigid. Thus small menu costs do not deter them from adjusting.

These results establish the need for the third foundation of sticky-price models: real rigidities. There are many plausible theories of why relative prices and real wages are insensitive to shifts in demand. (Our favorites include efficiency wage models, Okun's (1982) customer-market theory, and Woglom's (1982) model of kinked product demand based on imperfect information.) As discussed above, real rigidities alone do not yield nominal rigidity. But, Ball and Romer (1990) show, real rigidities magnify the nominal rigidities arising from menu costs. The reason is that real rigidities reduce the private cost of nominal rigidity. If a firm wishes to maintain a stable relative price, and if other nominal prices are rigid, then the firm desires at most a small nominal adjustment when the money supply falls. The private cost of forgoing this small adjustment is likely to be smaller than a modest menu cost. Thus adding real rigidities to imperfect competition and menu costs helps to explain why firms fail to make the price adjustments that would neutralize a monetary shock.

A final feature of recent models is important for answering a common heretical criticism. The argument is that sticky prices are irrelevant in practice because there is little correspondence between the sectors of the economy with sticky prices and those that are most sensitive to monetary shocks. Ahmed (1987), for example, demonstrates the lack of correlation across industries between nominal wage rigidity (measured by the extent of indexation) and
employment variability. Car prices may be more flexible than haircut prices because they are set by bargaining; yet the auto industry is more cyclically sensitive than the haircutting industry. These facts are, however, consistent with current sticky-price models because stickiness affects the economy through aggregate demand externalities. The effects of monetary shocks on aggregate spending, and hence on demand in a given sector, depend on aggregate nominal rigidity, not rigidity in that sector. It may be the barbers' fault that the aggregate price level adjusts sluggishly, but this does not help the car salesmen when demand drops in a recession. Differences in the cyclicality of industries arise from other factors; for example, autos are hit hard by tight money because spending on them is sensitive to interest rates, whereas spending on haircuts is not.

Making the Models Dynamic

In our view, the ideas discussed above add up to a complete theory of nominal rigidities in a static setting. Yet real economies are not static. A basic tenet of traditional macroeconomics is that money is non-neutral in the short run—a period of a few years—but neutral in the long run. Can dynamic models with costly price adjustment generate the time-series responses of prices and output to monetary shocks that we observe in actual economies? In contrast to static theory, there is considerable disagreement among sticky-price researchers about the right approach to dynamics.

Before discussing the difficulties in this area, we should emphasize a basic insight about dynamics: the role of staggered price adjustment in generating price-level inertia. Sticky-price models, like all macro models, face the challenge of explaining the persistence of output
fluctuations. When Paul Volcker tightened monetary policy in October 1979, why was output depressed until 1984? The particular problem for sticky-price models is that the period for which downturns last exceeds the period for which most individual prices are sticky. If most prices are adjusted within a year, as Blinder finds, why doesn't the long-run of monetary neutrality arrive within a year? The answer, first presented by Taylor (1979) and Blanchard (1983), is that different firms adjust prices at different times. With synchronized adjustment, all firms would adjust fully to a monetary shock as soon as their next adjustment date arrived. With staggering, however, some group of firms must go first in adjusting prices, which means accepting a lower relative price. With strong real rigidity, no firm is willing to accept a large relative cut; instead, different firms take turns making small adjustments, and it takes many rounds of adjustment for the aggregate price level to adjust fully. Thus, full adjustment to a monetary shock can take much longer than the period for which each price is fixed.

Beyond this qualitative insight, can we solve dynamic models of price adjustment? The answer depends on what simplifying assumptions we make. The crucial modelling choice is between "time-contingent" and "state-contingent" adjustment. Under time-contingent adjustment, a firm adjusts prices at intervals of a fixed length (which can be chosen optimally given the cost of adjustment). Under state-contingent adjustment, a firm adjusts whenever the state of the economy warrants it; usually, it is optimal to follow an "Ss" rule in which the firm's relative price is adjusted to a base level whenever it hits some bounds.

The behavior of economies with time-contingent adjustment is now well-understood. Taylor and Blanchard use time-contingent models to formalize the result that staggering produces inertia. More generally, researchers have derived the dynamic response of the
economy to a monetary shock under reasonable conditions. Typically, an innovation in nominal money produces smooth adjustment of the aggregate price level over time, with the speed depending inversely on the degree of real rigidity. The effect on real output is greatest when the innovation occurs, declines over time, and dies out asymptotically.

There is no simple summary of results for state-contingent models, because they have proven much more difficult to solve. Starting with Caplin and Spulber (1985), impressive progress has been made in analyzing state-contingent models; current state-of-the-art models include Caballero and Engel (1992) and Caplin and Leahy (1991a). Unfortunately, strong restrictions are still needed for tractability. In most models, for example, real rigidity is ruled out: a firm's desired price depends entirely on the money stock rather than on other firms' prices. (Caplin and Leahy (1991b) is an exception.) Thus the models do not include the key source of price inertia under staggered adjustment.

In any case, the bottom line of state-contingent models is similar to that of time-contingent models: the price level does not adjust immediately to a monetary shock (except in very special cases), and so money is not neutral. State-contingent models go beyond time-contingent models in generating subtle non-linearities. In Caplin and Leahy (1991a), for example, a positive monetary shock is less likely to raise output if it is preceded by other positive shocks. The empirical relevance of such non-linearities is unclear.

State-contingent models have recently been more popular than time-contingent models. The apparent reason is that state-contingent adjustment is optimal for a price setter who faces a fixed adjustment cost. In this setting, it is arbitrary to assume that the firm adjusts at fixed intervals, regardless of whether adjustment is warranted by changed circumstances. In our
view, however, the emphasis on exclusively state-contingent models is misplaced. As Caballero (1989) shows, time-contingent adjustment is optimal if the main cost of adjustment is gathering information about the state rather than making the actual adjustment. State-contingent models assume that firms continuously monitor the environment to determine when to adjust. Yet it may be less expensive to gather information at fixed intervals. If so, and if the pure adjustment cost is small, firms naturally adjust prices at these fixed intervals.

As an empirical matter, time-contingent adjustment is common. Nearly all wages are adjusted on a fixed time schedule. The same is true for many output prices; for example, many firms issue catalogs on a regular schedule. The use of time-contingent rules also explains the fact that firms often make small price changes (Kashyap, 1987). A small change can be optimal when the time for adjustment arrives, whereas small changes never happen with state-contingent adjustment.

In addition, even if many firms make state-contingent adjustments, the behavior of the aggregate price level can be close to the pure time-contingent case as long as there is some time-contingent adjustment. This point is made by Bonomo (1992), who studies an economy with both kinds of adjustment. The result follows from the idea, discussed above, that flexible parts of the economy inherit rigidities from the rigid part of the economy when firms care about relative prices. In this case, the relevant rigidity is the fixed timing of price adjustment. The fact that time-contingent firms cannot adjust immediately to a large shock means that state-contingent firms do not want to adjust either.

Research on dynamic price adjustment is far from complete. In particular, the literature has produced some surprising results that warrant further investigation. Caplin and Spulber
show that monetary neutrality can arise even with sticky prices if firms follow one-sided Ss rules. Ball (1994a) shows that if firms follow time-contingent rules, then a fully credible slowdown in money growth can cause an output boom. These theoretical results are important, even though they have little direct empirical relevance. The Caplin-Spulber result shows that explanations for non-neutrality must go beyond the simplest state-contingent models, for example by introducing idiosyncratic shocks that cause prices to fall as well as rise. The Ball result suggests that theories must include credibility problems as well as frictions in price adjustment. Research on these issues is continuing.

What Are Menu Costs?

The new work on sticky prices has established that menu costs—the underlying nominal friction producing non-neutrality—can be of trivial size relative to the macroeconomic effects of non-neutrality. To be sure, they must be strictly positive: there must be some cost to adjusting prices. The resistance to sticky-price models in some quarters appears to arise from an unwillingness to allow any cost of nominal adjustment. Is the idea of menu costs plausible?

Certainly "menu costs" should be interpreted more broadly than the physical costs of changing price tags. In this sense, the term may be unfortunate. "Menu costs" are a metaphor like "shoe leather costs". Some critics have suggested testing menu-cost models by seeing whether rigidities arise in circumstances of higher technological costs of price changes—for example, when price tags must be replaced by hand rather than electronically. This is a bit like testing whether consumers dislike inflation more in countries with less durable footwear.

We suspect that the most important costs of price adjustment are the time and attention
required of managers to gather the relevant information and make and implement decisions. The fact that many wages and prices are adjusted at fixed intervals of time suggests that information-gathering costs are important. At a common-sense level, it seems obvious that it is more convenient for busy managers to decide on new prices once a year rather than once a day. This fact leads to infrequent adjustment if (as suggested by recent research) the private costs of non-adjustment are small.

In any case, a literal account of menu costs is not necessary for studying most issues in macroeconomics. Menu costs should be viewed as a parable—a convenient formalization that captures the facts that prices are not adjusted continuously, and that they tend to adjust more quickly to large than to small shocks. In resting our theories on a parable, we have an irreproachable precedent: price theory under perfect competition, which rests on the parable of the Walrasian auctioneer. Walras observed that prices move to equilibrate supply and demand, and he captured this tendency with the parable of an auctioneer. Similarly, macroeconomists have noted that many prices are sticky in the short run, and they capture this fact with the parable of menu costs. It is no more appropriate to insist on an exact identification of menu costs than it is to demand the social security number of the Walrasian auctioneer.

Of course, it is still interesting to go beyond the parable to better understand the foundations of nominal frictions. Future research could examine information gathering and processing costs in actual firms, for example. The analogue in competitive economics is research that examines the foundations of the auctioneer story, such as the literature on the convergence of Nash games to perfect competition. Note that microeconomists are rarely reproached for studying competitive models despite the incompleteness of this research on
foundations. Similarly, macroeconomists can use menu-cost models without a complete, literal account of menu costs.

A New View of Aggregate Supply

The previous section describes recent progress in developing models of sticky prices, and the challenges that remain. We now ask what the payoff is to all this work. One accomplishment is to put the assumption of price rigidity on firmer foundations. We can now write down models with sticky prices and teach them to undergraduates without the guilty feeling that we are doing violence to microeconomics. But beyond providing microeconomic foundations, have we learned anything about macroeconomics?

Early menu-cost models were sometimes criticized for failing to generate new empirical predictions, or to explain previously mysterious phenomena (e.g. Summers, 1988). The models were designed to produce monetary non-neutrality, and that is all they did. In our view, the last five years of research have made this criticism obsolete: an important new branch of macroeconomics has grown out of models with costly price adjustment. In particular, the models lead to a new theory of the supply side of the economy in the short run, a theory with several novel (and apparently correct) empirical predictions. The simplest way to summarize the new theory is in terms of the undergraduate model of aggregate demand and aggregate supply. Sticky-price models provide new answers to several old questions about the short-run aggregate supply curve:
Why does the aggregate supply curve slope up?

Equivalently, why do shifts in monetary policy or other determinants of aggregate demand affect real output, rather than prices alone? This question provided the original motivation for research on sticky prices. As described above, the answer rests on a combination of imperfect competition, menu costs, and real rigidities.

Why does the slope of the aggregate supply curve differ across countries and time periods?

One benefit of deriving rather than assuming nominal price rigidity is that one can discuss why the degree of rigidity might vary. The new research has shown that a key determinant of the degree of rigidity—and thus the slope of the aggregate supply curve—is the level of trend inflation (Ball, Mankiw, and Romer 1988). High inflation leads to more frequent price changes for given adjustment costs, and more frequent adjustment increases the responsiveness of prices to monetary shocks: the aggregate supply curve becomes steeper. Ball, Mankiw, and Romer report strong empirical support for this prediction: average inflation explains much of the variation in the slope of the aggregate supply curve across countries and over different eras in a given country. In addition, DeFina (1991) reports evidence of a higher-frequency time-series relation between trend inflation and the slope of aggregate supply. The effects of trend inflation are quantitatively important: the Ball-Mankiw-Romer estimates imply that reducing inflation from ten to five percent more than doubles the output effect of a monetary shock. Thus, for example, the output-inflation tradeoff facing U.S. policymakers is considerably less favorable today than at the outset of the Volcker disinflation. And attempts to achieve zero inflation, as in Canada today, are likely to be very costly.
Why does the aggregate supply curve shift?

Traditional macroeconomics fell into disfavor in the 1970s in part because of its inability to explain the stagflation arising from supply shocks. These shocks were apparent movements in the aggregate supply curve, rather than movements along it resulting from changes in aggregate demand. At one level, this problem was solved when researchers added "supply shifters" such as world oil prices to empirical Phillips curves. At a deeper level, shifts in aggregate supply have remained puzzling. A shift in aggregate supply means a change in the aggregate price level corresponding to given output. As a theoretical matter, it is not clear why the aggregate price level should be influenced by changes in relative prices, such as an increase in oil prices arising from OPEC collusion. Indeed, classical theory draws a sharp dichotomy between relative prices, which depend on real microeconomic factors, and the overall price level, which depends on the supply and demand for money.

A new explanation for shifts in aggregate supply, consistent with traditional macroeconomics, is developed by Ball and Mankiw (1992). The classical argument that relative prices are unrelated to the aggregate price level implicitly assumes flexibility of nominal prices. By contrast, we assume that it is costly to adjust prices. The key implication is that firms adjust prices in response to large shocks, but that it is not worth adjusting to small shocks. Consequently, large shocks have disproportionately large effects on actual price adjustment.

In this setting, the overall price level depends on the distribution of shocks to desired relative prices. To see this point, consider an example in which the frictionless relative price of oil rises by 50%. By the definition of "relative," other relative prices must go down to
balance this increase. Yet it is not usual for some other sector to experience a 50% relative decrease when oil prices rise. Instead, the relative decreases are spread across the non-oil economy, with many equilibrium prices falling by small amounts. Thus the distribution of desired price changes is asymmetric. With full price flexibility, the increases in the oil sector and decreases in other sectors average to zero. But with menu costs, and hence disproportionate adjustment to large shocks, the asymmetry in shocks has aggregate effects. The actual increases in oil prices are larger than the decreases in other sectors, and the aggregate price level rises. Thus the model can explain an increase in the price level for given aggregate demand.

Ball and Mankiw show that this idea has wide applicability. Empirically, we measure the relative incidence of large positive and large negative shocks with the skewness of the distribution of price changes. If the distribution is skewed to the right, for example, there are large positive shocks, which tend to raise the price level. We find that movements in the skewness of price changes explain a large fraction of U.S. supply shocks, or shifts in the short-run Phillips curve, both in the OPEC era since the 1970s and in earlier periods. Indeed, our measures of skewness outperform traditional supply shifters such as food and energy: they fit better in sample, display greater subsample stability, and better explain various historical episodes.

Why might the aggregate supply curve be non-linear?

One theme in traditional thinking is that the slope of aggregate supply is likely to be different for increases and decreases in aggregate demand. Tobin (1972) and others argue that
decreases in demand have large output effects, whereas increases in demand trigger larger price responses and therefore have smaller output effects. Typically, traditionalists obtain this result simply by assuming that prices are more sticky downwards than upwards.

Recently, sticky-price theorists have shown that an asymmetric aggregate supply curve arises endogenously under natural conditions. Different versions of the argument appear in Tsiddon (1991), Ball and Mankiw (1994), and Caballero and Engel (1992). In all cases, the crucial assumption is positive trend inflation. In an environment with trend inflation, firms' relative prices fall automatically between nominal adjustments. In this setting, a firm does not need to make a special adjustment if a negative shock reduces its desired relative price: inflation automatically does the work. In contrast, a positive shock means that the firm's desired relative price is rising while its actual price is falling, creating a large gap between desired and actual prices. Thus a positive shock triggers quick nominal adjustment, whereas prices are sticky in response to substantial negative shocks — exactly the asymmetry that Tobin and others assume.

A number of empirical studies present evidence that monetary shocks indeed have asymmetric effects on real output. See Cover (1992), DeLong and Summers (1988), and Morgan (1993). This finding is yet another empirical result that is explained by new research on sticky prices.

Conclusion

A scientific theory should be judged not only by the intrinsic appeal of its assumptions, but also by its ability to explain observed facts—especially ones that it was not explicitly
designed to explain. Menu-cost models were introduced to explain the first feature of the aggregate supply curve—its upward slope—but Mankiw and Akerlof-Yellen were not thinking of asymmetries or cross-country differences in aggregate supply, and certainly not the relation between inflation and the skewness of price changes. As menu-cost models have developed, they have yielded a unified explanation for many of the empirical features of aggregate supply.
REFERENCES

Ahmed, S.

Akerlof, G. A., and Yellen, J. L.

Ball, L.


Ball, L., and Mankiw, N. G.


Ball, L., Mankiw, N. G., and Romer, D.

Ball, L., and Romer, D.


Barro, R. J.

Blanchard, O. J.

32
Blanchard, O.J., and Fischer, S.

Blanchard, O.J., and Kiyotaki, N.

Blinder, A. S.

Bonomo, M.

Brunner, K., Cukierman, A., and Meltzer, A. H.

Caballero, R.

Caballero, R., and Engel, E. M. R. A.

Caplin, A. S., and Leahy, J.

(1991b) Aggregation and Optimization with State-Dependent Pricing. Columbia University and Harvard University.

Caplin, A. S., and Spulber, D. F.

Carlton, D.
Cecchetti, S. G.

Chadha, B., and Prasad, E.

Christiano, L. J., and Eichenbaum, M.

and

Cooley, T., and Ohanian, L.

Cover, J. P.

DeFina, R. H.

DeLong, J. B., and Summers, L. H.

Eichengreen, B.

Friedman, M., and Schwartz, A. J.

Fuerst, T. S.
Gray, J., and Spencer, D. E.

Haltiwanger, J., and Waldman, M.

Hume, D.
(1752) Of Money.

Kashyap, A. K.
(1987) Sticky Prices: New Evidence from Retail Catalogs. MIT.

Krugman, P.
(1993) What Do We Need to Know About the International Monetary System? Princeton Essays in International Finance, #190.

Kydland, F. E., and Prescott, E. C.

Lucas, R. E., Jr.


Mankiw, N. G.

Morgan, D.

Okun, A. M.
Plosser, C. I.  

Romer, D.  

Romer, C. D., and Romer, D.  

———  

Romer, C. D., and Romer, D.  

Rotemberg, J.  

Summers, L. H.  

Taylor, J. B.  

Tsiddon, D.  

Tobin, J.  

Woglom, G.  