

The History of the Phillips Curve: Consensus and Bifurcation

By ROBERT J. GORDON

Northwestern University, NBER and CEPR

Final version received 15 December 2008.

While the early history of the Phillips curve up to 1975 is well known, less well understood is the post-1975 fork in the road. The left fork developed a theory of policy responses to supply shocks in the context of price stickiness in the non-shocked sector. Its econometric implementation interacts shocks with backward-looking inertia. The right fork approach emphasizes forward-looking expectations that can jump in response to anticipated policy changes. The left fork approach is better suited to explaining the postwar US inflation process, while the right fork approach is essential for understanding behaviour in economies with unstable macroeconomic environments.

INTRODUCTION

The history of the Phillips curve (PC) has evolved in two phases, before and after 1975, with a widespread consensus about the pre-1975 evolution, which is well understood. Bifurcation begins in 1975, when the PC literature split down two forks of the road, with little communication or interaction between the two forks. The major contribution of this paper, and hence the source of ‘bifurcation’ in its subtitle, is to examine, contrast and test the contributions of the two post-1975 forks.

The pre-1975 history is straightforward and is covered in Section I. The initial discovery of the negative inflation–unemployment relation by Phillips, popularized by Samuelson and Solow, was followed by a brief period in which policy-makers assumed that they could exploit the trade-off to reduce unemployment at a small cost of additional inflation. Then the natural rate revolution of Friedman, Phelps and Lucas overturned the policy-exploitable trade-off in favour of long-run monetary neutrality. Those who had implemented the econometric version of the trade-off PC in the 1960s reeled in disbelief when Sargent demonstrated the logical failure of their test of neutrality, and finally were condemned to the ‘wreckage’ of Keynesian economics by Lucas and Sargent following the twist of the inflation–unemployment correlation from negative in the 1960s to positive in the 1970s. The architects of neutrality and the opponents of the Keynesian trade-off emerged triumphant, with two major *caveats* that their own models based on information barriers were unconvincing, and that their core result, that business cycles were driven by monetary or price surprises, floundered without supporting evidence.

After 1975 the evolution of the PC literature split in two directions, each of which has largely failed to recognize the other’s contributions. Section II reviews the ‘left fork of the road’, the revival of the PC trade-off in a coherent and integrated dynamic aggregate supply and demand framework that emerged in the late 1970s in econometric tests, in theoretical contributions, and in intermediate macro textbooks. This approach, which I have called ‘mainstream’, is resolutely Keynesian, because the inflation rate is dominated by persistence and inertia in the form of long lags on past inflation. An important difference between the mainstream approach and other post-1975 developments is that the role of past inflation is not limited to the formation of expectations, but also includes a pure persistence effect due to fixed-duration wage and price contracts, and lags between

changes in crude materials and final product prices. Inflation is dislodged from its past inertial values by demand and supply shocks.

The econometric implementation of this approach is sometimes called the ‘triangle’ model, reflecting its three-cornered dependence on demand, supply and inertia. Demand is proxied by the unemployment or output gap, and explicit supply shock variables include changes in the relative prices of food, energy and imports, changes in the trend growth of productivity, and the effect of Nixon-era price controls. The triangle approach explains the twin peaks of inflation and unemployment in the 1970s and early 1980s as the result of supply shocks, and provides a symmetric analysis of the ‘valley’ of low inflation and unemployment in the late 1990s. It emphasizes that inflation and unemployment can be either positively or negatively correlated, depending on the source of the shocks, the policy response and the length of lagged responses.

The right fork in the road is represented by models in which expectations are not anchored in backward-looking behaviour but can jump in response to current and anticipated changes in policy. Reviewed in Section III, important elements in this second literature include policy credibility, models of the game played by policy-makers and private agents forming expectations, and the new Keynesian Phillips curve (NKPC), which derives a forward-looking PC from alternative theories of price stickiness. The common feature of these theories is the absence of inertia, the exclusion of any explicit treatment of supply shock variables, the ability of expected inflation to jump in response to new information, and alternative barriers to accurate expectation formation due to such frictions as ‘rational inattention’.

Which post-1975 approach is right? Models in which expectations can jump in response to policy are essential to understanding Sargent’s (1982) ends of four big inflations and other relatively rapid inflations in nations with a history of monetary instability, e.g. Argentina. But the mainstream/triangle approach is unambiguously the right econometric framework in which to understand the evolution of postwar US inflation, and the NKPC alternative has been an empirical failure as it has been applied to US data.

Section IV develops and tests the triangle econometric specification alongside one recently published version of the NKPC approach. The latter can be shown to be nested in the former model and to differ by excluding particular variables and lags, and these differences are all rejected by tests of exclusion restrictions. The triangle model outperforms the NKPC variant by orders of magnitude, not only in standard goodness-of-fit statistics, but also in post-sample dynamic simulations.

The scope of this paper is limited to the American theoretical and empirical literature, with the exception of Phillips’ (1958) article itself. There are three main interrelated themes in this paper that have not previously received enough attention. First, two quite legitimate responses occurred after 1975 to the chaotic state of the PC. Second, each response is important and helps us to understand how inflation behaves, albeit in different environments. Third, the two approaches need to pay more attention to each other, and this paper represents a start toward that reconciliation.

I. CHANGING INTERPRETATIONS OF THE PHILLIPS CURVE, 1958–75

We begin by reviewing the evolution of the PC from Phillips’ 1958 article through the development of the Friedman and Phelps natural rate hypothesis and Lucas’ introduction of rational expectations. Beyond the scope of this paper are developments before 1958, in particular the many references ably surveyed by Humphrey (1991) dating

back to Hume in the mid-eighteenth century regarding the long-run neutrality and short-run non-neutrality of money. The only exception to the 1958 starting cut-off in this paper is Fisher's 1926 article, which anticipates Phillips' relation, albeit interpreting it with the reverse direction of causation.

The Phillips curve is born: Phillips and Samuelson–Solow

The acceptance of new ideas and doctrines is often facilitated if they help to elucidate an outstanding empirical puzzle. Thus the acceptance in the late 1960s of Friedman's natural rate hypothesis occurred rapidly, because it helped to explain the ongoing acceleration of the US inflation rate far beyond the rate forecast by previous research. Likewise, the acceptance of the negative PC a decade earlier was almost immediate, since the PC appeared to resolve an ongoing puzzle about the interpretation of American inflation in the 1950s.

Implicit in pre-Phillips views of US inflation was a 'reverse L' aggregate supply curve, with the joint of the reverse 'L' at a level of economic activity often called 'full employment'. Sustained increases of 'demand-pull' inflation would occur when the economy was operating at a higher level of activity than full employment. But below full employment the inflation rate would be near zero or, at very low levels of activity, even negative as occurred between 1929 and 1933. The early history of the postwar era was reassuring, in that during the recession of 1949 the inflation rate was negative (-2.0% at an annual rate for the GDP deflator between 1948(IV) and 1950(I)). Then inflation returned during the low-unemployment Korean War years 1950–53 to an extent that had to be suppressed by price controls.

Doubts emerged beginning with the failure of the inflation rate to decline for a single quarter during the 1953–54 recession, followed by its inexorable rise during 1955–57, 'despite growing overcapacity, slack labor markets, slow real growth, and no apparent great buoyancy in over-all demand' (Samuelson and Solow 1960, p. 177). No consensus emerged on the right combination of demand-pull with alternative supply-driven explanations, variously named 'cost-push', 'wage-push' and 'demand-shift'. Into this fractured intellectual atmosphere, the remarkable Phillips (1958) article replaced discontinuous and qualitative descriptions by a quantitative hypothesis based on an unusually long history of evidence. Since 1861 there had been a regular negative relationship in Britain between the unemployment rate and the growth rate of the nominal wage rate. By implication, since the inflation rate would be expected to equal the growth rate of wages minus the long-term growth rate of productivity, there was a regular negative relationship between the unemployment rate and the inflation rate.

Before examining the data, Phillips makes two important theoretical observations. First, the negative relationship between the unemployment rate and the rate of nominal wage change should be 'highly non-linear' due to downward wage rigidity that reflects in turn the reluctance of workers 'to offer their services at less than the prevailing rates when the demand for labor is low and unemployment is high' (1958, p. 283). Second, the rate of change of wages may depend not just on the level of unemployment but also on its rate of change, and subsequently we will discuss the role of this 'rate of change' effect in the context of US postwar models and of the interpretation of the Great Depression.

However, Phillips surprisingly debunks a third possible correlation, that between the rate of change of wages and the retail inflation rate ('working through cost of living adjustments'). He was thinking of a world in which wage rates represented four-fifths of factor costs and import prices the other one-fifth, and normally wage rates and import prices would rise at the same rate. Only when import prices rise five times as fast as

productivity growth would retail prices influence wage rates. An interesting note is that Phillips was already thinking of a world in which demand shocks (the level and change of unemployment) and supply shocks (the rate of change of import prices relative to final goods prices) both mattered in determining wage and price changes. However, the role of supply shocks was not fully integrated into PC analysis until the late 1970s.

Most of Phillips' article consists of a set of 11 graphs displaying the rate of change of the nominal wage on the vertical axis and the unemployment rate on the horizontal axis. Graphs are shown for the major sub-periods (1861–1913, 1913–48 and 1948–57) and for each business cycle within the first sub-period. The accompanying text provides an explanation for each point that lies off the fitted regression line, which for 1861–1913 is

$$(1) \quad w_t = -0.90 + 9.64U_t^{-1.39},$$

where, as in the rest of this paper, upper-case letters are levels, lower-case letters are rates of change, w_t is the rate of change of the nominal wage rate, and U_t is the unemployment rate. Points above the line are identified as years of declining unemployment or rapidly rising import prices, and vice versa. Note the nonlinear formulation and the fact that neither the rate of change effect nor the import price effect is explicitly incorporated into the equation. An econometric representation that included both the level and rate of change effect was soon provided by Lipsey (1960).

Recall that equation (1) is estimated for data only from 1861–1913, and the remaining post-1913 data are plotted against this curve in order to locate episodes when the actual data lie away from the curve. The change in wage rates is remarkably close to the prediction of the 1861–1913 curve except for the two years 1951–52, which were influenced by rapid increases in import prices in 1950–51 resulting from the 1949 devaluation of sterling.

Phillips concludes by translating the fitted curve for wage change into an unemployment–inflation relationship by subtracting long-term productivity growth; it appears that stable prices require an unemployment rate of roughly 2.5%. Notably, Phillips does not conjecture about circumstances in which the apparently stable 1861–1913 curve might shift up or down in the long run. Also, Phillips does not mention policy implications at all, and this provides the setting in which Samuelson and Solow (1960) christen the relationship as the ‘Phillips’ curve and explore its policy implications.

So widely read and discussed was the Samuelson–Solow article that the term ‘PC’ entered the language of macroeconomics almost immediately and soon became a lynchpin of the large-scale macroeconomic models which were the focus of research activity in the 1960s. Much of the Samuelson–Solow article provides a critique of the pre-Phillips hypotheses and the difficulty of identifying them.

Then, turning to the Phillips evidence, Samuelson and Solow lament the absence of a similar study for the USA and extract some observations from a scatter plot of US data. First, the US relationship does not work for the 1930s and the two world wars. Second, the implied zero-inflation rate of unemployment is about 3% for the remaining prewar years, similar to Phillips' estimate of 2.5%. Third, there is a clear upward shift in the relationship from the prewar years to the 1950s, and the zero-inflation unemployment rate for the 1950s had risen from 3% to ‘5 to 6 percent’.

They struggle to explain the postwar upward shift by invoking powerful trade unions that are less ‘responsible’ than their UK counterparts, and/or the expectation of permanent full employment in the USA. Another conjecture is that the compact size of the UK compared to the USA makes labour markets in the former more flexible. One

policy conclusion is that anything that makes US labour markets more flexible will help to shift the PC downwards.

Samuelson and Solow have rightly been criticized for posing a long-run inflation–unemployment trade-off available for exploitation by policy-makers. As the authors conclude: ‘We rather expect that the tug of war of politics will end us up in the next few years somewhere in between their selected points. We shall probably have some price rise and some excess unemployment’ (Samuelson and Solow 1960, p. 193).

While Samuelson and Solow conclude by warning that the PC relationship could shift over the longer run, their example involves a ‘low-pressure’ (i.e. high-unemployment) economy in which expectations of low inflation could shift the PC down or could aggravate structural unemployment, thus shifting the PC up. They regard either outcome as possible and notably fail to reason through the long-run implications of a high-pressure economy with its implications of a steady increase in inflation expectations and an associated steady upward shift in the PC. That inference had to wait another eight years for the contributions of Friedman and Phelps.

An interesting side issue is the antecedent of Phillips’ article published by Irving Fisher in an obscure journal in 1926, reprinted and brought to a wider audience in 1973.¹ Recall that Samuelson and Solow lament the availability of a detailed statistical study of the USA analogous to Phillips’ UK research, yet Fisher had already provided such research more than 30 years earlier.² A notable difference with Phillips is that Fisher reverses the direction of causation, so that changes in the rate of inflation cause changes in the level of the unemployment rate. Fisher explains the mechanism in modern textbook terms—because costs of production (including interest, rent, salaries and wages) are fixed in the short run ‘by contract or by custom’, a faster rate of inflation raises business profits and provides an incentive to raise output. ‘Employment is then stimulated—for a time at least’ (1973 version, p. 498). Because of the lag of costs behind prices, Fisher emphasizes that the relationship is between unemployment and the inflation rate, not the price level, and that the price level has ‘nothing to do with employment’. He uses the analogy of driving, in which it takes more fuel per mile to climb a hill than descend it, but exactly the same amount to navigate a ‘high plateau as on the lowlands’.

Fisher’s statistical study is limited to monthly data for the years 1915–25. When the influence of inflation is represented by a short distributed lag over five months, the correlation coefficient is 90% between the unemployment rate and the short distributed lag of inflation. An important weakness of the Fisher study is evident in his Chart II but is not discussed by the author. The 90% correlation applies to 1915–25 but his chart extends back to 1903. During the period 1903–15, unemployment is almost as volatile as during 1915–25 but the variance of inflation is much lower, implying that the relationship is not stable and that Fisher’s main result may be picking up the special features of the First World War and its aftermath (just as Phillips’ UK correlation is strong during the First World War).

Aspects of Phillips curve economics in the 1960s

During the early to mid-1960s, at least three aspects of the PC emerged that would have subsequent consequences. First, the PC trade-off appeared to provide policy-makers with a menu of options. The policy advisors of the Kennedy and Johnson administrations, led by Walter Heller with support roles by Robert Solow and James Tobin, argued that the previous Republican administration had chosen a point too far south-east along the PC trade-off, and that it was time to ‘get the country moving again’ by moving to the north-

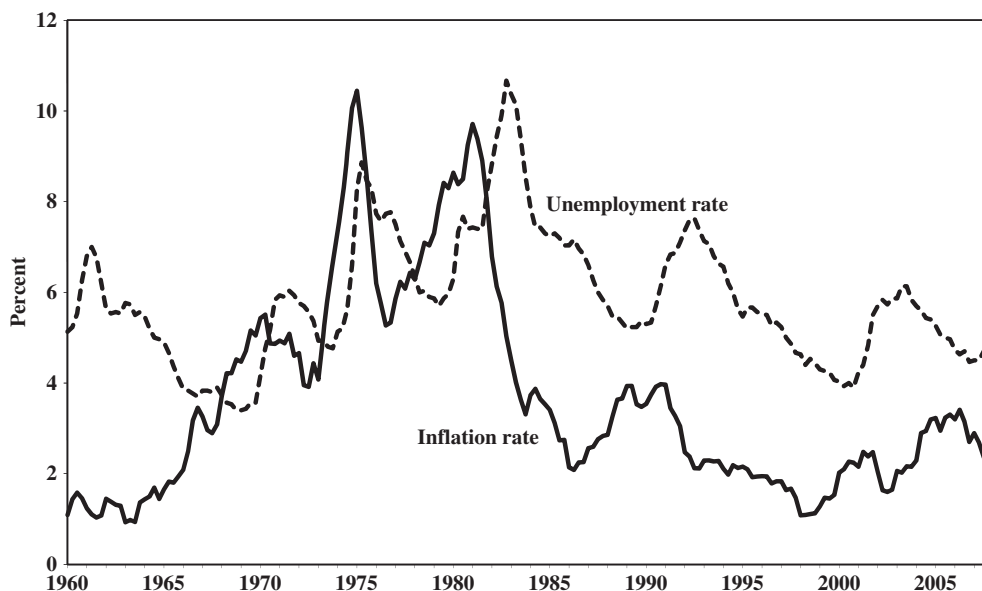


FIGURE 1. The unemployment and inflation rates, quarterly data, 1960–2007. (Source: US Bureau of Labor Statistics (www.bls.gov) and US Bureau of Economic Analysis (www.bea.gov)).

west. Heller's group convinced President Kennedy to recommend major cuts in Federal income taxes, and these were implemented after his death by the Johnson administration in two phases during 1964 and 1965. However, in late 1963 the economy was already operating at an unemployment rate of 5.5% that Samuelson and Solow had calculated was consistent with zero inflation, and so the expansionary Kennedy–Johnson fiscal policy would have implied an acceleration of inflation even without the further loosening of the fiscal floodgates due to the Vietnam War.

Figure 1 plots the US inflation and unemployment rates in quarterly data since 1960, and we shall refer to it here to examine the period 1960–71 and then return to the same graph below to link the evolution of PC debates to the post-1971 behaviour of inflation and unemployment.³ The unemployment rate fell below 5.5% in 1964 and remained below 4% between 1966 and 1970. The sharp acceleration of inflation from less than 2% in 1963 to 5.5% in 1970 is consistent with current econometric estimates of the 1963 natural rate of unemployment (the rate that is consistent with steady inflation rather than zero inflation) in the range of 5.5% to 6.0% (see Figure 5 below).

A second aspect of this period was the development of mainframe electronic computers that made it practical for the first time to specify and estimate large-scale econometric models (a book-length policy analysis using the Brookings model is contained in Fromm and Taubman 1968). The specification of the inflation process in these models always consisted of at least two equations. The PC was embodied in an equation for the rate of change of the nominal wage in which the main explanatory variables were the unemployment rate, sometimes its rate of change, some measure of expected inflation based on a backward-looking set of lags, and perhaps various tax rates.

Then the estimated change in wages was typically translated into the inflation rate in an equation that related the price *level* to the wage *level* adjusted for the level of trend productivity, the so-called 'trend unit labour cost'. The price–labour cost ratio or 'mark-

up' was allowed to respond to a measure of demand, usually not the unemployment rate but rather a measure more directly related to the product market, such as the ratio of unfilled orders to shipments. The reduced form of this approach implied that the inflation rate depended on the level and rate of change of unemployment, perhaps other measures of demand, and lagged inflation. We return below to the problems encountered by these models in confronting the data of the late 1960s and in dealing with the challenge of the Friedman–Phelps natural rate hypothesis.

A third, albeit peripheral, feature of this era was the rivalry between the economics departments at the University of Chicago and MIT in general, and between Milton Friedman and Franco Modigliani in particular. In 1965 more than 100 pages in the *American Economic Review* were devoted to a debate between them and their co-authors over the issue of whether 'only monetary policy mattered' or 'only fiscal policy mattered', a debate that seemed bizarre when the consensus view based on the IS-LM model showed that both monetary and fiscal policy mattered except in certain extreme cases.⁴

The natural rate revolution

Prior to the publication of the Friedman and Phelps articles, theoretical questions had been raised about the PC framework. Why did the nominal wage adjust slowly, particularly in a downward direction, and what determined the speed with which it responded to inadequate demand? Why did the PC lie so far to the right, that is, why did nominal wages rise so fast at a low unemployment rate, and why was such a high unemployment rate required to maintain zero inflation? Perhaps most relevant in anticipation of Friedman and Phelps, how could the PC be so stable over history when there were so many episodes of hyperinflations fuelled by permissive monetary and fiscal policy? I have always thought that the development of the natural rate hypothesis at Chicago, rather than at Harvard or MIT, reflected the deep involvement of several Chicago economists as advisers to several countries in Latin America, where the lack of correlation between inflation and unemployment was obvious.

Friedman's (1968) Presidential Address contained two sections that each had a main point, closely interrelated. First, the central bank could not control the nominal interest rate if that implied faster inflation, because the implied reduction in the real interest rate would add fuel to the inflationary fire. The second section was most important for the PC debate, his then-startling conclusion that policy-makers had no ability to choose any unemployment rate in the long run other than the natural rate of unemployment, the rate that would be 'ground out' by the microeconomic structure of labour and product markets. A more practical interpretation of the natural rate was the unemployment rate consistent with accurate inflation expectations, which implied a steady rate of inflation.

Conventional analysis based on a policy trade-off ignored the adjustment of expectations. Consider an economy operating at the natural rate of unemployment and with an initial inflation rate of 1% that was accurately anticipated. Any policy-maker attempting to reduce the actual unemployment rate below the natural rate would move the economy north-west along the short-run PC, pushing the unemployment rate lower but the actual inflation rate higher. Once agents notice that the actual inflation rate is higher than the initially anticipated rate of 1%, expectations will adjust upward and shift the entire short-run PC higher. This process will continue until the unemployment rate rises back to the natural rate of unemployment.

The timing of Friedman's address was impeccable and even uncanny. The Kennedy–Johnson fiscal expansion, including both the tax cuts and Vietnam War spending,

accompanied by monetary accommodation, had pushed the unemployment rate down from 5.5% to 3.5%, and each year between 1963 and 1969 the inflation rate accelerated, just as Friedman's verbal model would have predicted. The acceleration of inflation bewildered the large-scale econometricians, who had previously estimated a 'full employment' unemployment rate of 4% and whose forecasts of inflation had been exceeded by the actual outcome year after year.

Well aware of their own failure to forecast the late 1960s acceleration of US inflation, Friedman's detractors attacked the verbal model that Friedman used to motivate the natural rate hypothesis. In what later became known as the 'fooling' model, Friedman postulates employers with expectations of the price level that are always accurate, but workers with an expected price level that does not respond until after a substantial lag to a higher actual price level. In a business expansion, firms raise the wage but raise the price level by more, thus reducing the real wage as needed to provide the incentive to hire additional workers. But workers see the higher nominal wage and interpret it as a higher actual real wage, because they fail to adjust their expectation of the price level. Friedman's model was attacked as grossly implausible, because workers have access to monthly announcements of the Consumer Price Index (CPI) and indeed observe actual prices as they shop almost every day. In Friedman's world, there could be no business cycle.

Phelps (1967, 1968) is credited with co-discovering the natural rate hypothesis. In contrast to Friedman's distinction between smart firms and dumb workers, in Phelps's world everyone is dumb, i.e. equally fooled. Both firms and workers see the price rise in their industry and produce more, not realizing that the general price level has risen in the rest of the economy. Phelps developed one model in which workers are isolated from information about the rest of the economy. Normally there is frictional unemployment, as workers regularly quit one firm to go look for more highly paid work at other firms. But in a situation in which their own firm raises the wage, they stay with that firm instead of quitting. Thus the unemployment rate decreases even though, without their knowledge, all other firms in the economy have raised the wage by the same amount at the same time. The workers are fooled into a reduction in frictional unemployment, and the macroeconomic data register a decline in the unemployment rate. Hence there is a short-term correlation between the rate of wage change and the unemployment rate, but this lasts only as long as expectations are incorrect.

Whether firms or workers or both are fooled, the criticisms directed against the Friedman fooling model apply to Phelps as well. Workers and their employers buy many goods and services frequently; they obtain news on the CPI every month; and perhaps most important if periods of high real GDP and low unemployment had always been accompanied by an increase in the aggregate price level, workers and firms learn from these past episodes and use their experience to form expectations accurately.

Rational expectations and the 'policy ineffectiveness proposition'

Both the Friedman and Phelps models were based on the twin assumptions of continuous market clearing and imperfect information. Soon thereafter, in two influential articles, Lucas (1972, 1973) extended their model by adding a third component: rational expectations. Workers and firms use their knowledge of past history to work out the implications of an observed fall or rise in wages on the overall wage level. Rational expectations imply that erroneous expectations errors are not repeated.

Lucas collapsed the distinction between firms and workers and treated all economic agents as 'yeoman farmers' who face both idiosyncratic shocks to their own *relative* price

and macro shocks caused by fluctuations in monetary growth and other factors. The agents use rational expectations to deduce from past history how much of an observed change in the local price represents an idiosyncratic shock and how much represents a macro shock. When local price shocks have a high correlation with macro shocks, agents do not adjust production, knowing that no change in relative prices has occurred. Lucas used this insight to explain why the PC in a country like Argentina with high macro volatility would be much steeper than in a country like the US with low macro volatility.

The concept of rational expectations led Lucas and his followers to make a startling prediction. He argued that *anticipated monetary policy cannot change real GDP in a regular or predictable way*, a result soon known as the ‘policy ineffectiveness proposition’. In common with Friedman and Phelps, the Lucas approach implied that movements of output away from the natural level require a price surprise, so that the central bank can alter output not by carrying out a predictable change in monetary policy but only by creating a surprise. (The formal development of the proposition was carried out in Sargent and Wallace 1975.)

By the end of the 1970s the Lucas approach was widely criticized. The problem was not Lucas’ introduction of rational expectations, but rather the twin assumptions inherited from Friedman and Phelps, namely continuous market clearing and imperfect information. Deviations of the current actual price level from the expected price were the *only* allowable source of business cycle movements in real GDP. Thus, despite the widespread appeal of the Friedman–Phelps–Lucas approach, it ran aground on the shoals of an inadequate theory of business cycles. With monthly information available on the aggregate price level, the business cycle could last no more than one month.⁵ In the recent evaluation of Sims (2008, p. 4), the microeconomic underpinnings of the Lucas supply curve were ‘highly abstract and unrealistic—for example models of “island economies” in which people had to infer the value of the economy-wide interest rate or money stock from the price level on their own island’. Even Lucas later confessed that: ‘Monetary shocks just aren’t that important. That’s the view I’ve been driven to. There’s no question that’s a retreat in my views’ (Cassidy 1996, p. 53).

Rejection of the empirical case against monetary neutrality

Whatever the model used to explain the business cycle, the natural rate hypothesis and long-run monetary neutrality are intact if empirical coefficients imply that a reduction of the unemployment rate below a certain level (whether it is called the natural rate or the full employment rate) leads to continuously accelerating inflation. In the first few years after the Friedman and Phelps articles, those who had developed econometric models supporting a permanent long-run trade-off claimed that the validity of long-run neutrality could be tested by estimating whether the sum of coefficients on the lagged dependent variable in an inflation equation was equal to unity or was significantly below unity. Here we ignore the distinction between wage and price changes and examine the relationship between the inflation rate (p_t), its lagged value (p_{t-1}), and the unemployment rate (U_t):

$$(2) \quad p_t = \alpha p_{t-1} + \beta U_t + e_t.$$

Here the response of inflation to unemployment is negative ($\beta < 1$). If the sum of coefficients on lagged inflation is significantly below unity, then in the long run when $p_t = p_{t-1}$ there is a long-run trade-off between inflation and unemployment:

$$(3) \quad p_t = \beta U_t / (1 - \alpha).$$

Numerous research papers written in the late 1960s and early 1970s placed major emphasis on the finding that the α coefficient was significantly below unity, implying a permanent trade-off as in equation (3). However, these results were ephemeral and quickly abandoned for two reasons. First, as the sample period extended over more of the period of accelerating wage and price change in the late 1960s, the α coefficient kept creeping up and by 1972 had reached unity, particularly when the coefficient was allowed to vary over time.⁶

The second and more important reason to abandon this test of the long-run trade-off was Sargent's simple but devastating econometric point. Here we simplify Sargent's exposition by suppressing the difference between wages and prices, and by making expected inflation depend on only a single lag of inflation rather than a distributed lag. The original specification is not (2) but rather

$$(4) \quad p_t = \alpha E p_t + \beta U_t + e_t,$$

where $E p_t$ is the expected rate of inflation. An observable proxy for expected inflation must be obtained, and this requirement is satisfied by backward-looking or adaptive expectations:

$$(5) \quad E p_t = \nu p_{t-1}.$$

When (5) is substituted into (4), we obtain

$$(6) \quad p_t = \alpha \nu p_{t-1} + \beta U_t + e_t.$$

Now Sargent's point becomes clear: the single equation (6) cannot be used to estimate both α and ν . The only way that α can be interpreted as the coefficient on expected inflation is for an extraneous assumption to be introduced, in particular that $\nu = 1$.

Yet, Sargent argues, there is no reason for ν to be unity, and rather if the inflation rate can be approximated as a covariance-stationary stochastic process, ν must be less than unity. For ν to be unity, the inflation rate would display extremely strong serial correlation or 'drift', but during much of US history before 1950 the inflation rate displayed relatively little serial correlation. Thus it is quite possible that α was equal to unity throughout the postwar era but that ν gradually increased with the serial correlation of inflation in the 1960s that was higher than in the 1940s and 1950s. In short, Sargent made a convincing case that the previous econometric estimates of α in the context of equation (2) had no relevance to the validity of the natural rate hypothesis. Not surprisingly, such econometric exercises ceased quite abruptly after 1972.

Sargent's observation that the ν coefficient should be smaller in periods with less serial correlation of the inflation rate was subsequently validated. Gordon developed quarterly data back to 1892 and showed that the sum of coefficients on lagged inflation rose from 0.40 in 1892–1929 to 0.60 in 1929–53 and then to 1.06 in 1954–80 (Gordon 1982a, Table 3). We return to his results below, because they directly address the shifting form of the PC relationship during the two world wars and during the Great Depression that was originally noticed by Samuelson and Solow (1960).⁷

II. THE POST-1975 LEFT FORK IN THE ROAD: THE DYNAMIC DEMAND–SUPPLY MODEL WITH INERTIA

The 1960s were the glory years of the PC's interpretation as a negative correlation between inflation and unemployment, initially as incorporating a permanent negative trade-off, and subsequently as a significant short-run trade-off subject to the longer-run

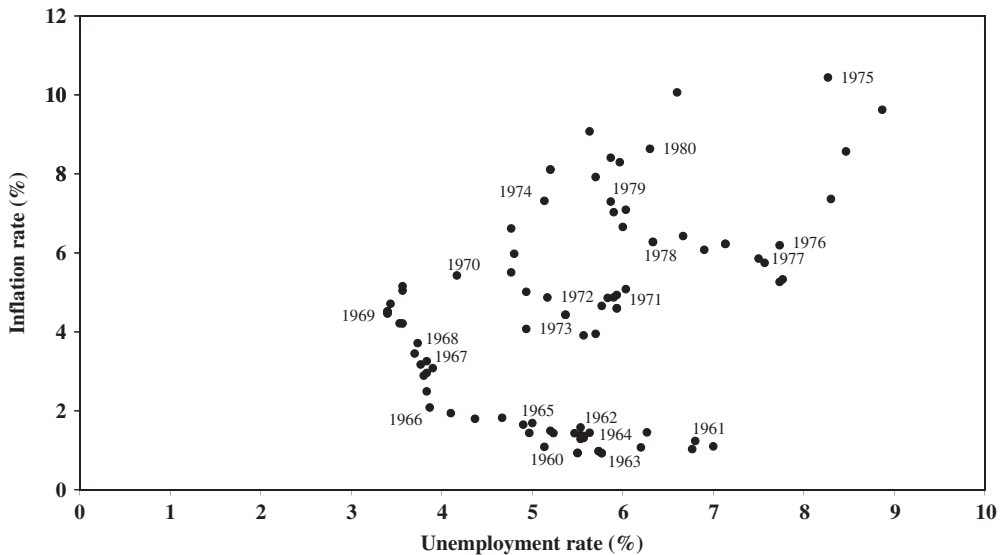


FIGURE 2. Scatter plot of the unemployment and inflation rates, quarterly data, 1960–80.

adjustment of expectations in the natural rate PC. But almost from the beginning, the decade of the 1970s seemed to overturn any thought that the negative PC trade-off was intact or stable. The nature of the problem is evident when we look again at Figure 1, which plots the inflation and unemployment rates in quarterly data since 1960, with the four-quarter change in the deflator for personal consumption expenditures (PCE) used to represent inflation. For the 1970s as a whole, the inflation–unemployment correlation is strongly positive, not negative, and in Figure 1, sharp changes in the inflation rate appear to *lead* by about one year changes in the same direction of the unemployment rate.

When plotted in Figure 2 on a scatter plot from 1960 to 1980, the inflation and unemployment rates are uncorrelated, with a combination of negative and positive correlations that range all over the map. The negative PC trade-off appeared to be utterly defunct. In flowery language that amounted to a simultaneous declaration of war and announcement of victory, Lucas and Sargent (1978, pp. 49–50) described ‘the task which faces contemporary students of the business cycle [is] that of sorting through the wreckage . . . of that remarkable intellectual event called the Keynesian Revolution’.

The year 1975 marks a clear break in the history of the PC. Surveys written at the time focus on the demise of the short-run trade-off and the emergence of the consensus expectational natural rate PC (see, for instance Laidler and Parkin 1975). Two complementary reasons lead us to mark 1975 as the transition year for PC doctrine. First, it was the year of the publication of the policy ineffectiveness proposition summarized above, which was the beginning of the end of business cycle theory based on expectation errors. Second, 1975 was a year in which both the US inflation and unemployment rates experienced the maximum impact (at least up to that time) of supply shocks, calling for a revised PC theory that explicitly incorporated supply shocks.

The demise of the Friedman–Phelps–Lucas information barriers model occurred in two stages. First, the theory was flawed by its inability to reconcile multi-year business cycles with one-month lags faced by agents in obtaining complete information about the aggregate price level. Second, the attempt to develop an empirical counterpart of the

policy-ineffectiveness proposition was a research failure. It floundered on the inability to develop a symmetric explanation of output and price behaviour. Barro (1977) showed that output was not related to anticipated monetary changes but could not demonstrate the required corollary—the full and prompt responsiveness of price changes to anticipated nominal disturbances. This failure reflected the fundamental conflict between the fully flexible prices required by the information barriers model and the inflation inertia deeply embedded into the US inflation process, a conflict that has returned to haunt the application of the NKPC approach in the past decade. Soon Mishkin (1982) and Gordon (1982a) showed that anticipated monetary changes had a strong effect on output in the short run and on inflation in the long run, preserving long-run but not short-run neutrality.

Since 1975 the development of PC doctrine has bifurcated into two divergent paths, called here the ‘left fork’ and ‘right fork’ of the road, with no sign of convergence. The left fork in the road, treated in this section, is the resurrection of Keynesian economics in the form of what I call the ‘mainstream’ PC model that incorporates long-run neutrality, that incorporates explicitly the role of supply shocks in shifting the PC up or down, and that interprets the influence of past inflation as reflecting generalized inertia rather than expected inflation. The right fork in the road of the post-1975 evolution, examined in Section III, features an approach developed by Kydland, Prescott and Sargent, and more recently by Galí, Gertler and others. Inflation depends on forward-looking expectations, and expectations respond rationally to actual and expected changes in monetary and fiscal policy. This two-way game between policy and expectation formation leaves no room for supply shocks or inertia.

The resurrection of the PC

Several years *before* the famous ‘wreckage’ pronouncement by Lucas and Sargent, the resurrection of the PC began. The first and perhaps most important element was the new theory of policy responses to supply shocks, developed independently by Gordon (1975) and Phelps (1978) in two slightly different models that were later merged by Gordon (1984). The ‘Gordon–Phelps’ model starts from the proposition that the price elasticity of demand of the commodity experiencing the adverse supply shock, e.g. oil, is less than unity, so that following an increase in the relative price of oil, the expenditure share of that commodity must increase and the expenditure share of all other components of spending must decrease. For instance, energy’s share of nominal US GDP tripled between 1972 and 1981.⁸

The required condition for continued full employment is the opening of a gap between the growth rate of nominal GDP and the growth rate of the nominal wage (Gordon 1984, p. 40) to make room for the increased nominal spending on oil. If nominal wages are flexible, one option is for the growth rate of wages to become negative, allowing the growth rate of nominal GDP to remain fixed. At the alternative extreme with rigid wages, to avoid a decline in non-energy output, an accommodating monetary policy must boost nominal GDP growth by the amount needed to ‘pay for’ the extra spending on oil, but this will lead to an inflationary spiral if expectations respond to the observed increase in the inflation rate. A third alternative, and the one that actually occurred in the 1970s, was a combination of wage rigidity with a partial response of nominal GDP growth, pushing down both real non-energy spending and employment.

By 1976 this model had made its way into the popular press when a *New York Times* headline announced, ‘A new theory: inflation triggers recession’ (18 July 1976, p. F13). Indeed, we can see in Figure 1 that throughout the period 1974–81, there was a time

lead of roughly one year of inflation relative to unemployment. This real-world result, that an adverse supply shock can depress real output and employment in a world of sticky non-oil prices, had been christened by Okun in 1974 conversations as a ‘macroeconomic externality’.⁹

Sometimes the output effect of the supply shock in the Gordon–Phelps framework is likened to an ‘oil tax’ that reduces non-oil real consumption by more, the smaller is the price elasticity of demand for oil. The extent of the resulting decline in real output depends not only on that elasticity but on the response of nominal demand, which in turn depends not just on the response of monetary policy but also on additional factors listed by Blinder and Rudd (2008)—bracket creep in a non-indexed tax system, negative wealth effects, scrapping of obsolete capital, and the effect of uncertainty in dampening demand.

By 1977 supply shocks had been incorporated into the natural rate expectational Phillips curve. This theoretical formulation (Gordon 1977a, equation 13), except for the absence of explicit lagged terms, is identical to the econometric ‘mainstream’ model developed subsequently and described below:

$$(7) \quad p_t = Ep_t + b(U_t - U_t^N) + z_t + e_t,$$

where the notation is the same as above with the addition of U_t^N to represent the natural rate of unemployment and z_t to represent ‘cost-push pressure by unions, oil sheiks, or bauxite barons’ (Gordon 1977a, p. 133). Other types of supply shocks include the imposition and termination of price controls (as in the US in 1971–74), changes in the relative price of imports, and changes in the trend growth of productivity. Episodes in which political events cause sharp changes in wages, such as the French general strike of 1968, also qualify as adverse supply shocks. A detailed narrative of the role of food, oil and price-control shocks in the inflation of the 1970s is provided by Blinder (1979, 1982).

The process of integrating supply shocks into macroeconomics took place during 1975–78 simultaneously on three fronts: theoretical as described above, empirical as described below, and in an unusual development, through a new generation of intermediate macroeconomic textbooks. An explanation was needed to reconcile the dominant role of demand shocks as the explanation of the Great Contraction of 1929–33 in the same model as would explain the positive correlation of inflation and unemployment in 1974–75. Once recognized, that explanation became obvious. Just as the output and price of corn or wheat could be positively or negatively correlated depending on the importance of micro demand or supply shocks, so aggregate output and the rate of inflation could be positively or negatively correlated, depending on the relative importance of aggregate demand or supply shocks.

The textbooks appeared simultaneously in 1978, and both used alternative versions of a simple diagram that can be traced back to a classroom handout used by Dornbusch at the Chicago Business School in early 1975.¹⁰ The diagram, which has the inflation rate on the vertical axis and either the unemployment or output gap on the horizontal axis, combines three elements—the expectational PC, shifts in that PC caused by supply shocks, and an identity that decomposes nominal GDP growth into inflation and output growth. The textbook version shows that the dynamic aggregate demand–supply model implies a simple first-order difference equation. Following a permanent upward or downward shift in nominal GDP growth, any lags in the formation of expected inflation cause the economy to cycle through loops to its new long-run equilibrium at a zero value of the unemployment or output gap and a permanently higher or lower rate of inflation.

Econometric implementation of the mainstream model

As in equation (7) above, the mainstream specification of the inflation process contains three sets of explanatory variables representing inertia, demand and supply, leading me to call it the ‘triangle’ model.¹¹ Replacing the expected inflation term is a set of long lags on past inflation, reflecting the view that the influence of past inflation reflects generalized backward-looking inertia, not just the formation of expectations. Important sources of inertia include the set of explicit and implicit contracts that dampen short-term changes in prices and wages (as recognized explicitly by Fisher 1926), and the input–output supply chain that creates thousands of links of unknown magnitude and duration between changes in crude and intermediate goods prices and the prices of final goods, as emphasized by Blanchard (1987). All of these channels interact to create the ‘inertia’ effect, the first leg of the triangle.

This approach is Keynesian because the role of inertia is to make the inflation rate slow to adjust to changes in nominal demand, and as a result real GDP emerges as a residual, not as an object of choice as in the Friedman–Phelps–Lucas model. A vast theoretical literature under the rubric of ‘new Keynesian economics’ (NKE), starting in the late 1970s with Fischer (1977) and Taylor (1980), provided numerous models to motivate the inertia mechanism by explaining real and nominal rigidity of wages and/or prices, and many of these explicitly incorporated rational expectations.¹² In our discussion in Section III, we will be careful to distinguish between the theoretical models of the NKE and the empirical application of the NKPC.

In the triangle model, the speed of price adjustment and the speed of expectation formation are two totally different issues. Price adjustment can be delayed by wage and price contracts, and by the time needed for cost increases to percolate through the input–output table, and yet everyone can form expectations promptly and rationally based on full information about the historical response of prices to its own lagged values, to demand shocks and to supply shocks.

The demand side of the model is represented by the level and change of the output gap or alternatively the unemployment gap. As we have seen above, Phillips recognized the role of the ‘rate of change’ effect that at any given unemployment rate makes the inflation rate higher when the unemployment rate is falling than when it was rising. Because the unemployment gap is always entered in the triangle model with both the current value and with additional lags, the zig-zag of the estimated lagged coefficients between negative and positive incorporates the rate of change effect.

The supply side of the model is represented by a set of explicit supply shock variables, establishing a contrast between the mainstream approach and the recent NKPC literature where the supply shock effects are always hidden in the error term. The explicit supply shock variables are all defined so that the absence of supply shocks is represented in (7) as $z_t = 0$. Such variables, for instance, include changes in the *relative* price of oil and changes in the *relative* price of non-oil imports; when these relative prices exhibit zero change, there is no upward or downward pressure on the inflation rate from supply shocks. The list of supply shock variables includes dummy variables which measure the impact of the 1971–74 Nixon-era price controls in holding down inflation in 1971–73 and then adding to the supply shock impact on inflation in 1974–75.¹³ Earlier versions (Gordon 1982b; Gordon and King 1982) included changes in the real exchange rate in place of real import prices.

Unfortunately, the essential role of sticky wages and/or inflexible non-oil prices was missed by many analysts who later tried to model the impact of oil prices on real output. For instance, Bruno and Sachs (1985) use a neoclassical production function to show that

the elasticity of output with respect to the real energy price is the energy share in gross output, thus missing the macroeconomic externality. Hamilton (1983), in a much-cited paper, showed that oil prices Granger-cause changes in real output in all but one of the recessions that occurred between 1948 and 1980. Hamilton's results cannot be compared to those of Bruno and Sachs, or those implied by the triangle inflation equation, because he provides no elasticity estimates and no analysis of the extent to which the oil price effect works through overall inflation, as in the Gordon–Phelps model, or through a direct impact of oil on output via the production function, as in Bruno–Sachs.

Since the 1970s the literature on supply shocks has come increasingly to focus on oil and to neglect shocks related to changes in the relative price of food, of non-oil non-food imports, and the effects of the Nixon-era price controls. In fact, food rather than oil was the example used in the initial development of the macroexternality theory in Gordon (1975). Bosworth and Lawrence (1982) provide ample background on the reasons for sustained increases in the real price of food in 1973–74 and in 1978–79 (another such episode occurred in 2007–08)

Blinder and Rudd (2008) revisit the supply shock explanation of the 'Great Stagflation' of the 1970s and early 1980s, and confirm its central role. They summarize a set of arguments against the supply shock explanation, and refute each. To those (including Barsky and Kilian 2002) who cannot understand why a change in a *relative* price would be relevant for overall inflation, they point to the rigidity of prices in the non-shocked sector. They also assess arguments (like those of Bernanke *et al.* 1997) claiming that the impact of oil shocks on the economy actually represents the effects of the central bank response rather than the oil shock itself. They provide new evidence from a structural vector autoregressive regression (VAR) model reaffirming an independent role for oil shocks. In fact, given ample empirical evidence of long lags in the response of output and unemployment to monetary policy actions, the Blinder–Rudd results make perfect sense—an adverse supply shock causes an initial spike of unemployment, and the monetary policy response then determines by how much unemployment declines in the subsequent years after the shock.

Since the original Phillips (1958) article was about wage changes, not price changes, it is noteworthy that the triangle model is a single reduced-form equation for the inflation rate, with no mention of wage changes. Starting from separate wage change and price mark-up equations, as had been standard in the PC econometrics literature up to that time, Gordon (1982b) merged the two and discussed the simplifying assumptions needed to perform the merger, particularly the absence of wage–wage inertia.

The usual assumption that inflation is equal to nominal wage changes minus productivity growth assumes a fixed value of labour's share in national income. But labour's income share rose sharply in the late 1960s and has drifted down since then. The goal of the central bank is to control inflation, not wage changes, so changes in labour's income share across business cycles imply a loose relation between inflation and wage changes that is fruitfully ignored. An important contribution to the demise of the wage equation was made by Sims (1987), who argued that wage and price equations have no separate structural interpretations, and that a price equation is a wage equation stood on its head, and vice versa.

Empirical results: strengths and weaknesses

The current econometric version of the mainstream or triangle model was originally developed in the late 1970s (Gordon 1977b) and published in its current form (as a single

reduced-form price-on-price equation with no wages) in Gordon (1982b). It has been maintained essentially intact since then, with the same set of explanatory variables and lag lengths, in order to allow post-sample simulations to identify forecasting errors that may call for rethinking the specification. The first challenge to the model arrived almost immediately in the form of the Volcker disinflation of 1979–86. As shown in Figure 1, the inflation rate collapsed from nearly 10% in 1981 to only 3% in 1983–84, much faster than had been forecast by commentators using an expectational PC with a heavy emphasis on wage rigidity.

The ‘sacrifice ratio’ is a convenient summary measure of the speed of inflation adjustment in response to high unemployment and low output. This ratio is defined as the cumulative years of output gap during the disinflation divided by the permanent reduction of inflation expressed as an absolute value. Some *ex ante* forecasts of the sacrifice ratio made in 1980–81 were as high as 10, but the actual sacrifice ratio in retrospect turned out to be between 3.5 and 4.5.¹⁴ The key to the surprisingly low sacrifice ratio turned out to be the role of supply shocks, and in particular the 1981–86 decline in the relative price of energy, and the 1980–85 appreciation of the dollar that reduced the relative price of imports. In a remarkable forecasting success achieved in the middle of the disinflation, Gordon and King (1982) estimated a six-equation VAR model that combined the triangle inflation equation with equations that allowed monetary policy to influence endogenous oil prices and the exchange rate, and their main result was a sacrifice ratio in the range of 3.0 to 3.5, much below the prevailing wisdom of the time.¹⁵

The Gordon–King result is consistent with the Kydland–Prescott–Sargent interpretation—reviewed in the next section—that makes no mention of supply shocks but rather emphasizes the interplay between the credibility of the central bank and the expectations of the public. No doubt a major role in the speed of the disinflation, and the resulting relatively small sacrifice ratio, was the widespread perception that the Fed’s monetary policy changed after 1979 and its anti-inflation stance became more credible than before. The advantage of the Gordon–King method is that the channels of monetary policy are explicitly traced, not just through high unemployment but also through the effect of the monetary–fiscal policy mix in causing an appreciation of the dollar in 1980–85, with an accompanying decline in the relative price of oil and of non-oil imports.

Returning now to Figure 1, we see that the Volcker disinflation was followed in the late 1980s by a repeat of the negative inflation–unemployment trade-off already experienced in the 1960s, albeit with a smaller acceleration of inflation and a higher level of unemployment. Similarly, the negative trade-off is evident in the slowdown of inflation in 1990–93 in response to a marked increase in the unemployment rate during the same period.

At first glance, the behaviour of the PC in the 1990s appears to be puzzling. Unemployment in the late 1990s fell to the lowest rate since the 1960s, but there was no parallel acceleration of inflation. Instead, inflation was lower in 2000 than in 1993. As shown by Gordon (1998), low inflation in the late 1990s can be explained by beneficial supply shocks that pushed the PC down in contrast to the adverse supply shocks of the 1970s; the beneficial shocks of the 1996–99 period included lower real energy prices, lower relative import prices, and faster trend productivity growth. As shown in Figure 1, the ‘twin peaks’ of inflation and unemployment were joined by the ‘valley’ of inflation and unemployment during 1997–2000.

Despite these research successes, the evolution of the data required one change in the 1982 specification of the triangle model. Post-sample simulations in 1994–95 revealed that the model’s predictions had started to drift in the direction of predicting too much inflation, given actual values of the unemployment gap. These errors turned out to be due

not to a flaw in the model but to a data choice, that is, the false assumption that the natural rate of unemployment was fixed, allowed to change only in response to the demographic composition of the unemployment rate.¹⁶

For several decades the natural rate of unemployment has been called by its nickname, the 'NAIRU', standing for 'non-accelerating inflation rate of unemployment'. The time-varying NAIRU (or TV-NAIRU) combined an econometric method introduced by Staiger *et al.* (1997) that was applied to a version of my triangle model, and simultaneously I published a paper which used their method applied to my model (Gordon 1997). The estimated TV-NAIRU exhibited a pronounced downward drift after 1990 that explained in a mechanical way why the inflation rate was lower in the 1990s than had previously been predicted with a fixed NAIRU. An initial set of substantive explanations of this decline in the NAIRU was provided by Katz and Krueger (1999).

III. THE POST-1975 RIGHT FORK IN THE ROAD: JUMPING AND FORWARD-LOOKING POLICY-RESPONSIVE EXPECTATIONS

The alternative post-1975 research approach, the right fork of the road, emphasizes jumps in expectations in response to policy actions and implicitly incorporates price flexibility, market clearing and an absence of backward-looking inflation inertia. The central idea that expected inflation can jump in response to actual or anticipated policy changes is crucial to understanding the ends of hyperinflations (Sargent 1982). It begins with the basic proposition, already embedded in the Friedman–Phelps natural rate hypothesis, that the choice by a policy-maker of a particular short-run combination of inflation and unemployment rates can alter expectations, causing the trade-off to change.

The policy game

Kydland and Prescott (1977) distinguished between policy discretion and rules, contrasting discretionary policy-makers who reassess the desired response to alternative inflation rates in each successive time period, with rule-following policy-makers who adhere to a rule which is fixed for all future time periods. They show, not surprisingly, that the long-run inflation rate is higher under a discretionary policy than under a rules-based policy. How does this approach explain the positive correlation of inflation and unemployment in the 1970s without mention of supply shocks? Papers written by Sargent (1999), Cogley and Sargent (2005) and Sargent *et al.* (2006) begin with the standard presumption that choices by discretionary policy-makers will cause the PC to shift and policy options to change. The attempt to conduct policy without knowledge of the current position of the Phillips curve can lead a policy-maker to make choices that yield persistently high inflation outcomes.

'Credibility' is an important concept in the game involving policy-makers and private agents. Because expectations can jump in response to changes in policy-makers' actions and perceived intentions, the outcome of actual inflation is higher if agents infer that the policy-maker is trying to manipulate unemployment along the short-run PC trade-off. A credible policy is one which promises to maintain a low inflation rate in the long run; agents are convinced that a policy is credible if the policy-makers pursue an inflation target and regularly raise the interest rate when inflation exceeds its target but do not lower interest rates in response to an increase in unemployment. Doubts by agents that

the policy-maker is committed to low inflation in the long run can raise the unemployment cost of reducing inflation, i.e. the sacrifice ratio.

One problem with this line of research is that it ignores additional information available to policy-makers—that oil or farm prices have risen, that the dollar has been devalued, that price controls have been imposed or ended, or that trend productivity growth has slowed or revived. Indeed, it is striking that Sargent *et al.* (2006) claim to be able to explain the entire upsurge of inflation in the 1970s and early 1980s without any mention of supply shocks, despite the fact that the word ‘shocks’ appears in the title of their paper: ‘allow the model to reverse engineer a sequence of government beliefs about the Phillips curve which, through the intermediation of the Phelps problem, capture both the acceleration of U.S. inflation in the 1970s and its rapid decline in the early 1980s’.

Another problem with the policy game approach is that it ignores the policy-maker’s fundamental dilemma in the face of an adverse supply shock. As shown in Gordon (1975, 1984) and Phelps (1978), unless wages are perfectly flexible, the policy-maker cannot escape a choice between holding inflation constant at the cost of substantial extra unemployment, or holding unemployment constant at a cost of higher and accelerating inflation, or something in between. In fact, because of long lags in the impact of monetary policy on unemployment and inflation, in reality the policy-maker is incapable of holding either inflation or unemployment constant following a supply shock.

Related work by Primiceri (2006) includes the government’s underestimate of the NAIRU in the 1970s as a cause of high inflation, but he does not provide any explicit analysis of supply shocks as the cause of this underestimate. Sims (2008) has suggested that Primiceri is guilty of an asymmetry, because he allows only for uncertainty about coefficient values in a model that policy-makers assume is correct, instead of allowing for the fact that the model may be wrong. In fact, the discussion above of the mainstream model suggests that Primiceri and others working on policy–expectations interactions may indeed have chosen the wrong model, at least for the USA, by assuming that expectations can jump in response to policy announcements and ignoring the role of backward-looking inertia and supply shocks.

The new Keynesian Phillips curve

The NKPC model has emerged in the past decade as the centrepiece of macro conference and journal discussions of inflation dynamics and as what Blanchard and others have called the ‘workhorse’ of the evaluation of monetary policy.¹⁷ The point of the NKPC is to derive an empirical description of inflation dynamics that is ‘derived from first principles in an environment of dynamically optimizing agents’ (Bårdsen *et al.* 2002).

The theoretical background is that monopolistically competitive firms have control over their own prices due to product differentiation. They are constrained by a friction in the setting of prices, of which there are many possible justifications that are inherited from the theoretical NKE literature. For instance, we have already cited Taylor’s (1980) model which merges rational expectations with fixed-duration contracts. More frequently cited, as in Mankiw’s (2001) exposition, is Calvo’s (1983) model of random price adjustment, in which prices are fixed for random periods. The firm’s desired price depends on the overall price level and the unemployment gap. Firms change their price only infrequently, but when they do, they set their price equal to the average desired price until the time of the next price adjustment. The actual price level, in turn, is equal to a weighted average of all prices that firms have set in the past. The first-order conditions for optimization then imply that expected future market conditions matter for today’s

pricing decision. The model can be solved to yield the standard NKPC specification in which the inflation rate (p_t) depends on expected future inflation ($E_t p_{t+1}$) and the unemployment (or output) gap:

$$(8) \quad p_t = \alpha E_t p_{t+1} + \beta(U_t - U_t^*) + e_t.$$

The constant term is suppressed, so the NKPC has the interpretation that if $\alpha = 1$, then U_t^* represents the NAIRU.

Notice that the NKPC in equation (8) is identical to the post-1975 mainstream PC written in (7) above, with two differences. First, there is no explicit treatment of supply shocks; these are suppressed into the error term. Second, expectations are explicitly forward-looking in equation (8), whereas in (7) expectations could be either forward-looking or backward-looking, or both. Because of frictions of the Taylor or Calvo type, policy changes that raise or lower the inflation rate have short-run effects on the unemployment or output gap. The Taylor NKE framework assumes fixed contract lengths of pricing intervals, while the Calvo model makes price changes dependent on a fixed gap between the actual and desired price levels. But Sims (2008) points out that this ‘theory has simply moved the non-neutrality from agent behavior itself into the constraints the agent faces, the frictions’. In real-world situations in which macro shocks create Argentina-like instability, contract lengths would surely change in response to the expected inflation rate.

NKPC models vary in their inclusion of the single variable that supplements future expected inflation. This is modelled sometimes as the unemployment gap, as in (8), and sometimes as the closely related output gap. Mankiw’s (2001) exposition, followed below, uses the unemployment gap.

Another version of the NKPC replaces either gap with real marginal cost. This is always proxied by real average cost, that is, the real wage divided by the average product of labour $(W/P)/(Y/N)$, which is by definition equal to labour’s share in national income (WN/PY) . Some papers in the NKPC literature treat real marginal cost as exogenous, but this is unacceptable because labour’s income share is inherently endogenous and requires a multi-equation model with separate equations for the level of the wage rate, the price level and the level of labour productivity. Thus far empirical implementation of the marginal cost version of the NKPC has swept under the rug the endogeneity of labour’s income share. In contrast, Dew-Becker and Gordon (2005) have examined joint feedback between prices and wages by endogenizing changes in labour’s share.¹⁸

This section treats the ‘right fork’, with its absence of inertia and expectations that can jump in response to anticipated policy changes, as a fruitful development in macroeconomics when applied to rapid inflation episodes, whether in Germany in 1922–23 or in Brazil or Argentina more recently. Unfortunately, the empirical implementation of the NKPC has been almost entirely to data for the postwar USA, where it is the wrong model. This can be easily seen for both the ‘gap’ and ‘real marginal cost’ versions of the NKPC. The ‘gap’ version as written above in equation (8) drives changes in the inflation rate only with changes in the unemployment gap, or equivalently with the output gap. Its prediction is that the coefficient (β) on the unemployment gap is negative. But we have already seen in Figure 1 that the correlation between inflation and the unemployment rate is both negative and positive, with a positive correlation between 1971 and 1982 when the variance of inflation was greatest. The NKPC contains no element to capture the switch from negative to positive correlation and back again.

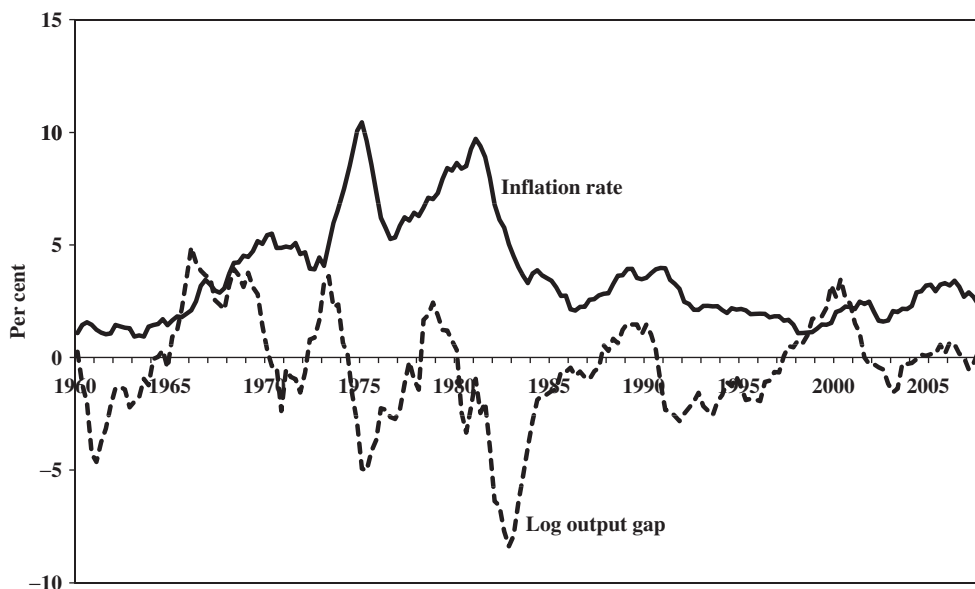


FIGURE 3. The inflation rate and the per cent log output gap, quarterly data, 1960–2007.

Correspondingly, the model's prediction is that when the unemployment gap is replaced by the output gap, the correlation should be positive. But as shown in Figure 3, the correlation between inflation and the output gap is strongly negative between 1971 and 1982. Thus it is not surprising that Rudd and Whelan (2005b, Table 1) show that the estimated coefficient on the output gap is significantly *negative*. Their result is obtained by the usual procedure of replacing the unknown expectation of future inflation with two-stage least squares estimation in which the first stage regresses the actual inflation rate on a set of instruments.¹⁹ This apparent conundrum is resolved in the mainstream triangle model in which the output gap leads inflation positively (as in 1965–69 and 1986–89) while inflation leads the output gap negatively (as in 1971–82) due to the influence of supply shocks.

A look at the data also predicts a failure of the version of the NKPC that uses real marginal cost as the variable that drives inflation. As noted above, real marginal cost is always proxied by real average cost, which is the same as labour's income share, and this share is plotted against the inflation rate in Figure 4.²⁰ Labour's share exhibits one big upward jump in 1967–70, at least four years too early to explain the first inflation peak in 1974–75. After 1970, labour's share is essentially trendless, varying only between 70% and 75%, with no movements that would help to explain the second inflation peak in 1979–81 nor the Volcker disinflation of 1981–84. Accordingly, it is not surprising that Rudd and Whelan (2005b, Table 1) estimate an insignificant coefficient in equation (8) when real marginal cost replaces the unemployment gap.

The NKPC literature seems to be just as confused by the behaviour of real marginal cost as by the negative correlation of inflation with the output gap. As Woodford (2003) has pointed out, the standard model predicts that increases in output tend to be accompanied by higher real marginal cost as workers move out of a positively sloped labour supply curve, as overtime premia rise, and as input materials costs respond positively. However, Figure 4 indicates that, at least before 1990, labour's income share peaked in recessions and appears to be countercyclical. This has an easy explanation that

TABLE 1
ESTIMATED EQUATIONS FOR QUARTERLY CHANGES IN THE PCE DEFLATOR, 1962(I) TO 2007(IV)

Variable	Roberts		
	Lags	NKPC	Triangle
Constant		1.16**	
Lagged dependent variable	1–24 ^a	1.01**	
	1–4	0.95**	
Unemployment gap	0–4		– 0.56**
Unemployment rate	0	– 1.17*	
Relative price of imports	1–4		0.06**
Food–energy effect	0–4		0.89**
Productivity trend change	1–5		– 0.95**
Nixon controls ‘on’	0		– 1.56**
Nixon controls ‘off’	0		1.78**
R^2		0.78	0.93
SEE		1.17	0.64
SSR		244.0	64.6
Dynamic simulation			
1998(I) to 2007(IV)	Note b		
Mean error		– 2.75	0.29
Root mean-square error		3.20	0.70

Notes

**Indicates significance at 1%; *indicates significance at 5%.

^aLagged dependent variable is entered as the four-quarter moving average for lags 1, 5, 9, 13, 17 and 21, respectively.

^bDynamic simulations are based on regressions for the sample period 1962(I) to 1997(IV) in which the coefficients on the lagged dependent variable are constrained to sum to unity.

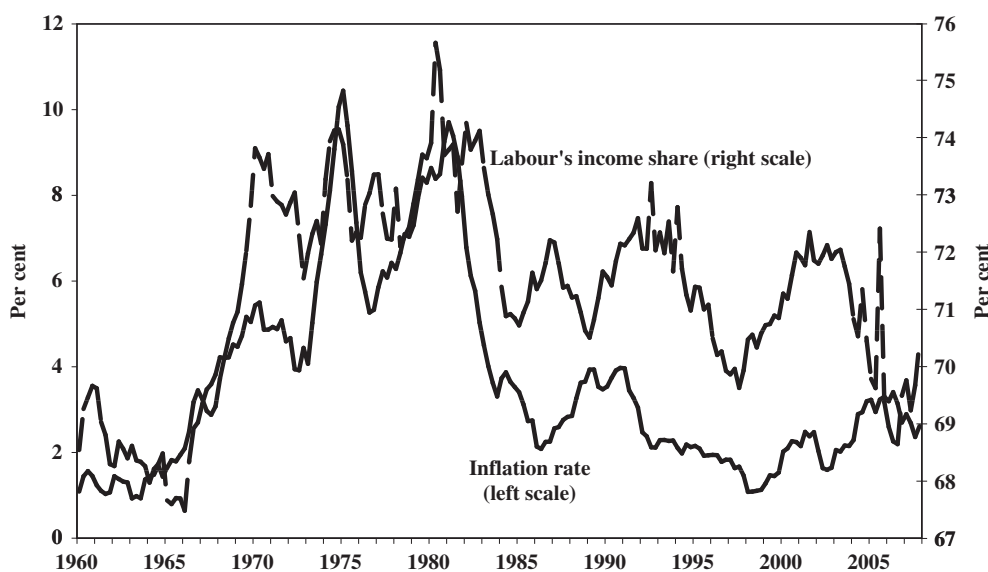


FIGURE 4. The inflation rate and labour's share in domestic net factor income, quarterly data, 1960–2007.

has apparently been neglected in many NKPC discussions: the procyclicality of labour productivity, which appears in the denominator of real average cost. Rudd and Whelan (2005b) also discuss the problem that labour's share, which equals real average cost, may be a poor proxy for real marginal cost.

The challenge of persistence

On the surface, the NKPC as written in equation (8) appears similar, except for the omission of explicit supply shock variables, to the mainstream PC as written in equation (7). But its policy implications are radically different from the mainstream model, with its costly disinflation and significant sacrifice ratio. This occurs because in the NKPC model there is no backward-looking inertia, that is, no structural dependence of inflation on its own lagged values. Instead, inflation is entirely driven by forward-looking expectations, and equation (8) can be solved forward to set the inflation rate equal to an infinite sum of expected future output gaps. Inflation can be costlessly controlled by a credible commitment to follow policies that minimize the output gap forever into the future.

However, as we shall see in Section IV, inflation persistence in the form of long lags on past inflation rates is a central feature of postwar US inflation behaviour. As a result, in the US environment expectations are unlikely to jump except in response to widely recognized supply shocks, such as the surge of oil prices in 1973–75, 1979–81 or 2006–08. The recognition that, in the absence of supply shocks, the inflation rate is dominated by persistence creates a challenge for policy-makers to reduce inflation by altering public expectations *directly*. How can policy-makers convince the public that inflation will spontaneously decrease, without any cost of higher unemployment or lost output, when the public knows that inflation behaviour is dominated by persistence?

As we show in the first subsection of Section IV, in practice the NKPC is simply a regression of the inflation rate on a few lags of inflation and the unemployment gap. As pointed out by Fuhrer (1997), the only sense in which models including future expectations differ from purely backward-looking models is that they place restrictions on the coefficients of the backward-looking variables that are used in the first stage of two-stage least squares estimation as proxies for the unobservable future expectations. In Fuhrer's words:

Of course, some restrictions are necessary in order to separately identify the effects of expected future variables. If the model is specified with unconstrained leads and lags, it will be difficult for the data to distinguish between the leads, which solve out as restricted combinations of lag variables, and unrestricted lags. (Fuhrer 1997, p. 338)

Subtleties in the interpretation of the 'hybrid' NKPC

Gali and Gertler (1999), two of the inventors of the NKPC approach, have introduced a 'hybrid' NKPC model in which the public consists of both forward-looking and backward-looking agents, and in their empirical version current inflation depends on both expected future inflation and past inflation. However, since future inflation is always proxied by some transformation of past inflation, there is little difference in practice between the 'pure' forward-looking NKPC and the hybrid version, except for the form of the restrictions that emerge. Further, if there are enough backward-looking members of the population, then forward-looking members cannot ignore the persistence introduced by backward-looking agents. This dependence of future contract outcomes on the inheritance of ongoing contracts with staggered expiration dates has been explicit in the theoretical NKE literature since its introduction by Taylor (1980).

The hybrid NKPC model is the same as equation (8) above, except that the influence of inflation is divided between future expected inflation and lagged inflation rather than being channelled exclusively through future expected inflation:

$$(9) \quad p_t = \alpha_f E_t p_{t+1} + \alpha_b p_{t-1} + \beta(U_t - U_t^*) + e_t.$$

The central issue is the relative size of the forward-looking and backward-looking coefficients (α_f and α_b). Galí and Gertler (1999) and Galí *et al.* (2005) have reported estimates of equation (9) with the unemployment gap replaced by labour's income share and conclude that 'forward-looking behavior is dominant', i.e. α_f is estimated to be much larger than α_b . This conclusion is important, since it appears to justify the original formulation of the NKPC (equation (8) above) and makes the role of lagged inflation appear to be a minor distraction of little empirical importance.

However, as pointed out by Rudd and Whelan (2005a, b), these estimates do not actually distinguish between forward-looking and backward-looking behaviour due to the nature of the two-stage least squares exercise. The second-stage equation (9) omits variables that belong in the true model of inflation—e.g. additional lags on inflation itself as well as explicit supply shock variables like commodity prices—but includes them in the first stage as a proxy for expected future inflation. Indeed, anything that is correlated with current inflation but not included in the second stage will serve as a good instrument for future expected inflation and thus raise α_f relative to α_b . These omitted variables boost the coefficient on expected future inflation even if expected future inflation has no influence at all on inflation itself, as occurs when Rudd and Whelan estimate a pure backward-looking model that includes some of the additional variables included as instruments in the two-stage procedure. Overall, the NKPC approach to date has delivered no evidence that expectations are forward-looking. The instruments used in the first stage are incompatible with the theory posited in the second stage. If lagged inflation and commodity prices matter for inflation, then why are they omitted from the NKPC inflation equation?

Constraints on the formation of expectations

Recent research has revived the discussion of barriers to the formation of expectations. As we have seen, the original formulations of Friedman, Phelps and Lucas were based on information barriers that prevented one set of agents (Friedman's workers) or all agents (Phelps' desert island residents) from having access to government data on income, output, money and prices that are released frequently at zero cost to all agents in the economy. That literature was flawed because it placed the information barriers in the wrong place, in an inability to perceive costless macro information, instead of where the information barriers really exist, at the micro level of costs and supplier–producer relationships.

Producers of final goods are unable to perceive cost increases of crude and intermediate materials that may be in the pipeline, and they have no choice but to wait until they receive notification of actual cost changes (with the exception of crude materials like oil where prices are determined in public auction markets). This approach, based on supplier–producer arrangements, was introduced in Blanchard (1987) and was christened the 'input–output' approach in Gordon (1990), who suggests a four-cell matrix of information barriers of supply and demand shocks at both the macro and micro level. A fundamental source of persistence is not just explicit wage contracts as analysed by Taylor, but also explicit or implicit price contracts between suppliers and

producers of final goods. Even without contracts, persistence and inertia are introduced by lags between price changes of crude materials, intermediate goods and final goods. For some goods, e.g. cars or aircraft, there are literally thousands of separate intermediate goods, and most of these are made up of further layers of intermediate goods.

The recent literature has largely ignored the *micro* uncertainty embodied in the input–output approach and instead has attempted to find credible explanations for imperfect *macro* information. One approach is that agents take time to learn about the structure of the economy (see Orphanides and Williams 2005). This information barrier is consistent with the triangle approach, in which changes in the TV-NAIRU are observed not in real time but only after the fact, as are changes in coefficients on the PC slope or the coefficients on such supply shocks as oil prices.

A second barrier may be imperfect information regarding the goals of the central bank (see, among others, Kiley 2007). Clearly, in the US context the Fed has changed goals several times, and this became evident only after the fact. The Volcker policy shift in 1979–80 was widely noticed at the time, but there was no historical antecedent to allow predictions of its consequences. Likewise, studies of the Taylor rule indicate that the Fed shifted around 1990 from a policy that mainly responded to inflation to a policy that mainly responded to the output gap, and no empirical Taylor rule can explain why the Federal Funds rate was so low in 2001–04.

The third barrier consists of costs or constraints on information acquisition and processing. One version of this approach emphasizes costs of obtaining information that lead to infrequent adjustments in expectations (Reis 2006). Another approach (see Sims 2006) is called ‘rational inattention’ and also emphasizes constraints on information processing capabilities. However, all of these barriers concern constraints on the ability of private agents to adjust their expectations accurately to reflect the current stance of monetary policy and anticipated future changes in policy, and none reflects any of the sources of persistence and inertia, particularly lack of information at the micro level of the input–output table. Rational inattention makes sense at the micro level, when translated into the minimization of managerial cost by avoiding daily deliberations about price changes required by changes in supplier costs and instead making decisions infrequently.

Which model applies to which episodes?

This survey has contrasted the inertia-bound triangle approach to explaining US inflation with alternative frameworks in which the expectations of private agents can jump in response to perceived changes in monetary policy. Which model best describes which historical situations? As indicated above, the mutual interplay between policy decisions and expectations formation is essential to understanding episodes of rapid inflation, including Sargent’s (1982) ends of four big inflations. This approach is in fact essential to an understanding of the inflation process in any nation that has experienced high inflation volatility in the past due to shifts in policy (as contrasted with the influence of supply shocks). A prime example would be a country like Argentina in which private agents know that the government’s ability to restrain monetary growth depends on fiscal decisions made at the level of states and localities.

Relatively little research has been done to establish a dividing line between situations suitable for analysis with the policy–expectations game approach vs the inertia-bound triangle approach. The convergence of inflation rates within the European Monetary Union between 1980 and 1998 provides another example in which an inertia-dominated

PC is inadequate, as countries with similar unemployment rates experienced very different time paths of inflation. After experiencing inflation rates of over 20% to 25% in the mid-1970s, Italy and the UK converged to low single digits of inflation, and an explanation of this convergence requires a model in which agents formed expectations based in part on the monetary policy of the Bundesbank, not just that of their own national central bank.²¹

Another issue in extending the PC framework to fit the postwar European experience is the question of whether the standard PC relation between the inflation rate and the level of unemployment needs to be supplemented by a hysteresis mechanism between the inflation rate and the change of unemployment. Recently Ball (2008) has suggested that the hysteresis idea, after languishing since its invention by Blanchard and Summers (1986), should be revived. Some versions of hysteresis imply that inflation depends not on the level of unemployment but on its rate of change, that ancient idea supported in the results of both Fisher (1926) and Phillips (1958) and incorporated in the triangle model specification.

The empirical results presented in Section IV suggest that the triangle model, which combines demand and supply shocks with inertia, does a much better job in explaining postwar US inflation than does the NKPC approach that omits lags and supply shocks. However, how far back can the triangle-type PC specification be pushed in US data? Samuelson and Solow (1960) had already noticed that the American PC relationship does not work in the Great Depression and during the two world wars. A quantitative answer to this question was provided by Gordon (1982a) in a unique set of interpolated quarterly data extending back to 1892. His results (Table 3) estimate PC equations for 1892–1929, 1929–53 and 1954–80, in a framework which allows the inflation rate to depend not just on the level of the output gap, but on changes in expected and unexpected nominal GDP, lagged inflation and a series of dummy variables.

Gordon's results suggest that prior to 1954 there were substantial shifts in the American PC process that are consistent with a role for an interplay between the expectations of private agents and perceived changes in policy and the macroeconomic environment more generally. The PC relation, in the form of the coefficient on the output gap, has roughly the same coefficient before 1929 as after 1954, but is zero during the middle period. In all three periods the anticipated and unanticipated change in nominal GDP is highly significant, and this 'rate of change effect' dominates the explanation of inflation in the middle period when the PC relationship is absent (see also Romer 1996, 1999 for an analysis of price changes in the 1930s).

The role of policy in shifting the inflation rate (whether this is perceived as working through expectations or not) is evident in large shift coefficients on the impact of nominal GDP changes on inflation in the First World War, and in large negative dummy variables for price controls in the First and Second World Wars, the Korean War and the 1971–74 Nixon episode, as well as a large positive coefficient for the New Deal's National Recovery Administration.²²

IV. THE NEW KEYNESIAN AND TRIANGLE PHILLIPS CURVES: SPECIFICATION, ESTIMATES AND SIMULATIONS

What difference does it make if we explain US inflation using the mainstream triangle model or the NKPC? We now turn to the detailed specification of the NKPC and triangle alternatives. Then we examine their performance in US data spanning 1962–2007.

The NKPC model

A central challenge to the NKPC approach is to find a proxy for the forward-looking expectations term in equation (8) above ($E_t p_{t+1}$). Surprisingly, there is little discussion in the literature of this aspect, or the implications of the usual solution, which is to use instrumental variables or two-stage least squares (2SLS) to estimate (8). The following treatment is consistent by including in the first stage only the variables that are part of the basic theory in the second stage. The first-stage equation to be included in the 2SLS estimation explains expected future inflation by recent lags of inflation and by the current unemployment gap:

$$(10) \quad E_t p_{t+1} = \sum_{i=1}^4 \lambda_i p_{t-i} + \phi(U_t - U_t^*).$$

When the first-stage equation (10) is substituted into the second-stage equation (8), we obtain the reduced form

$$(11) \quad p_t = \alpha \sum_{i=1}^4 \lambda_i p_{t-i} + (\alpha N + \beta)(U_t - U_t^*) + e_t.$$

Thus in practice the NKPC is simply a regression of the inflation rate on a few lags of inflation and the unemployment gap. We have already cited Fuhrer (1997) as pointing out that the only sense in which models including future expectations differ from purely backward-looking models is that they place restrictions on the coefficients of the backward-looking variables, as in (11). The procedure of Galí and Gertler (1999) and many others of adding additional variables like commodity prices and wage changes to the first-stage equation is entirely *ad hoc*, as pointed out by Rudd and Whelan (2005a, b) because any relevance of these variables to the forecasting of future inflation is inconsistent with the basic second-stage NKPC inflation model of equation (8), which omits these additional variables.

The Roberts (2006) version of the NKPC is of particular interest here, because of his finding that the slope of the PC has declined by more than half since the mid-1980s. Roberts describes his equation as a ‘reduced form’ NKPC, and indeed it is identical to equation (11) above with two differences: the NAIRU is assumed to be constant, and the sum of coefficients on lagged inflation is assumed to be unity. Thus the Roberts (2006, equation 2, p. 199) version of (11) is

$$(12) \quad p_t = \sum_{i=1}^4 \lambda_i p_{t-i} + \gamma + \beta U_t + e_t,$$

where the implied constant NAIRU is $-\gamma/\beta$.

The triangle model of inflation and the role of demand and supply shocks

The inflation equation used in this paper is almost identical to that developed 25 years ago (Gordon 1982b). When the influence of demand is proxied by the unemployment gap, the triangle model can be written as (13), which is identical to (7) above except for the introduction of lags. This general framework can be written as

$$(13) \quad p_t = a(L)p_{t-1} + b(L)(U_t - U_t^N) + c(L)z_t + e_t.$$

As before, lower-case letters designate first differences of logarithms, upper-case letters designate logarithms of levels, and L is a polynomial in the lag operator.²³

As in the NKPC and Roberts approaches, the dependent variable p_t is the inflation rate. Inertia is conveyed by a series of lags on the inflation rate (p_{t-1}). The term z_t is a vector of supply shock variables (normalized so that $z_t = 0$ indicates an absence of supply shocks), and e_t is a serially uncorrelated error term. Distinguishing features in the implementation of this model include unusually long lags on the dependent variable, and a set of supply shock variables that are uniformly defined so that a zero value indicates no upward or downward pressure on inflation. Current and lagged values of the unemployment gap serve as a proxy for the influence of demand, where the unemployment gap is defined as the difference between the actual rate of unemployment and the NAIRU, and the NAIRU is allowed to vary over time. This specification predicts steady inflation when the unemployment gap and the supply shock terms are all zero, and hence it is always estimated without a constant term.

The estimation of the time-varying (TV) NAIRU combines the above inflation equation (12) with a second equation that explicitly allows the NAIRU to vary with time:

$$(14) \quad U_t^N = U_{t-1}^N + \eta_t, \quad E\eta_t = 0, \quad \text{var}(\eta_t) = \tau^2.$$

In this formulation, the disturbance term η_t in the second equation is serially uncorrelated and is uncorrelated with e_t . When its standard deviation $\tau_\eta = 0$, the natural rate is constant, and when τ_η is positive, the model allows the NAIRU to vary by a limited amount each quarter. If no limit were placed on the variance of the NAIRU, then the TV-NAIRU would jump up and down and soak up all the residual variation in the inflation equation (13).²⁴ In practice, the smoothness criterion is chosen to avoid negatively correlated zig-zags in the estimated NAIRU, to be consistent with Friedman's original (1968) idea of the NAIRU as slowly changing in response to underlying microeconomic structural factors.

The triangle approach differs from the NKPC and Roberts approaches by including long lags on the dependent variable, additional lags on the unemployment gap that incorporate a rate-of-change effect, and explicit variables to represent the supply shocks (the z_t variables in (13) above), namely the change in the relative price of non-food non-oil imports, the effect on inflation of changes in the relative price of food and energy, the change in the trend rate of productivity growth, and dummy variables for the effect of the 1971–74 Nixon-era price controls.²⁵

Estimating the TV-NAIRU

The time-varying NAIRU is estimated simultaneously with the inflation equation (12) above. For each set of dependent variables and explanatory variables, there is a different TV-NAIRU. For instance, when supply shock variables are omitted, the TV-NAIRU soars to 8% and above in the mid-1970s, since this is the only way the inflation equation can 'explain' why inflation was so high in the 1970s. However, when the full set of supply shocks is included in the inflation equation, the TV-NAIRU is quite stable, shown by the dashed line plotted in Figure 5, remaining within the range 5.7% to 6.5% over the period between 1962 and 1988.

Beginning in the late 1980s, the TV-NAIRU drifts downwards until it reaches 5.3% in 1998, and then it displays a further dip in 2004–06 to 4.8%. One hypothesis to be explored below is that the Roberts NKPC implementation reaches the conclusion that

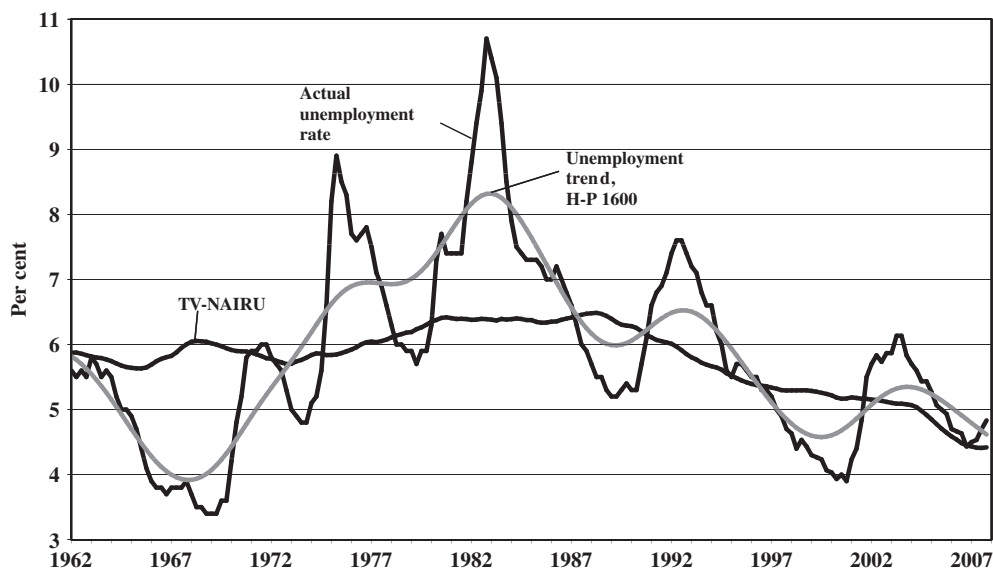


FIGURE 5. Actual unemployment rate vs time-varying NAIRU, 1962–2007.

the Phillips curve has flattened because the NAIRU is forced to be constant, and that a decline in the TV-NAIRU is an alternative to a flatter PC in explaining why inflation has been relatively well behaved in the past 20 years.

Some of the NKPC literature estimates the TV-NAIRU by directly applying a Hodrick–Prescott (H–P) filter to the time series of the unemployment rate.²⁶ As shown in Figure 5 using the traditional H–P parameter of 1600, this ‘direct’ approach to estimating the TV-NAIRU results in an unexplained increase in the TV-NAIRU from 3.9% in 1968 to 8.3% in 1985, whereas the triangle approach has no such unexplained increase because of its introduction of explicit supply shock variables.²⁷ In contrast, the Roberts implementation of the NKPC forces the NAIRU to be constant at an estimated 7.0%.

How much difference do the explicit supply shock variables make in the predictions of the triangle specification? Figure 6 displays predictions made with the actual supply shock variables and with the supply shock variables zeroed out (but with the other variables, including the unemployment gap, taking their historical values and estimated coefficients). Evidently, the supply shock variables explain *all* of the twin peaks of inflation in the 1973–81 period, and in addition they explain more than half of the Volcker disinflation (predicted inflation drops by 7 percentage points between 1980(I) and 1985(I) when supply shocks are included, but by only 3 percentage points when supply shocks are excluded). Notice also that without (beneficial) supply shocks, the influence of low unemployment would have caused inflation to rise by 1.3 percentage points between 1994 and 2001, whereas with supply shocks inflation is predicted to be roughly constant.

Roberts NKPC vs triangle: coefficients and simulation performance

We next turn to the estimated coefficients, goodness of fit and simulation performance of the Roberts NKPC and triangle PC specifications. Table 1 displays the estimated sums of coefficients and their significance levels for both the Roberts NKPC and triangle

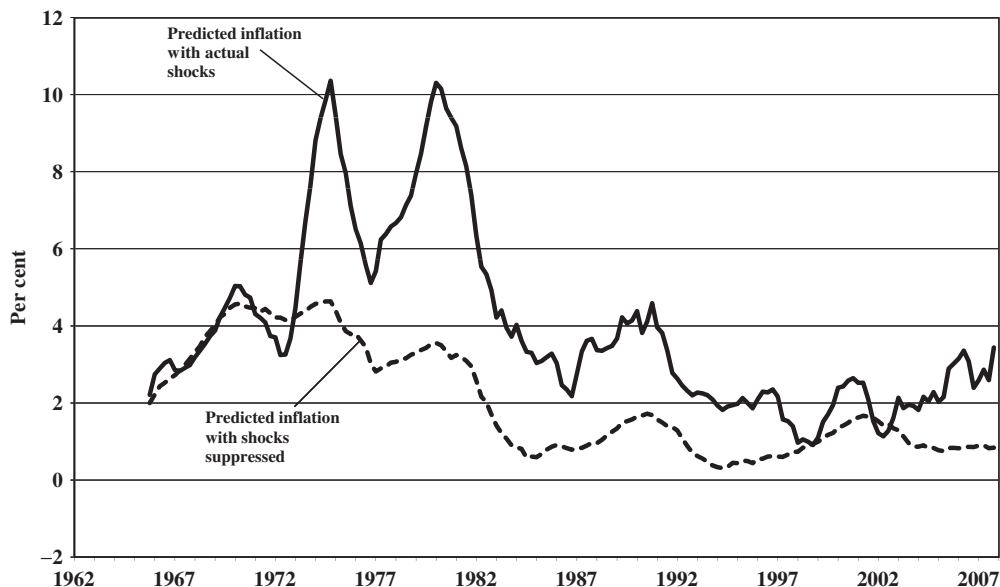


FIGURE 6. Predicted inflation in triangle mode with and without supply shocks, 1962(I) to 2007(IV).

specifications for equations in which the dependent variable is the quarterly change in the headline PCE deflator. In both specifications the sum of coefficients on the lagged inflation terms is very close to unity, as in previous research.²⁸ The sum of the unemployment gap variables in the triangle approach is around -0.6 , which is consistent with a stylized fact first noticed by Samuelson and Solow (1960) that the slope of the short-run Phillips curve is roughly minus one-half. Why is the Roberts NKPC unemployment coefficient lower than in the triangle specification? The excluded supply shock variables are positively correlated with inflation and positively correlated with the unemployment gap, so the omission of these supply shock variables causes the negative coefficient on the unemployment gap to be biased towards zero. We note that the sum of squared residuals (SSR) for the triangle model is barely one-quarter that of the Roberts NKPC specification.

The explicit supply shock variables in the triangle model are all highly significant and have the correct signs; except for the productivity trend variable, all of these enter exactly as specified in 1982 and thus their significance has not been diminished by an extra 25 years of data. The change in the relative import price effect has a highly significant coefficient of 0.06.²⁹ The food–energy effect has a coefficient of 0.89, close to the expected value of 1.0. The productivity trend variable has the expected negative coefficient and helps to explain why inflation accelerated in 1965–80 and was so well behaved in 1995–2000. The Nixon-era control coefficients, as in previous research, indicate a significant impact of the controls in holding down inflation by a cumulative -1.6% in 1971–72 and boosting inflation by 1.8% in 1974–75.

Rather than relying on the usual statistical measures of goodness of fit, a dynamic model heavily dependent on the contribution of the lagged dependent variable is best tested by the technique of dynamic simulations. These generate the predictions of the equation with the lagged dependent variable generated endogenously rather than taking the actual values of lagged inflation. To run such simulations, the sample period is

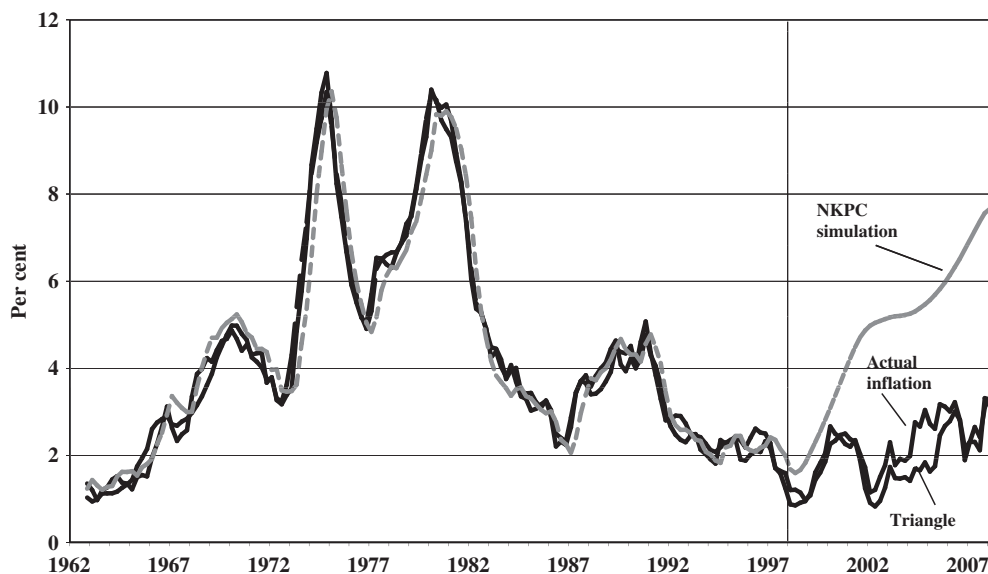


FIGURE 7. Predicted and simulated values of inflation from triangle and NKPC equations 1962(I) to 2007(IV).

truncated ten years before the end of the data interval, and the estimated coefficients through 1997(IV) are used to simulate the performance of the equation for 1998–2007, generating the lagged dependent variables endogenously. Since the simulation has no information on the actual value of the inflation rate, there is nothing to keep the simulated inflation rate from drifting away from the actual rate. These simulations have been criticized because they use the actual values of the explanatory variables other than lagged inflation, but this is an exercise not in forecasting but rather in determining whether a particular set of variables and lags adds to the explanatory power of the equation.

The bottom section of Table 1 displays results of a dynamic simulation for 1998(I) to 2007(IV) based on a sample period that ends in 1997(IV). Two statistics on simulation errors are provided: the mean error (ME) and the root mean-squared error (RMSE). The simulated values of inflation in the triangle model are close to the actual values, with a mean error over 40 quarters of only 0.29, meaning that over the forty quarters the actual inflation rate on average is 0.29 percentage points higher than the predicted value. However, the mean error for the Roberts NKPC model is a huge -2.75 , and as displayed in Figure 7, this error reflects that model's wild overprediction that the inflation rate should have reached nearly 8% by late 2007. The RMSE of the triangle simulation is a bit above the standard error of estimate (SEE) for the 1962–1997 sample period, whereas for the Roberts NKPC model the RMSE is almost three times as large.

The Roberts NKPC and triangle results agree on only one aspect of the inflation process, that the sum of coefficients on the lagged inflation terms is always very close to unity. However, the Roberts coefficients on the unemployment rate are much lower than the triangle coefficients on the unemployment gap. This is an artifact of the exclusion restrictions in the Roberts approach which are statistically rejected in the triangle approach.

The triangle model outperforms the Roberts NKPC model by several orders of magnitude, as displayed in Table 1 and Figure 7. This raises a question central to future research on the US Phillips curve: what are the crucial differences that contribute to the

superior performance of the triangle model? The three key differences are the inclusion in the triangle model of longer lags on both inflation and the unemployment gap, the inclusion of explicit supply shock variables, and the allowance for a time-varying (TV) NAIRU in place of the Roberts NKPC assumption of a fixed NAIRU. In the Appendix we quantify the role of these differences, taking advantage of the fact that the Roberts NKPC model is fully nested in a version of the triangle model that assumes a constant NAIRU. Each exclusion restriction in the Roberts model can be tested by standard statistical exclusion criteria, and every one of the Roberts NKPC exclusion criteria is rejected at high levels of statistical significance.³⁰

Has the PC slope become flatter?

The NKPC research of Roberts and others has concluded that the Phillips curve has become flatter over the past several decades. Yet we have seen that every aspect of the Roberts NKPC specification is rejected at high levels of statistical significance.

Has the Phillips curve flattened? The Roberts NKPC specification says ‘yes’ and the triangle specification says ‘no’. Figure 8 evaluates changes in coefficients by Roberts’ own preferred method (2006, Figure 2, p. 202): rolling regressions that shift the sample period of the regression through time in order to reveal changes in coefficients. The number of quarters in our basic results in Table 1 is 184 (1962(I) to 2007(IV)), and we cut this roughly in half to 90 quarters and run rolling 90-quarter regressions which alternatively start in each quarter from 1962(I) to 1985(III).

As shown in Figure 8, the Roberts NKPC unemployment coefficient rises from -0.17 in 1962 to a peak value of -0.41 in 1974, and then declines back to roughly zero in 1982–84. This appears to support his basic conclusion that the Phillips curve has

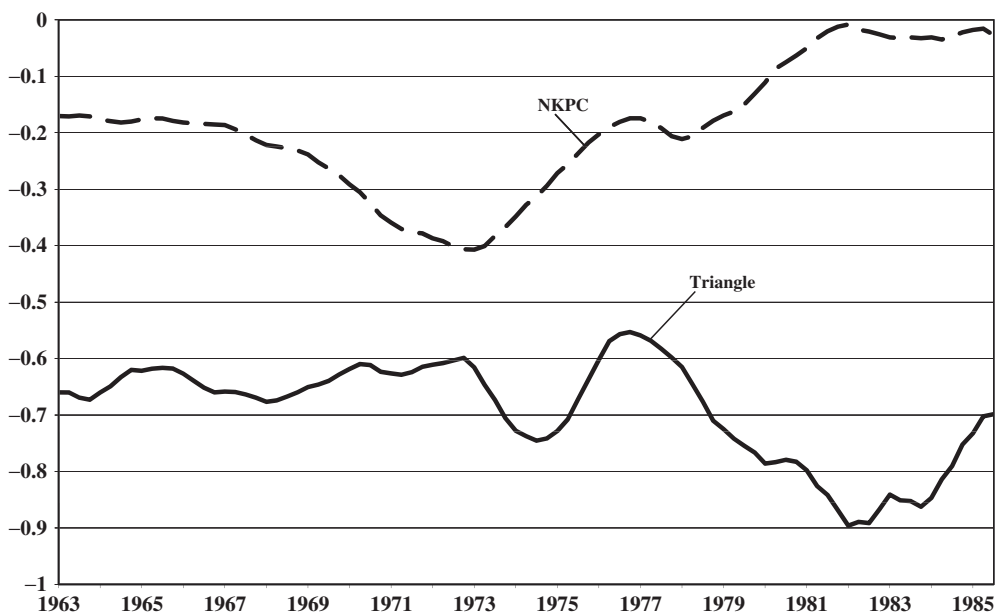


FIGURE 8. NKPC vs triangle unemployment coefficients in 90-quarter rolling regressions from 1962(I) to 1985(III).

flattened. Yet the triangle model reveals no evidence of a decline in the slope of the Phillips curve. The Phillips curve based on the triangle model has a roughly stable PC slope of about -0.6 to -0.7 from 1963 to 1977, and then the slope *rises* towards about -0.7 to about -0.9 in the simulations starting in 1982, then drifts back to -0.7 in the final year. As indicated above, the NKPC slope estimate is biased towards zero by differing amounts in each period, due to the omission of supply shocks.

Has the impact of supply shocks become less important?

Hooker (1996) was among the first to notice that the macroeconomic impact of oil shocks became smaller at some point between the mid-1970s and early 1980s. Since his work, a substantial literature has arisen to debate the sources of the decline in the impact of oil prices on real output and/or inflation. The most obvious cause of the decreased macroeconomic impact of oil originates in the shrinking input of energy in GDP, down by half since 1969.³¹ Kilian (2008) provides a set of reasons that go beyond the declining input share of energy. Part of the answer lies in the role of global demand in causing much of the recent 2004–08 rise in oil prices; a demand-driven increase in oil prices may raise rather than reduce real GDP as would occur in the case of a supply disruption.

Blanchard and Galí (2007) go more deeply into the sources of the declining macroeconomic importance of oil beyond its shrinking input share. Their first reason is that the 1970s oil shocks had a big impact because they were accompanied by other significant shocks that all had the effect of raising the rate of inflation and reducing output. These other shocks, embedded in the triangle inflation equation since the beginning, include adverse food price shocks, the depreciation of the dollar after the breakdown of Bretton Woods, and the unwinding of the Nixon-era price controls. The 2004–08 rise in oil prices had a smaller impact because it was not accompanied as in the 1970s by these other adverse shocks. There was an increase in the real price of food only in 2007–08, and the post-2002 decline in the dollar was not sufficient to cause any sustained increase in the relative price of imports.

Blanchard and Galí also consider two additional factors that may have reduced oil's impact: a decrease in the extent of real wage rigidity and the increased credibility of monetary policy. Using survey data on expected inflation, they show a sharply diminished response of expectations to a given size of oil shock after the mid-1980s, which they attribute to increased credibility.

Does the triangle model confirm a reduced macroeconomic role for oil shocks? Part of oil's impact is disguised in the triangle specification, which enters not the change in the real price of oil, but rather the 'food–energy effect' defined as the difference between headline and core inflation. Thus the declining share of energy input in GDP would cause a reduced response of the food–energy effect to any given change in the real price of oil, not a change in the coefficient on the food–energy effect in the triangle inflation equation itself.

However, it does appear that the coefficient on the food–energy effect has declined, as suggested by the other literature surveyed in this section. Figure 9 plots the sum of coefficients on the food–energy effect in rolling regressions computed by the same technique and for the same time period as for the unemployment coefficient already displayed in Figure 8. Two lines are shown. The first shows changes in the food–energy sum of coefficients in the equation for headline inflation, the same as Table 1. In contrast to a coefficient sum of 0.89 when that sum is held constant over the full 1962–2007 period as in Table 1, Figure 9 shows that the sum of coefficients is at or above 1.0 in 90-quarter regressions from 1962 to 1980, and then the sum declines to about 0.6 in the final 90-quarter

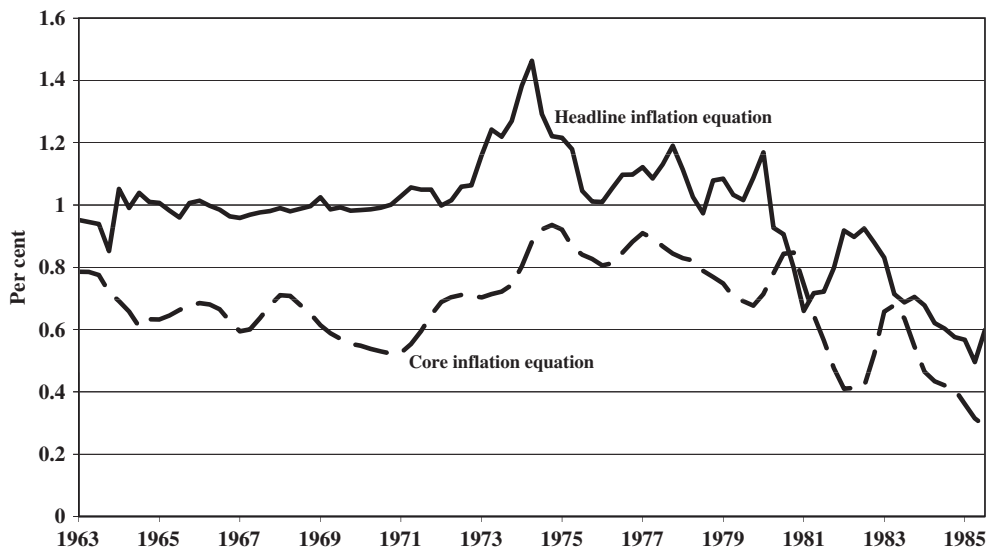


FIGURE 9. Sum of coefficients on food–energy effect in headline and core triangle inflation equations, 90-quarter rolling regressions from 1962(I) to 1985(III).

regression estimated for 1985(III) to 2007(IV). The second line shows a similar decline in the food–energy effect in the core inflation equation, from an average of 0.72 for 90-quarter regressions from 1962 to 1980 down to only 0.3 for the final 90-quarter regression.³²

There is likely to be an interplay between shifts in coefficients and the estimated TV-NAIRU displayed in Figure 5. For instance, the sudden decline in the TV-NAIRU from 5.0% to 4.5% in 2006–07 may be an artifact of the assumed fixity of the food–energy coefficient. If the sum of coefficients on the food–energy effect in 2006–07 were allowed to be smaller, there would be less of a puzzle as to why inflation was so low during that period and hence no reason for the estimation to ‘force’ a decline in the NAIRU. In current research I am looking more closely at changes over time in all the sets of coefficients in the specification of Table 1 and their implication for post-sample simulation performance and the time series behaviour of the TV-NAIRU.

V. CONCLUSION

This paper makes several unique contributions. It contrasts the consensus history of the PC before 1975 with the bifurcated split in PC research since 1975. The evolution of PC doctrine before 1975 is widely accepted and no longer elicits much debate. The discovery by Phillips and his disciples Samuelson and Solow of an inverse relationship between inflation and unemployment briefly suggested an exploitable policy trade-off that was destroyed by the Friedman–Phelps natural rate hypothesis of the late 1960s. Exploitable trade-offs were out, and long-run neutrality was in (it had never disappeared in many environments, including Latin America and the University of Chicago). The econometric models developed in the 1960s to support the policy trade-off were rejected both empirically and logically by Sargent’s trenchant identification argument.

Debates in the early 1970s centred on the models in which Friedman and Phelps had embedded the natural rate hypothesis, and particularly the assumption of arbitrary

barriers that prevented individual workers or agents in general from learning the values of macro data—output, money and prices—provided costlessly by the government. There was also controversy about the implications of the further development of the Friedman–Phelps paradigm by Lucas, who introduced rational expectations into macroeconomics. The Lucas model implied the policy ineffectiveness proposition, which held that anticipated changes in money had no effect on output and were entirely reflected in price changes. Empirical work rejected this framework, showing that monetary surprises had little effect on output, were incapable of explaining the serial correlation of output, and were inconsistent with the persistence of inflation.

After 1975 the PC literature bifurcated into two lines of research, which since then have communicated little with each other. Along the ‘left fork in the road’, the PC was revived by importing micro demand and supply analysis into macroeconomics. There was no assumption that unemployment and inflation are negatively correlated. Demand shocks create an initial and temporary negative correlation, and supply shocks create an initial positive correlation that then evolves according to the policy response. As early as 1975, the theoretical literature on policy responses to supply shocks was developed and showed that adverse supply shocks force policy-makers to choose between higher inflation, lower output, or a combination. By the early 1980s an econometric specification of this AD–AS framework was available that joined demand and supply shocks with long-run neutrality and a strong role for persistence and inertia.

An important difference between the mainstream approach and other post-1975 developments is that the role of past inflation is not limited to the formation of expectations, but also includes pure persistence due to fixed-duration wage and price contracts, and lags between changes in intermediate goods and final product prices. Inflation is dislodged from its past inertial values by demand shocks proxied by the unemployment or output gap, and explicit supply shock variables including changes in the relative prices of food, energy and imports, and the role of changes in trend growth of productivity. The econometric implementation of this approach is sometimes called the ‘triangle’ model, reflecting its three-cornered dependence on demand, supply and inertia.

After 1975, the ‘right fork in the road’ built models in which expectations are not anchored in backward-looking behaviour but can jump in response to current and anticipated changes in policy. Important elements of this literature include policy credibility, models of the game played by policy-makers and private agents forming expectations, and the new Keynesian Phillips curve (NKPC), which derives a forward-looking PC from alternative theories of price stickiness. The common feature of these theories is the absence of inertia, the exclusion of any explicit supply shock variables, the ability of expected inflation to jump in response to new information, and alternative barriers to accurate expectation formation due to such frictions as ‘rational inattention’.

Which post-1975 approach is right? Models in which expectations can jump in response to policy are essential to understanding Sargent’s (1982) ends of four big inflations and other relatively rapid inflations in nations with a history of monetary instability, e.g. Argentina. But the mainstream/triangle approach is the right econometric framework in which to understand the evolution of postwar US inflation. The paper tests the triangle econometric specification alongside one recently published version of the NKPC approach. The latter can be shown to be nested in the former model and to differ by excluding particular variables and lags, and these differences are all rejected by tests of exclusion restrictions. The triangle model outperforms the NKPC variant by orders of magnitude, not only in standard goodness-of-fit statistics, but also in post-sample dynamic simulations. The triangle estimates show that the slope of the Phillips curve has

not become appreciably flatter in the past two decades, a conclusion reached in the NKPC framework due to the specification error of omitting explicit supply shock variables. The triangle estimates do, however, confirm other work indicating that the impact of oil shocks on inflation has diminished.

Thus there are three main interrelated themes in this paper that have not previously received sufficient attention. First, two quite legitimate responses, the left and right forks, occurred after 1975 to the chaotic state of the PC literature at that time. Second, each response is important and helps us to understand how inflation behaves, albeit in different environments. Third, the two approaches need to pay more attention to each other and to engage in a dialogue about which models apply to which episodes, and what factors would motivate a shift in relevance between the alternative models. This paper represents a start towards that reconciliation.

APPENDIX: THE 'TRANSLATION MATRIX' BETWEEN THE ROBERTS NKPC AND TRIANGLE SPECIFICATIONS

Which differences matter in explaining the poor performance of the Roberts NKPC specification in Table 1 and Figure 7? In this appendix we start with the Roberts NKPC specification and gradually change, step by step, to the triangle specification, allowing the NAIRU alternatively to be constant and to vary over time. In everything that follows, the sample period is 1962(I) to 2007(IV).

Table A1 provides the 'translation matrix' that guides us between the Roberts NKPC specification and the triangle specification. There are 24 rows that allow us to trace the role of each specification difference, and the individual rows of alternative specification are evaluated based not just on the SSR measure of goodness of fit, but also on the post-sample simulation performance in 1998–2007 based on coefficient estimates for 1962–97.

We have already seen in Table 1 that the performance of the Roberts NKPC specification for the PCE deflator is inferior to that of the triangle specification by both the criterion of goodness of fit (SSR) and also the less conventional criterion of dynamic simulation performance (ME and RMSE). In Table A1 the basic Roberts variant is in row 1 and the basic triangle variant is in row 21. Roberts' row 1 and the triangle row 21 have SSRs of 244.0 and 64.6, exactly the same as in Table 1.

Table A1 allows the three main differences between the Roberts NKPC and triangle specifications to be evaluated, step by step. Is the crucial difference the longer lags, the supply shocks, the TV-NAIRU, or an interaction of these differences?

In the 24 rows of Table A1, the first 12 rows exclude supply shock variables, and rows 13–24 include the supply shock variables. Scanning down the column for 'SSR', we find that the variants in rows 13–24 including supply shocks all have SSRs below 100, while most of the SSRs that exclude supply shocks have values above 200. Thus our first conclusion is that the exclusion of explicit supply shocks in the Roberts NKPC research is the central reason for its empirical failure either to explain postwar inflation or to track the evolution of inflation in post-sample 1998–2007 simulations. This finding applies to all previous NKPC research, all of which excludes explicit SS variables.

What difference is made by long lags and by the TV-NAIRU? When supply shocks are omitted as in rows 1–12 of Table A1, there is little difference among the alternative variants, which yield SSRs ranging from 183.8 to 244.0. Simulation mean errors (MEs) range from -2.04 to -2.75 when the NAIRU is fixed. Much lower MEs are obtained when the NAIRU is allowed to vary over time.

The results that include supply shocks are displayed in rows 13–24 of Table A1. When supply shocks are included but lag lengths are short, as in rows 13–14, 17–19 and 22, the post-sample simulation errors are very large. When supply shocks are included, the best results are in rows 15–16 with a fixed NAIRU and in rows 20–21 with a TV-NAIRU. Long lags on the dependent variable (inflation) matter in the specification of a PC including supply shocks.

The right-hand section of Table A1 contains a large number of significance tests on the exclusion of variables that are omitted in the Roberts NKPC specification and included in the triangle specification. Starting in row 3, even without supply shock variables, the significance value of excluding lags 9–24 on the lagged dependent variable is 0.01, and on lags 1–4 of the

TABLE A1
 TRANSFORMATION OF THE PHILLIPS CURVE FROM ROBERTS NKPC (ROW 1) TO TRIANGLE (ROW 21), 1962(I) TO 2007(IV)

	Inf lag	U lag length	Fixed or TV NAIRU	Supply shocks	Regression statistics			Simulation errors			Exclude inf lags 5-8			Exclude inf lags 9-24			Exclude U lags 1-4			Exclude supply shocks		
					SEE	SSR	ME	RMSE	F-stat.	Sig. Level	F-stat.	Sig. Level	F-stat.	Sig. Level	F-stat.	Sig. Level	F-stat.	Sig. Level	F-stat.	Sig. Level	F-stat.	Sig. Level
1	1 to 4	0	Fixed	No	0.78	1.17	244.0	-2.75	3.20													
2	1 to 8	0	Fixed	No	0.78	1.17	237.9	-2.73	3.15	1.11	0.35											
3	1 to 24	0	Fixed	No	0.80	1.11	195.0	-2.04	2.41			2.18	0.01									
4	1 5 9 13 17 21	0	Fixed	No	0.78	1.16	238.5	-2.19	2.53													
5	1 to 4	0 to 4	Fixed	No	0.79	1.15	229.2	-2.27	2.68							2.81	0.03					
6	1 to 4	0 to 4	TV	No	0.79	1.13	224.8	0.26	1.43													
7	1 to 8	0 to 4	TV	No	0.79	1.14	222.0	0.36	1.46	0.55	0.70											
8	1 to 24	0 to 4	TV	No	0.80	1.10	188.7	0.15	1.33			1.71	0.05									
9	1 5 9 13 17 21	0 to 4	TV	No	0.79	1.14	225.6	0.03	1.33													
10	1 to 8	0 to 4	Fixed	No	0.78	1.15	226.0	-2.11	2.53	0.61	0.66											
11	1 to 24	0 to 4	Fixed	No	0.81	1.09	183.8	-2.18	2.54			2.21	0.01									
12	1 5 9 13 17 21	0 to 4	Fixed	No	0.80	1.12	216.7	-2.23	2.56													
13	1 to 4	0	Fixed	Yes	0.90	0.78	100.3	-2.45	3.06	0.52	0.72										18.20	0.00
14	1 to 8	0	Fixed	Yes	0.90	0.78	99.0	-2.60	3.15												17.39	0.00
15	1 to 24	0	Fixed	Yes	0.92	0.71	73.8	-0.18	1.00			3.09	0.00								18.31	0.00
16	1 5 9 13 17 21	0	Fixed	Yes	0.92	0.71	82.2	-0.24	0.97												23.85	0.00
17	1 to 4	0 to 4	Fixed	Yes	0.90	0.78	97.3	-2.11	2.72						1.22	0.31					16.79	0.00
18	1 to 4	0 to 4	TV	Yes	0.91	0.76	94.1	-1.14	1.47												17.31	0.00
19	1 to 8	0 to 4	TV	Yes	0.91	0.76	92.1	-1.15	1.40	0.84	0.50										17.12	0.00
20	1 to 24	0 to 4	TV	Yes	0.93	0.65	59.6	0.30	0.74			4.84	0.00								23.66	0.00
21	1 5 9 13 17 21	0 to 4	TV	Yes	0.93	0.64	64.6	0.29	0.70												30.70	0.00
22	1 to 8	0 to 4	Fixed	Yes	0.90	0.78	96.6	-2.23	2.77	0.30	0.88										16.18	0.00
23	1 to 24	0 to 4	Fixed	Yes	0.92	0.69	67.2	-0.50	1.08			3.85	0.00								18.81	0.00
24	1 5 9 13 17 21	0 to 4	Fixed	Yes	0.93	0.68	73.4	-0.56	1.06												23.88	0.00

unemployment gap is 0.03. Throughout rows 1–12 of Table A1, we learn that excluding short lags (e.g. excluding lags 5–8 from equations containing inflation lags 1–4) is insignificant, whereas excluding lags 9–24 yields highly significant exclusion tests.

Rows 13–24, which all include the full set of supply shock variables, differ only in the length of lags included on the lagged dependent (inflation) variable and on lagged unemployment, and also in whether the NAIRU is forced to be fixed or is allowed to vary over time. We can interpret the bottom half of Table A1 by looking at blocks of four rows.

The first group of four rows, 13–16, share in common the inclusion of supply shocks, the assumption of a fixed NAIRU, and alternative lags on the dependent variable. The mean error in the dynamic simulations falls by 80% when lags up to 24 are included, and the exclusion of lags 9–24 is rejected at a 0.00 significance value. The same result occurs in rows 22–25 when with a time-varying NAIRU the significance of long lags on the dependent variable is strongly supported at a significance level of 0.00.

ACKNOWLEDGMENTS

This paper was written on the occasion of the 50th anniversary of Phillips' original article. An earlier version was given as the A. W. H. Phillips Invited Lecture at the Australasian Meetings of the Econometric Society in Wellington, New Zealand, on 9 July 2008. Neil Sarkar was the exemplary research assistant on the empirical section of this paper, and Lucas Zaldueño helped to prepare the manuscript for publication. In writing the early history of the Phillips curve, I found a useful source to be Lacker and Weinberg (2007). I am grateful to Bob Buckle for suggesting the topic and to David Hendry, Giorgio Primiceri and participants in the Northwestern macro workshop for helpful comments.

NOTES

1. The article was unearthed by Jacob Mincer, and as co-Editor of the *Journal of Political Economy* during 1971–73, I was responsible for the decision to reprint it and give it the dramatic title shown in the reference list of this paper.
2. An amusing commentary on the research technology of the 1920s is Fisher's comment that: 'During the last three years in particular I have had at least one computer in my office almost constantly at work on this problem' (1926, p. 786).
3. In Figure 1 the inflation rate is the four-quarter change in the deflator for personal consumption expenditures.
4. Spectators at the time called the *American Economic Review* debate the 'battle of the radio stations', after the AM–FM initials of the protagonists, Ando and Modigliani vs Friedman and Meiselman.
5. Lucas (1973, equation (3), p. 327) created a serially correlated business cycle by introducing a lagged value of cyclical output into the equation explaining cyclical output by the price surprise. This lagged term is gratuitous and neither called for by the theory nor consistent with it.
6. Eckstein and Brinner (1972) produced the first paper to emerge with a specification in which the α coefficient was unity. Gordon (1972) concurred, based on a different parameterization of a time-varying coefficient, and provided comparisons of his approach with those of Eckstein–Brinner and Perry (1970).
7. See also Pakko (2000) for a study of differences in the shape of frequency distributions of inflation and output over the pre- and post-1929 period.
8. While the Gordon and Phelps papers were the first to develop the theory of policy responses to supply shocks, Pierce and Enzler (1974) had previously run simulations with a large-scale econometric model that showed large macroeconomic impacts of commodity price shocks, with a trade-off between the output and inflation effects as the policy response varied.
9. Blinder (1981) subsequently extended the theoretical analysis to allow for rational expectations in the formation of wages and for a distinction between anticipated and unanticipated shocks. The topic was more recently revisited by Ball and Mankiw (1995).
10. The 1978 rival textbooks were by Dornbusch and Fischer and by myself. The dynamic version of the demand–supply model in the form of a first-order difference equation was confined to intermediate level textbooks, with many imitators published soon after. Elementary macro principles textbooks limited themselves to the display of static aggregate demand and supply curves, starting with Baumol and Blinder in 1979.
11. The 'triangle' nomenclature has been picked up by a several authors, including Rudd and Whelan (2005b) and Fitzenberger *et al.* (2009).
12. The theoretical NKE research literature is surveyed and placed in the perspective of historical puzzles of price and wage behaviour in Gordon (1990).

13. The study of wage and price control effects began with Gordon (1972). Nixon-era control dummies were included in all the econometric tests of the Gordon mainstream model, beginning with Gordon (1977b). The current version of the triangle model (see Table 1) estimates that the imposition of controls reduced the price level by a cumulative 1.6%, and the removal of the controls raised the price level by 1.8%. Using CPI rather than PCE data, and with a slightly different method, Blinder and Newton (1981, Table 2, Model 1) estimated that the controls held down the price level by a cumulative 3.1% through early 1974, followed by a 3.2% bounce-back in 1974–75.
14. The concept of the sacrifice ratio was developed by Okun (1978), and his preferred estimate of the sacrifice ratio was 10 (p. 348).
15. In a paper written in early 1981, Gordon (1982b, Table 10) integrated an endogenous flexible exchange rate into the effects of a hypothetical Volcker disinflation and simulated a sacrifice ratio of 4.8, a relatively low number that resulted in part from a 33% exchange rate appreciation.
16. The pre-1994 estimate of the NAIRU incorporated changes in the difference between the official unemployment rate and the demographically-adjusted unemployment rate, originally introduced by Perry (1970). The Perry-weighted NAIRU was assumed to be fixed.
17. Blanchard (2008, p. 8) uses the ‘workhorse’ label for a small three-equation macro model of which the NKPC is one equation, even though he calls the NKPC equation ‘patently false’ (p. 9). We return in Section IV to quantify the extent to which the NKPC is ‘patently false’ and leave to the reader the puzzle as to how this model could have become a ‘workhorse’.
18. The insight that feedback between wage and price equations is transmitted through labour’s share dates back to Franz and Gordon (1993).
19. They use the same list of instruments as in Galí and Gertler (1999), with two lags instead of four lags on each.
20. Labour’s share in Figure 4 is defined as employee compensation divided by domestic net factor income (www.bea.gov, NIPA Table 1.10: GDP minus consumption of fixed capital minus taxes on production and imports less subsidies).
21. See Del Boca *et al.* (2008) for a wide-ranging econometric assessment of the inflation rate of the Italian lira from 1861 to 1998. They conclude that Italy has had a ‘conventional inflation–output trade-off only during times of low inflation and stable aggregate supply’.
22. All four price control dummy variables, as well as the NRA dummy, are entered in the form of ‘on effects’ summing to 1.0 followed by ‘off effects’ summing to -1.0 , implying no permanent impact of the controls. The coefficients on these dummy variables measure the cumulative impact of the controls in displacing the price level prior to the reverse snap-back effect.
23. This triangle or three-term PC equation, with each term including lags and a lag operator, was introduced in Gordon and King (1982, equation 13). The notation has been used by many authors in the mainstream ‘left-fork’ tradition, most recently by Blinder and Rudd (2008, unnumbered equation, p. 73)
24. This method of estimating the TV-NAIRU was introduced in simultaneous papers by Gordon (1997) and Staiger *et al.* (1997).
25. The relative import price variable is defined as the rate of change of the non-food non-oil import deflator minus the rate of change of the core PCE deflator. The relative food–energy variable is defined as the difference between the rates of change of the overall PCE deflator and the core PCE deflator. The Nixon-era control variables and the lag lengths (shown explicitly in Table 1) remain the same as originally specified in Gordon (1982b). The productivity variable is the eight-quarter change in a Hodrick–Prescott filtered trend of the change in non-farm private business output per hour (using 6400 as the Hodrick–Prescott smoothness parameter).
26. In subsequent work, Staiger *et al.* (2001) and Stock and Watson (2007) abandoned the attempt to estimate the TV-NAIRU as a byproduct of the inflation equation, and Blinder and Rudd (2008) follow their lead. Now they estimate the NAIRU as a trend on the actual unemployment rate, taking no account of the role of supply shocks in making this trend unusually high in 1974–75 and 1981–82, or unusually low in 1997–2000. As a result, their version of the unemployment gap ($U - U^N$) greatly understates the size of that gap and its influence on inflation during the key supply shock episodes.
27. Basistha and Nelson (2007) are among those authors who exclude explicit supply shock variables from their equations and derive estimates of the TV-NAIRU that are extremely high, e.g. 8% in 1975 and 10% in 1981 (2007, Figure 6, p. 509).
28. The inclusion of lags 13–24 (years four to six) in an exclusion test is strongly significant at the 0.00 confidence level. As stated in the notes to Table 1, we conserve on degrees of freedom by including six successive four-quarter moving averages of the lagged dependent variable at lags 1, 5, 9, 13, 17 and 21, rather than including all 24 lags separately.
29. This can be compared to an import share in nominal GDP of 10% at the midpoint of the sample period in 1985.
30. Dew-Becker (2006) has previously traced the statistical significance of stripped-down Phillips curves and reached conclusions that are similar to those arrayed in Appendix Table A1.
31. Blinder and Rudd (2008, Figure 19) plot a ratio of British Thermal Units (BTUs) to real GDP that declines from 18 in 1969 to 8.7 in 2007.

32. The core inflation equation is specified exactly as the headline equation in Table 1, with the exception of the food–energy effect. To allow for longer lags in the impact of energy on core inflation, the food–energy effect is measured as the four-quarter moving average of the difference between headline and core inflation, and is entered at lags 1, 5 and 9.

REFERENCES

- BALL, L. (2008). Hysteresis in unemployment. Paper presented at Federal Reserve Bank of Boston Conference, 10–11 June.
- and MANKIW, G. N. (1995). Relative-price changes as aggregate supply shocks. *Quarterly Journal of Economics*, **110**, 161–93.
- BÅRDSEN, G., JANSEN, E. S. and NYMOEN, R. (2002). Testing the new Keynesian Phillips curve. Paper presented at Workshop on the Phillips curve: new theory and evidence, Facföreningsrörelsens Institut För Ekonomisk Forskning (Trade Union Institute For Economic Research), Stockholm, 25–6 May.
- BARRO, R. J. (1977). Unanticipated money growth and unemployment in the United States. *American Economic Review*, **67**, 101–15.
- BARSKY, R. B. and KILIAN, L. (2002). Do we really know that oil caused the great stagflation? A monetary alternative. *NBER Macroeconomics Annual 2001*, 137–83.
- BASISTHA, A. and NELSON, C. R. (2007). New measures of the output gap based on the forward-looking new Keynesian Phillips curve. *Journal of Monetary Economics*, **54**, 498–511.
- BERNANKE, B. S., GERTLER, M. and WATSON, M. (1997). Systematic monetary policy and the effects of oil price shocks. *Brookings Papers on Economic Activity*, **28**(1), 91–142.
- BLANCHARD, O. J. (1987). Aggregate and individual price adjustment. *Brookings Papers on Economic Activity*, **18**(1), 57–109.
- (2008). The state of macro. NBER Working Paper no. 14259.
- and GALI, J. (2007). The macroeconomic effects of oil shocks: why are the 2000s so different from the 1970s? NBER Working Paper no. 13368.
- and SUMMERS, L. (1986). Hysteresis and the European unemployment problem. *NBER Macro Annual 1985*, 15–77.
- BLINDER, A. S. (1979). *Economic Policy and the Great Stagflation*. New York: Academic Press.
- (1981). Monetary accommodation of supply shocks under rational expectations. *Journal of Money, Credit, and Banking*, **13**, 425–38.
- (1982). The anatomy of double-digit inflation in the 1970s. In R. E. Hall (ed.), *Inflation: Causes and Effects*. Chicago: University of Chicago Press for NBER, pp. 261–82.
- and NEWTON, W. J. (1981). The 1971–74 controls program and the price level: an econometric post-mortem. *Journal of Monetary Economics*, **8**, 1–23.
- and RUDD, J. B. (2008). The supply-shock explanation of the great stagflation revisited. NBER Working Paper no. 14563.
- BOSWORTH, B. P. and LAWRENCE, R. Z. (1982). *Commodity Prices and the New Inflation*. Washington DC: Brookings Institution.
- BRUNO, M. and SACHS, J. D. (1985). *Economics of Worldwide Stagflation*. Cambridge, MA: Harvard University Press.
- CALVO, G. A. (1983). Staggered price in a utility maximizing framework. *Journal of Monetary Economics*, **12**, 383–98.
- CASSIDY, J. (1996). The decline of economics. *New Yorker*, 2 December, pp. 54–7.
- COGLEY, T. and SARGENT, T. J. (2005). The conquest of U.S. inflation: learning and robustness to model uncertainty. *Review of Economic Dynamics*, **8**, 528–63.
- DEL BOCA, A., FRATIANNI, M., SPINELLI, F. and TRECROCI, C. (2008). The Phillips curve and the Italian lira, 1861–1998. University of Brescia MoFiR working paper no. 8.
- DEW-BECKER, I. (2006). Was Lucas right? Evidence on structural shift in the Phillips curve. Senior honours thesis submitted to Northwestern Economics Department, May.
- and GORDON, R. J. (2005). Where did the productivity growth go? Inflation dynamics and the distribution of income. *Brookings Papers on Economic Activity*, **36**, 67–127.
- ECKSTEIN, O. and BRINNER, R. (1972). The inflation process in the United States. A study prepared for the use of the Joint Economic Committee, 92 Congress 2 session.
- FISCHER, S. (1977). Long-term contracts, rational expectations, and the optimal money supply rule. *Journal of Political Economy*, **85**, 191–205.
- FISHER, I. (1926). A statistical relation between unemployment and price changes. *International Labor Review*, **13**, 785–92. (Reprinted as Fisher, I. (1973). I discovered the Phillips curve. *Journal of Political Economy*, **81**, 496–502.)

- FITZENBERGER, B., FRANZ, W. and BODE, O. (2009). The Phillips curve and NAIRU revisited: new estimates for Germany. *Journal of Economics and Statistics*, forthcoming.
- FRANZ, W. and GORDON, R. J. (1993). Wage and price dynamics in Germany and America: differences and common themes. *European Economic Review*, **37**, 719–54.
- FRIEDMAN, M. (1968). The role of monetary policy. *American Economic Review*, **58**, 1–17.
- FROMM, G. and TAUBMAN, P. (1968). *Policy Simulations with an Econometric Model*. Washington: Brookings Institution.
- FUHRER, J. C. (1997). The (un)importance of forward-looking behavior in price specifications. *Journal of Money, Credit, and Banking*, **29**, 338–50.
- GALÍ, J. and GERTLER, M. (1999). Inflation dynamics: a structural econometric analysis. *Journal of Monetary Economics*, **44**, 195–222.
- , ——— and LOPEZ-SALIDO, D. (2005). Robustness of estimates of the hybrid new Keynesian Phillips curve. *Journal of Monetary Economics*, **52**, 1107–18.
- GORDON, R. J. (1972). Wage–price controls and the shifting Phillips curve. *Brookings Papers on Economic Activity*, **3**(2), 385–421.
- (1975). Alternative responses of policy to external supply shocks. *Brookings Papers on Economic Activity*, **6**(1), 183–206.
- (1977a). The theory of domestic inflation. *American Economic Review, Papers and Proceedings*, **67**, 128–34.
- (1977b). Can the inflation of the 1970s be explained? *Brookings Papers on Economic Activity*, **8**(1), 765–78.
- (1982a). Price inertia and policy ineffectiveness in the United States, 1890–1980. *Journal of Political Economy*, **90**, 1087–117.
- (1982b). Inflation, flexible exchange rates, and the natural rate of unemployment. In M. N. Baily (ed.), *Workers, Jobs, and Inflation*. Washington: Brookings, pp. 89–158.
- (1984). Supply shocks and monetary policy revisited. *American Economic Review, Papers and Proceedings*, **74**, 38–43.
- (1990). What is new-Keynesian economics? *Journal of Economic Literature*, **28**, 1115–71.
- (1997). The time-varying NAIRU and its implications for economic policy. *Journal of Economic Perspectives*, **11**, 11–32.
- (1998). Foundations of the Goldilocks economy: supply shocks and the time-varying NAIRU. *Brookings Papers on Economic Activity*, **29**, 297–333.
- and KING, S. R. (1982). The output cost of disinflation in traditional and vector autoregressive models. *Brookings Papers on Economic Activity*, **13**, 205–42.
- HAMILTON, J. D. (1983). Oil and the macroeconomy since World War II. *Journal of Political Economy*, **91**, 228–48.
- HOOKER, M. A. 1996. What happened to the oil price macroeconomy relationship? *Journal of Monetary Economics*, **38**, 195–213.
- HUMPHREY, T. M. (1991). Nonneutrality of money in classical monetary thought. *Federal Reserve Bank of Richmond Economic Review*, **77**, 3–15.
- KATZ, L. F. and KRUEGER, A. B. (1999). The high-pressure U.S. labor market of the 1990s. *Brookings Papers on Economic Activity*, **30**, 1–65.
- KILEY, M. T. (2007). An empirical comparison of sticky-price and sticky-information models of price setting. *Journal of Money, Credit, and Banking*, **39**, 101–25.
- KILIAN, L. (2008). The economic effects of energy price shocks. *Journal of Economic Literature*, **46**, 871–909.
- KYDLAND, F. E. and PRESCOTT, E. C. (1977). Rules rather than discretion: the inconsistency of optimal plans. *Journal of Political Economy*, **85**, 473–91.
- LACKER, J. M. and WEINBERG, J. A. (2007). Inflation and unemployment: a layperson’s guide to the Phillips curve. *Federal Reserve Bank of Richmond Economic Quarterly*, **93**, 201–27.
- LAIDLER, D. and PARKIN, M. J. (1975). Inflation: a survey. *Economic Journal*, **85**, 741–809.
- LIPSEY, R. G. (1960). The relation between unemployment and the rate of change of money wage rates: a further analysis. *Economica*, **27**, 1–31.
- LUCAS, R. E. Jr. (1972). Expectations and the neutrality of money. *Journal of Economic Theory*, **4**, 103–24.
- (1973). Some international evidence on output–inflation tradeoffs. *American Economic Review*, **63**, 326–34.
- and SARGENT, T. J. (1978). After Keynesian macroeconomics. In *After the Phillips Curve: Persistence of High Inflation and High Unemployment*. Boston: Federal Reserve Bank of Boston, pp. 49–72.
- MANKIW, G. N. (2001). The inexorable and mysterious tradeoff between inflation and unemployment. *Economic Journal*, **111**, 45–61.
- MISHKIN, F. S. (1982). Does anticipated monetary policy matter? An econometric investigation. *Journal of Political Economy*, **90**, 22–51.

- OKUN, A. M. (1978). Efficient disinflationary policies. *American Economic Review, Papers and Proceedings*, **68**, 348–52.
- ORPHANIDES, A. and WILLIAMS, J. C. (2005). Inflation scares and forecast-based monetary policy. *Review of Economic Dynamics*, **8**, 498–527.
- PAKKO, M. R. (2000). The cyclical relationship between output and prices: an analysis in the frequency domain. *Journal of Money, Credit, and Banking*, **32**, 382–99.
- PERRY, G. L. (1970). Changing labor markets and inflation. *Brookings Papers on Economic Activity*, **1**, 411–41.
- PHELPS, E. S. (1967). Phillips curves, expectations of inflation and optimal unemployment over time. *Economica*, **34**, 254–81.
- (1968). Money–wage dynamics and labor–market equilibrium. *Journal of Political Economy*, **76**, 678–711.
- (1978). Commodity-supply shock and full-employment monetary policy. *Journal of Money, Credit, and Banking*, **10**, 206–21.
- PHILLIPS, A. W. (1958). The relation between unemployment and the rate of change of money wage rates in the United Kingdom, 1861–1957. *Economica*, **25**, 283–99.
- PIERCE, J. L. and ENZLER, J. J. (1974). The effects of external inflationary shocks. *Brookings Papers on Economic Activity*, **1**, 13–54.
- PRIMICERI, G. (2006). Why inflation rose and fell: policymakers’ beliefs and US postwar stabilization policy. *Quarterly Journal of Economics*, **121**, 867–901.
- REIS, R. (2006). Inattentive consumers. *Journal of Monetary Economics*, **53**, 1761–800.
- ROBERTS, J. M. (2006). Monetary policy and inflation dynamics. *International Journal of Central Banking*, **2**, 193–230.
- ROMER, C. D. (1996). Inflation and the growth rate of output. NBER Working Paper no. 5575.
- (1999). Why did prices rise in the 1930s? *Journal of Economic History*, **59**, 167–99.
- RUDD, J. and WHELAN, K. (2005a). New tests of the new-Keynesian Phillips curve. *Journal of Monetary Economics*, **52**, 1167–81.
- and ——— (2005b). Modelling inflation dynamics: a critical review of recent research. Finance and Economics Discussion Series 2005–19, Board of Governors of the Federal Reserve System.
- SAMUELSON, P. A. and SOLOW, R. M. (1960). Analytical aspects of anti-inflation policy. *American Economic Review, Papers and Proceedings*, **50**, 177–94.
- SARGENT, T. J. (1982). The ends of four big inflations. In R. E. Hall (ed.), *Inflation*. Chicago: University of Chicago Press, pp. 41–97.
- (1999). *The Conquest of American Inflation*. Princeton, NJ: Princeton University Press.
- and WALLACE, N. (1975). ‘Rational’ expectations, the optimal monetary instrument, and the optimal money supply rule. *Journal of Political Economy*, **83**, 241–54.
- , WILLIAMS, N. and ZHA, T. (2006). Shocks and government beliefs: the rise and fall of American inflation. *American Economic Review*, **96**, 1193–224.
- SIMS, C. A. (1987). Comment. *Brookings Papers on Economic Activity*, **1**, 117–20.
- (2006). Rational inattention: beyond the linear-quadratic case. *American Economic Review*, **97**, 586–606.
- (2008). Inflation expectations, uncertainty, the Phillips curve, and monetary policy. Paper presented at Federal Reserve Bank of Boston Conference, 10–11 June.
- STAIGER, D., STOCK, J. H. and WATSON, M. W. (1997). The NAIRU, unemployment, and monetary policy. *Journal of Economic Perspectives*, **11**, 33–49.
- , ——— and ——— (2001). Prices, wages, and the U.S. NAIRU in the 1990s. In A. B. Krueger and R. M. Solow (eds), *The Roaring Nineties: Can Full Employment Be Sustained?* New York: Russell Sage Foundation and Century Foundation Press, pp. 3–60.
- STOCK, J. H. and WATSON, M. W. (2007). Why has U.S. inflation become harder to forecast? *Journal of Money, Credit, and Banking*, **39**, 3–34.
- TAYLOR, J. (1980). Aggregate dynamics and staggered contracts. *Journal of Political Economy*, **88**, 1–24.
- WOODFORD, M. (2003). *Interest and Prices*. Princeton, NJ: Princeton University Press.