

## Is Hysteresis Important for U.S. Unemployment?

John M. Roberts and Norman J. Morin  
Board of Governors of the Federal Reserve System  
Washington, D.C. 20551

December 1999

**Abstract:** We look for evidence of “hysteresis” in the U.S. unemployment rate — that is, that current labor-market outcomes can affect the future equilibrium level of unemployment. We first examine (using a variety of econometric tests for unit roots) whether the unemployment rate tends to come back to a long-run average over time. On balance, our results suggest that the unemployment rate tends to return to a long-run value, ruling out the possibility of *permanent* hysteresis. We look for evidence of *temporary* hysteresis by examining whether lagged unemployment enters a standard Phillips-curve model of U.S. inflation. We find weak evidence in support of temporary hysteresis, but the effect is not very large, suggesting that hysteresis is not very important for U.S. unemployment.

We are grateful to Carol Corrado, Dan Sichel, Dave Stockton, Sandy Struckmeyer, Bill Wascher, and seminar participants at the Federal Reserve Board for helpful comments. The views expressed in this paper are those of the authors and do not necessarily reflect the views of the Board of Governors of the Federal Reserve System or other members of its staff.

The unemployment rate recently hit its lowest level in twenty-eight years. As a consequence, many workers who would otherwise not be gaining labor market experience have the opportunity to work. Because workers with experience are preferred by employers, a natural question to ask is, will there be lasting benefits of the extra labor market experience workers are gaining in today's exceptional labor market?

The phenomenon of current labor market outcomes affecting the future labor market equilibrium has been labeled "hysteresis." With respect to the unemployment rate, the term has been used to mean that past experience of high unemployment will raise the rate of unemployment consistent with stable inflation (the "NAIRU"), and that past experience of low unemployment will lower the NAIRU. Hysteresis can also apply to labor force participation, when workers drawn into the labor force during a boom have higher labor force participation in later years than they would have had had the boom never occurred.

There are a number of mechanisms whereby past labor market experience could affect current labor market outcomes. One is the human capital story just mentioned: Unemployed people with recent labor market experience are more employable than those without recent experience. While employment may confer skills, skills gained during employment may erode during a period of unemployment. If the unemployment rate has been high in the recent past, a higher proportion of the unemployed may have suffered such skill erosion and so will be less employable, raising the NAIRU.

Another reason current unemployment may have lasting effects on labor market equilibria is that a prolonged period of high unemployment rates may reduce the social stigma of unemployment: If your neighbors have jobs, you may feel greater social pressure to find a job yourself than if your neighbors are unemployed.

Blanchard and Summers (1986), who introduced the term hysteresis to the discussion of employment dynamics, emphasized a model in which union behavior can lead to hysteresis. This model, sometimes referred to as the "insider-outsider" model, was originally suggested by Lindbeck and Snower (1988). Because it relies heavily on the behavior of labor unions, the model

may not be very important for the United States, but may be more important in Europe, where unions are more influential.<sup>1</sup>

The effects of hysteresis on the NAIRU may be either temporary or permanent. For example, the experience of high unemployment during a recession lowers the employability of those who have experienced unemployment during the recession. But several years later, it is plausible that those who worked during the recession and those who did not do not differ greatly in their employability (assuming that both have worked more recently). That is, losing a job in a recession may not lead to a permanent stain on one's employment record. On the other hand, Blanchard and Summers (1986) emphasized the possibility that the hysteresis generated by the insider-outsider hypothesis may be permanent. And social stigma may also have permanent effects, as a "culture" of unemployment may develop in some neighborhoods.

Our focus in this paper is on hysteresis in the unemployment rate. However, there may also be hysteresis in labor force participation. In our view, it is natural to separate the analysis of unemployment from that of labor force participation. One reason is that there are important testable implications of hysteresis in the unemployment rate within the context of the Phillips curve, where there are not natural implications for labor-force participation. And a thorough treatment of labor force participation involves consideration of demographic developments that are of only secondary importance in discussing unemployment.

In section one, we show the implications of hysteresis for the Phillips curve in general terms. We then switch to empirical work, beginning with an examination of the long-run properties of the unemployment rate. A necessary condition for permanent hysteresis is that there is a permanent component to shifts in the U.S. unemployment rate.<sup>2</sup> We therefore draw on the large econometric literature on testing for such permanent effects of shocks to a time series (often

---

<sup>1</sup>Røed (1997) surveys other potential explanations for hysteresis.

<sup>2</sup>Nonetheless, permanent shifts in the unemployment rate are not a sufficient condition for permanent hysteresis, since there may be permanent shifts that are not caused by the previous experience of high unemployment — for example, if there are changes to the minimum wage or to unemployment insurance. Nonetheless, determining whether there are permanent shifts in the unemployment rate is an important first step, since the presence of a permanent component will affect subsequent statistical analysis. Note that deterministic breaks in the unemployment rate are also inconsistent with hysteresis, since under hysteresis, it is the history of the unemployment rate itself that affects labor market equilibrium.

referred to as “unit root tests”). We apply a variety of such tests to the U.S. unemployment rate. Consistent with earlier work, our test results are superficially ambiguous: Some tests allow us to reject the presence of permanent components in unemployment while others do not. But a careful examination of the results suggests that there have *not* been permanent shocks to unemployment: In particular, we argue that test results that do not reject the permanent component are based on specifications that include too many regressors, reducing the econometric power of the tests.

Our finding that the unemployment rate in the United States tends to come back over time to a long-run average level rules out the possibility of *permanent* hysteresis. However, it is still possible that there is an important degree of *temporary* hysteresis. Under temporary hysteresis, the recent history of the unemployment rate causes the NAIRU to deviate temporarily from its long-run level.<sup>3</sup> We look for evidence of temporary hysteresis through estimates of an empirical Phillips curve. The Phillips curve is a natural place to look for hysteresis, because if lagged unemployment rates help lower the NAIRU, they should enter the Phillips curve with the opposite sign of the current unemployment rate.

As we show below, it can be difficult to disentangle temporary hysteresis from other sources of lagged effects of unemployment on inflation. We find that hysteresis appears to be associated only with the effects of *long* lags of the unemployment rate, of from two to six years. Even then, our results are mixed. The sum of coefficients on long lags of past unemployment is often insignificant; even when it is statistically significant, it is only marginally so. The results are sensitive to the measure of inflation used in the Phillips curve and on details of the specification of the Phillips curve. When the long lag of past unemployment is of the correct sign, the implied historical variation in the NAIRU is modest in size, with the NAIRU varying between 5 and 7 percent for most models.

The results imply that the NAIRU is currently not very far from its long-run average. But temporary hysteresis implies that, at least for the time being, the longer we stay at current low levels of unemployment, the lower the temporary NAIRU will become. However, the effects are

---

<sup>3</sup>Some would argue that the term “temporary hysteresis” is an oxymoron, since hysteresis was originally used to indicate permanent physical changes caused by passing phenomena. However, in economics, the term has come to be associated with the effects of the past on equilibrium, even when that effect is temporary. See, for example, Blanchard and Summers, 1986, p. 17.

not likely to be very large. For example, the results suggest that spending one year with the unemployment rate at 4-1/2 percent rather than 5-1/2 percent will reduce the NAIRU for the next several years by at most 0.2 percentage point, with most models that find a significant effect putting the estimate closer to 0.1 percentage point. But because hysteresis is temporary, the effects on the NAIRU will not be permanent. And, according to the model, that temporary benefit comes at the risk of permanently higher inflation.

### ***1. Hysteresis and the Phillips Curve***

A simple way to express that the NAIRU is affected by past unemployment rates is:

$$NR_t = \alpha(L)RU_{t-1} + [1-\alpha(1)]NR^*, \quad (1.1)$$

where  $NR_t$  is the current level of the unemployment rate consistent with steady inflation (the NAIRU),  $RU$  is the unemployment rate,  $\alpha(L)$  is a lag polynomial,  $\alpha(1)$  is the sum of the coefficients in the lag polynomial, and  $NR^*$  is the “long-run” NAIRU. If  $\alpha(1) = 1$ , then the NAIRU is a function of past unemployment only, the most extreme form of hysteresis. In this case, there will be permanent shifts in the NAIRU and in the unemployment rate itself, and no stable, long-run NAIRU. If  $\alpha(1) < 1$ , then the effect of recent labor market conditions on the NAIRU is only temporary, and the NAIRU will eventually gravitate to the long-run natural rate,  $NR^*$ .

Because the NAIRU is a key element of the Phillips curve, hysteresis has important implications for the Phillips curve. Consider the standard empirical Phillips curve:

$$\Delta p_t = \lambda(L) \Delta p_{t-1} + \gamma [NR_t - RU_t] + \beta X_t + \varepsilon_t \quad (1.2)$$

where  $\Delta p$  is inflation,  $\gamma$  is a constant greater than zero,  $\varepsilon$  is an error term, and  $X$  is a vector of supply shocks, each element of which has mean zero. Consistent with the evidence that inflation in the United States has not been stable over time, we impose the restriction that the sum of the coefficients on lagged inflation is one, which can be expressed as  $\lambda(1) = 1$ . The equation has the standard Phillips curve implication that when unemployment is below the natural rate, inflation increases.

When we substitute the hysteresis formulation of the NAIRU into the Phillips curve, we obtain:

$$\Delta p_t = \lambda(L) \Delta p_{t-1} - \gamma RU_t + \gamma \alpha(L) RU_{t-1} + \gamma [1-\alpha(1)] NR^* + \beta X_t + \varepsilon_t \quad (1.3)$$

So, if there is hysteresis in the unemployment rate, lagged unemployment will enter the Phillips curve, but with a positive sign, the opposite of that on contemporary unemployment. If the extreme version of hysteresis holds, so that  $\alpha(1) = 1$ , then the sum of the lagged coefficients will be equal and opposite to the coefficient on contemporaneous unemployment. In that case, the model can be rewritten so that only changes in unemployment appear.

One complication to the above model is that lagged unemployment may enter the Phillips curve for reasons other than hysteresis. For example, an implication of Taylor's (1979, 1980) staggered contracts model of the Phillips curve is that, because old contracts are still in effect, lagged unemployment will enter the equation as well as current unemployment. Furthermore, relative prices of some less-processed materials move strongly with economic activity. This phenomenon — sometimes referred to as the “speed” effect — may mean that the change in the unemployment rate enters the Phillips curve in addition to the level of the unemployment rate.

Modifying the model to allow for these near-term lagged effects implies:

$$\Delta p_t = \lambda(L) \Delta p_{t-1} + \gamma(1) NR_t - \gamma(L) RU_t + \beta X_t + \varepsilon_t, \quad (1.4)$$

where the notation  $\gamma(L) RU_t$  indicates that lags as well as the current level of RU enters the model. Substituting in equation 1.1 yields:

$$\Delta p_t = \lambda(L) \Delta p_{t-1} + \gamma(1) \alpha(L) RU_{t-1} - \gamma(L) RU_t + \beta X_t + \varepsilon_t. \quad (1.5)$$

Equation 1.5 suggests a potential identification problem, since lags in the unemployment rate appear not only on account of “hysteresis” but also because of near-term lags. If the near-term lags appear on account of “speed,” then it is not very important to distinguish them from hysteresis because although “speed” effects have a different economic motivation than hysteresis, their effect is mathematically equivalent: Both imply that lags of the unemployment rate will enter the Phillips curve with coefficients that are, on net, positive.

However, if there are lagged effects of the level of unemployment, as in the staggered contracts model, near-term lags in the unemployment rate may enter with the same sign as the current-period unemployment rate. If these coefficients were larger than coefficients on the lags implied by hysteresis, then the sum of coefficients on lagged unemployment could be negative, implying nonsensical results if interpreted as hysteresis. In this case, it may be useful to distinguish between short-term unemployment dynamics, which may be affected by factors such as staggered contracts that would imply negative coefficients, and longer-term lags, which are

unlikely to be affected by such considerations. Of course, identification in this case depends on lagged effects due to hysteresis being longer than lags from any near-term dynamics.<sup>4</sup>

## ***2. Have There Been Permanent Shifts in the U.S. Unemployment Rate?***

It is often important to determine whether the effects of shocks to a time series are permanent. If the long run effects of a shock are permanent, the series is said to be a “unit root” process or to contain a “permanent component,” and, for many purposes, standard statistical tools do not apply. Thus, prior to using the unemployment rate in a Phillips curve, it is important to decide whether it is a unit root process. If changes in the unemployment rate are indeed permanent, this would imply that there is no unemployment rate to which, in the absence of further innovations, the series would revert. That is, there would be no strict “natural rate” of unemployment.

### *2.1 — The augmented Dickey-Fuller test*

The most commonly used test for a unit root is the augmented Dickey-Fuller (ADF) test, originally developed by Dickey and Fuller (1979). In that test, the change in the variable in question is regressed on the lagged level of the variable and lagged changes in the variable. Test results have been derived so that a constant term and time trends may also be included, although we only consider the case of a constant term,

$$\Delta x_t = -\theta x_{t-1} + \varphi(L) \Delta x_{t-1} + d_t + u_t, \quad (2.1)$$

where  $x$  is the variable we are testing,  $\Delta x$  is its difference,  $d$  is the “deterministic” part, such as a constant or time trend, and  $u$  is an error term. The commonly used ADF test is a significance test on  $\theta$ ; if  $\theta$  is significantly different from zero, then  $x$  tends to come back to its deterministic trend over time, and so  $x$  is stationary. Because  $x$  is nonstationary under the null hypothesis, the distribution of  $\theta$  will be nonstandard. Fuller (1976) presents tables for determining the statistical significance of the t-ratio of  $\theta$ .

---

<sup>4</sup>One particular model of hysteresis, the insider-outsider model, has predictions about the role of long-term unemployment in the Phillips curve. The appendix explores these implications further. The results presented there are negative for a special role for long-term unemployment in the Phillips curve and thus for the insider-outsider model of hysteresis.

## 2.2 — *Choosing the lag length in the ADF and other tests*

The purpose of including lagged changes in the augmented Dickey-Fuller test is to render the error term white noise, since the tables for the distribution of  $\theta$  assume white-noise errors. However, choosing too many lags can reduce the statistical power of the test, leading us to fail to reject the unit root when there is in fact no unit root. Hence, choosing the number of lagged differences to include in the test is an important issue.

In many studies, lag lengths have been imposed based on the frequency of the data (four lagged differences for quarterly data, twelve for monthly data). Campbell and Perron (1991) suggest a strategy of specifying a large number of lagged differences and sequentially testing down until a significant lag is found via its  $t$ -statistic or via a likelihood ratio test — the so-called general-to-simple approach. Alternatively, a simple-to-general approach starts with no lagged differences and sequentially adds them until a test for residual serial correlation fails to reject no serial correlation. A final strategy is to choose a large maximum lag length and use a model selection criteria, such as the Akaike information criterion (AIC) or Schwartz criterion, to determine the number of lagged differences.

As is clear from equation 2.1, the adjustment for serial correlation with the ADF procedure implicitly fits the serial correlation of the data with an autoregressive (AR) model. Numerous studies have shown that among the generally used model selection procedures, the Schwartz criterion does the best at choosing the correct model when that model is a simple autoregressive one. So it should come as no surprise that several studies, including Elliot, Rothenberg, and Stock (1992), Hall (1994), and Ng and Perron (1995), have found that if the true model is autoregressive with continuous lags, the Schwartz criterion does better than the others, in that it rejects a unit root just about the “right” percent of the time (that is, 5 percent of the time for a test with a 5 percent confidence level), while having greater power than these other criteria.

However, these studies have found that choosing the number of lagged differences to include using the Schwartz criterion leads to over-rejection of a unit root when the true model includes a moving average (MA) process with strong negative autocorrelation. Intuitively, that’s because an MA process with a large negative coefficient is poorly approximated by a low-dimensional AR process, which the Schwartz criterion tends to choose. Criteria such as the

general-to-simple procedure or the AIC do a better job in this case because they allow for more parameters, and so do a better job at approximating the moving average process.

Hall (1994) has also shown that the Schwartz criterion can reject a unit root too often when the true model is an autoregressive model with skipped lags — in his examples, when only the fourth or eighth lags matter. The problem is especially severe when the autoregressive coefficients are negative. As with MA processes, approaches such as general-to-simple or the AIC do better because they include more parameters, and so are more likely to include the coefficient following the skipped lags.

In our empirical work, we've chosen to focus on the Schwartz criterion and the general-to-simple procedure. One reason is that Hall (1992) has shown that in large samples, both of these procedures yield the same distribution for the test statistics as choosing the correct lag length *a priori*. Also, as we've just discussed, the Schwartz criterion will do the best job of choosing a parsimonious model when such a model is correct, while the general-to-simple approach guards against the possibility that more-complex possibilities, such as a negative MA process or skipped lag models, will be overlooked.

### 2.3 — *Statistical properties of the ADF and other unit root tests*

Many studies have considered the finite sample properties of the ADF and other unit root tests. An influential early study is Schwert (1989); the most recent broad study is Elliott, Rothenberg, and Stock (1992), which is discussed extensively in Stock (1994). These studies investigate the statistical properties of the unit root tests under a variety of specifications for the dynamics of the stationary innovation process — for example, autoregressive innovation series with positive or negative serial correlation, moving average innovations with positive or negative serial correlation, and heteroskedastic innovations.

The results of these studies are broadly consistent. Based on Monte Carlo simulations, the ADF *t*-test described above has excellent “nominal size” properties — that is, its percentage of false rejections of the unit root null hypothesis based on critical values for a specific rejection rate (significance level) in finite samples is close to the specified rejection rate. A shortcoming of the ADF test is that it has poor power properties. That is, the ADF test is less likely than other tests to reject the null hypothesis of a unit root when it is indeed false (see Elliott, Rothenberg, and

Stock, 1992, and Stock, 1994).<sup>5</sup> But tests with better power properties, such as the Phillips-Perron (1988) test, don't have the good "size" properties of the ADF test.

In the unit root tests discussed above, the existence of a unit root is the null hypothesis. There is a recent literature concerning tests with the opposite null hypothesis — that the series is stationary (see Stock, 1994, and Hatanaka, 1996, for reviews). These tests generally seem to have poor size properties (that is, they reject stationarity too often), especially if the series has some degree of persistence, which, unfortunately, is just the sort of series one would be interested in testing.

Recently, Elliot, Rothenberg, and Stock (1996) have suggested a new test that has the excellent "size" properties of the ADF test but greater power. Referred to as the "DF-GLS" test, it requires quasi-differencing the data using a factor based on the sample size. Because of its possible advantages over the ADF test, we report results for this test below, although the results are often little different. The reader is referred to the original article for details on how this test is done.

#### *2.4 — Previous studies of unit roots in unemployment*

Studies of the stationarity of unemployment rates in European countries come to unambiguous conclusions: Using a wide variety of techniques and sample frequencies and periods, they almost uniformly fail to reject unit roots in the unemployment rate, suggesting that unemployment rates in Europe are nonstationary (for example, see Mitchell, 1993, and Røed, 1996).

The evidence for the United States is more mixed. Studies that suggest that U.S. unemployment is nonstationary include: Mitchell (1993), which fails to reject a unit root in yearly or quarterly data (1960:2 to 1991:3) for the U.S. using the ADF and other tests; Breitung (1994), which fails to reject a unit root in quarterly unemployment from 1960:1 to 1988:4 using ADF tests, and finds mixed results using other unit root tests; and Hatanaka (1996), which fails to reject a unit root in quarterly data from 1949:1 to 1991:4 using ADF tests.

---

<sup>5</sup>Another problem with the ADF test is that it can often overreject a unit root in data with strong negative autocorrelation. But that is not a concern for our application with the unemployment rate.

Other studies, however, find that the U.S. unemployment rate is stationary. The unemployment rate is the only series for which Nelson and Plosser (1982) reject a unit root with the ADF tests (using annual data from 1890 to 1970). Perron (1988) uses the Nelson and Plosser data with variants of the ADF tests and other tests and finds mixed evidence, but more in favor of rejecting a unit root. Using quarterly data from 1970:1 to 1994:4, Røed (1996) rejects a unit root using ADF tests, and using tests whose null hypothesis is stationarity, he fails to reject that the unemployment rate is stationary. Using variants of the ADF test and the DF-GLS test, Xiao and Phillips (1997) soundly reject a unit root in the unemployment rate.

One factor affecting the rejection or nonrejection of the unit root hypothesis seems to be the lag specification. Although the data sets vary (annual and quarterly, and across different samples), when lag specifications are imposed (*e.g.*, four lagged differences for quarterly data) or longer lag specifications are chosen with data-determined methods, a unit root is generally not rejected in the unemployment rate. When a more parsimonious lag structure is chosen, a unit root in the unemployment rate is rejected. (The exception is Nelson and Plosser (1982), who reject the unit root hypothesis while using four lagged differences.) One possible explanation for these results is the loss of econometric power when the lag structure is over-parameterized: If unemployment was near the boundary of the acceptance/rejection region, excess lags might lead one to a spurious failure to reject a unit root.

A different approach is taken by Song and Wu (1997). Using unemployment data from 48 U.S. states, they use the Levin-Lin (1992) panel-based unit-root test and strongly reject a unit root in the unemployment rate, although they fail to reject it individually for nearly all states. Using state-level panel data substantially increased the power to reject the unit root hypothesis.

### 2.5 — *Nonlinearities in unemployment*

Because unemployment is bounded between zero and one, it cannot, strictly speaking, have a unit root. Of course, a unit root may nonetheless be a good local approximation. But some business cycle theories imply that unemployment, considered as a univariate process, may be nonlinear. And empirical studies of the business cycle have found evidence of asymmetries in unemployment; see, for example, Sichel (1993), Peel and Speight (1995), and Koop and Potter (1998). Koop and Potter find that a logistic transformation can account for the bulk of the asymmetry in the unemployment rate.

It is possible that nonlinearities can affect the results of unit root tests. Simulations by Granger and Hallman (1991) suggest that nonlinear transformations of unit root processes may retain much of the persistence of a unit root, but unit roots in the transformed series often are nonetheless rejected by ADF tests. Less is known, however, about the converse, that is, how nonlinearity affects the results of unit root tests when the data are fairly persistent but nonetheless stationary.

One widely-used test for nonlinearity is Ramsey's (1969) regression error specification test (RESET). One fits a linear model and tests for the significance of powers of the in-sample predicted values via Lagrange multiplier, likelihood ratio, or Wald tests. In simulations, the RESET test performs well compared to other complicated tests (such as White's, 1987, 1992, dynamic information matrix test and Lee, White, and Granger's, 1993, neural network test).

#### 2.6 — *Empirical results, quarterly data*

We begin our study with an exploration of the post-war quarterly data. We begin with the quarterly data for two reasons. First, our Phillips curve analysis will be done with the post-war quarterly data, and it is useful to know the properties for the corresponding period. Second, we wish to examine the properties of the demographically adjusted unemployment rate, since it corrects for a readily identified factor that may have shifted the NAIRU, and it can only be constructed in the post-war period. (For more details on the demographically adjusted unemployment rate, see section 3.1, below.)

Because of the possibility of nonlinearity in the unemployment rate, we also examined logistically transformed versions of both the overall civilian and the demographically adjusted unemployment rates. The logistic transformation is  $\log[RU/(100-RU)]$ . As we discussed earlier, Koop and Potter (1998) found that it accounts for the asymmetries in the unemployment rate.

We report both the ADF t-test and the DF-GLS test. Recall that both the ADF and DF-GLS tests have good "size" properties, but that the DF-GLS test has better power properties. In practice, the two tests give very similar results, and because it is better known and intuitively simple to understand, we focus on the ADF test.<sup>6</sup> We also report the results of a version of the

---

<sup>6</sup>We also ran, but do not report, the Phillips-Perron (1988) test; the ADF Z-test; the Elliot, Rothenberg, and Stock (1992) point-optimal test; and the Zivot-Andrews (1992) extension of the ADF test to allow for a structural break. The results in all cases were qualitatively similar to the results we report here.

RESET test (specifically, the version of the test developed by Lee, White, and Granger, 1993) to detect nonlinearities in the unemployment rate series.

The two panels of table 1 show results using the Schwartz criterion and the general-to-simple approach to choosing the number of lagged changes in the unemployment rate to include in the test regression, as discussed in section 2.2. The procedures yielded strikingly different results: The Schwartz criterion preferred one lagged unemployment change for each measure of the unemployment rate, while the general-to-simple procedure chose twelve lagged differences in each case. Using the Schwartz criterion, the results in the top panel of table 1 are straightforward: The results suggest that we can reject a unit root at the 1 percent level for three of four unemployment series, and at the 5 percent level in the remaining case. Taken by themselves, these results imply that the four measures of the unemployment rate are stationary.

The RESET test for neglected nonlinearities strongly rejects the null hypothesis of linearity for the civilian unemployment rate and demographically-adjusted unemployment series. However, nonlinearities seem to be accounted for by the logistic transformation, as linearity is not rejected for the series which are logistically transformed, consistent with Koop and Potter (1998). Nonetheless, the nonlinearity in the unemployment rate does not appear to have affected the unit root test results, since the rejections are about as strong in the raw levels as in the logistically transformed data.

The bottom panel of the table 1 shows results using the general-to-simple lag length specification technique. With this technique, we cannot reject a unit root in the unemployment rate with either the ADF test or the more powerful DF-GLS test. The failure to reject a unit root by the general-to-simple technique is likely due to a loss of power associated with the longer lags.

In choosing which of these sets of results to believe, an important consideration is whether one of the “perverse cases” noted in the literature — namely, a large negative MA(1) component or the skipped-lag autoregressive model — is important for the unemployment rate. If not, then we would want to favor the results using the Schwartz criterion, since in the absence of such factors, it will generally pick the correct model.

An examination of the autocorrelation function of the differences of the unemployment rate quickly rules out the possibility of an large negative MA(1). However, when we examine the regression being run in the general-to-simple case, in which the change in the unemployment rate

is a function of the lagged level of the unemployment rate plus twelve lagged changes, we find that there are large negative coefficients on the lagged changes at lags four, eight, and twelve. This finding corresponds closely to the skipped lag case, where Hall (1994) found that the Schwartz criterion did poorly. But large negative autocorrelation at lags four, eight, and twelve may be an indication of problems with seasonal adjustment. In particular, if a seasonal adjustment procedure “overadjusts,” there will be negative autocorrelation at the “seasonal frequency.”

One way to escape problems with seasonal adjustment is to look at annual data. Since a unit root is a function of the long-run properties of the data, the move to annual data shouldn't matter, to a first approximation. (The main advantage of higher-frequency data, in principle, is that it may allow us more econometric power to estimate the lagged changes needed to absorb serial correlation.)

### 2.7 — *Empirical results, annual data*

BLS's annual unemployment rate data begin in 1929. Allowing for up to four lagged unemployment changes, the sample can begin in 1934.

As with the quarterly data, we applied both the Schwartz criterion and the general-to-simple procedure to choose the number of unemployment-change lags to include. As the first column of table 2 indicates, both methods suggested using just one lagged change in the unemployment rate to adjust for serial correlation. Since the studies we cited earlier suggest that the Schwartz criterion may pick too few lags while the general-to-simple procedure tends to pick too many, there seems to be little ambiguity that in this case, one lag is the preferred number. As the table indicates, the hypothesis of a unit root in the unemployment rate is rejected at the 5 percent level for the ADF test and at the 1 percent level for the more-powerful DF-GLS test.

Since a long sample is particularly important in unit root analysis, these results might be considered definitive. There are a couple of reasons for returning to the post-war sample we examined in the previous section, however. First, it is of interest to see whether the hypothesis that it was indeed seasonal adjustment problems that led to the inability to definitively reject the unit root in the quarterly data. Second, the period from 1934 to 1945 was dominated by World War Two, which was a peculiar period for the U.S. economy. Hence, it may be reasonable to exclude this period.

Looking at the period from 1949 to 1997, both the Schwartz criterion and the general-to-simple approach chose no lags. As indicated in the final column of table 2, the unit root is rejected at only the 10 percent level using the ADF test. However, it is rejected at the 5 percent level using the more-powerful DF-GLS test.

### 2.8 — *The null hypothesis of stationarity*

As we noted above, tests have been developed that allow stationarity, rather than nonstationarity, to be the null hypothesis, a switch from the augmented Dickey-Fuller test and the other tests we have discussed until now. Results of such tests could help confirm the results from the standard unit root tests: Because we have been able to reject the hypothesis of nonstationarity (at least for annual data), we would expect that we would *not* be able to reject the hypothesis of stationarity.

Two recent variants of stationarity tests are by Kwiatkowski et. al. (1992) (KPSS) and by Leybourne and McCabe (1994, 1999). They differ in their treatment of autocorrelation under the stationary hypothesis. The relationship between the KPSS and Leybourne-McCabe tests is analogous to that between the Phillips-Perron and ADF tests; the KPSS test makes nonparametric adjustments for autocorrelation via an estimate of the spectral density at the zero frequency (which involves choosing a lag truncation parameter), whereas the Leybourne-McCabe test involves fitting a parametric model to account for the dynamics (which involves choosing  $p$  in an ARMA( $p,1$ ) model).

For the KPSS test, the lag truncation parameter was chosen using Andrew's (1991) automatic bandwidth selection procedure; given the bandwidth, the spectral density was estimated using a quadratic spectral kernel (see den Haan and Levin, 1994 and 1996, for an extensive discussion of these estimation issues). The 10%, 5%, and 1% critical values for the null of stationarity are 0.347, 0.463, and 0.739, respectively. The KPSS test statistic was 0.262; the test statistic was 0.219 if one uses pre-whitening (see Andrews and Monohan, 1992). In both instances, one fails to reject the stationary null.

The critical values for the Leybourne and McCabe (1994, 1999) tests are the same as those for the KPSS test. Using a variant of a general-to-simple method they show is optimal for selecting the order of the autoregressive component in an ARMA( $p,1$ ) model, the statistic for their modified test (1999) is 0.091. Again, we fail to reject a stationary unemployment rate.

In annual data for the 1934-1997 period, the KPSS test statistics without pre-whitening and with pre-whitening are 0.221 and 0.257, respectively. The modified L-M statistic is 0.083. For the 1949-1997 period, the KPSS statistics are 0.300 and 0.199; the modified L-M statistic is 0.261. Using annual data, we fail to reject stationarity. The fact that we are unable to reject the stationary null in both quarterly and annual data confirms the impression from unit root tests that the unemployment rate is indeed stationary.

### *2.9 — Conclusions on permanent shocks to the U.S. unemployment rate*

On balance, we believe our results suggest that there is not a unit root in the U.S. unemployment rate. The main results that would indicate otherwise were those from the quarterly data using the general-to-simple model selection approach. But in that case, the selection criterion chose a model with many lags, and an examination of regression results suggested that those lags were important because of faulty seasonal adjustment. Annual data are not affected by seasonal adjustment issues, and with annual data, a unit root was rejected at significance levels of between 1 and 10 percent. Consistent with these results, when we used tests that made stationarity the null hypothesis, we were unable to reject the hypothesis of stationarity.

## ***3. Examining Temporary Hysteresis with the Phillips Curve***

### *3.1 — Model specification and data*

Our point of departure for the model of inflation is the empirical Phillips curve developed by Robert Gordon. A recent version of the model is described in Gordon (1997). Gordon's models relate inflation to the unemployment rate, real import prices, the relative price of food and energy, and a variable to pick up the swings in inflation associated with the Nixon-era wage-and-price controls. (Gordon also includes a variable to capture deviations in productivity growth from its trend. However, in the results Gordon presents, this variable is only significant in one of the three specifications he presents, and so we do not pursue it further.)

We examine several different measures of inflation, including the GDP chain price index, the nonfarm business sector deflator, and the CPI, adjusted to use a measure of rental equivalence for the pre-1983 period.

Gordon uses the BLS's main measure of the civilian unemployment rate in his model. However, shifts in the composition of the population lead to readily identifiable shifts in the NAIRU — for example, teenagers and young adults have higher average unemployment than

mature adults, and so the arrival of the baby boom in the labor market in the 1960s and 1970s lead to an increase in the NAIRU. To account for these shifts, we use a demographically adjusted unemployment rate in our analysis: We combine the unemployment rates of five demographic groups using labor-force weights for a constant base year. The five groups are teenagers of both genders; young women (20 to 24 years old); young men; mature adult women (25 and over); and mature adult men. The base year is 1993.<sup>7</sup>

For the relative prices of imports and of food and energy, we use the same definitions as Gordon: For the relative price of imports, we use the percent change in the chain-type price index for total imports less the percent change in the price index used in the relevant model. For the combined effect of the relative prices of food and energy, we use the percent change in the chain-type price index for personal consumption expenditures (PCE) less the percent change in the chain-type price index for PCE excluding food and energy.<sup>8</sup> Our Nixon-era price control variable is equal to one in 1971:Q3 to 1974:Q1 and equal to minus 11/3 in the period 1974:Q2 to 1974:Q4.

Previous work (Barsky, 1987), as well as our own investigations (not reported) using the ADF and DF-GLS tests, suggest that inflation does not return to its average value over the period we examine. As a consequence, we use the change in inflation as the dependent variable in our Phillips curve, which is equivalent to having the sum of the coefficients on lagged inflation be one in equation 1.2.

---

<sup>7</sup>Our measure of demographically adjusted unemployment also adjusts for two recent changes to the methodology BLS uses to compute the unemployment rate: Between 1989 and 1990, the BLS adopted 1990 Census weighting of population groups, which we estimate had the effect of raising the unemployment rate by 0.10 percentage point. Between 1993 and 1994, the BLS adopted a new survey of households, which we estimate raised the measured unemployment rate by a further 0.08 percentage point. We add the adjustments to the historical data, putting them on the currently defined basis.

<sup>8</sup>The price index for PCE excluding food and energy is not available prior to 1959. To construct the index, we use currently available national accounts data for prices and expenditures of food, gasoline, and fuel oil. The remaining items in the food and energy category are natural gas and electricity, for which data are not currently published before 1959. However, they were in the past, and so we used historical annual expenditure share data and consumer price indexes to create the necessary pre-1959 data for these categories.

Consistent with our results in section 2, we treat the unemployment rate as stationary. We therefore include the current level of the unemployment rate. As we discuss in more detail below, we also include long lags of past unemployment rates as our estimate of the time-varying NAIRU. In some of our sensitivity analysis, we replace the unemployment rate with its logistic transformation,  $\log[RU/(100-RU)]$ .

Our estimation period is 1956:Q2 to 1997:Q4.

### 3.2 — *Choosing the base model*

We begin by choosing the specification for the variables in the model other than unemployment before turning our attention to the role of unemployment.

We use a standard specification search criterion — the Schwartz criterion — to choose the best model specification for each measure of inflation. The idea of such criteria is to balance the fit of the model against the number of parameters. Statistical studies show that for simple models, the Schwartz criterion will more often pick the correct specification than competing criteria, such as the Akaike criterion, which often chooses models with too many parameters.<sup>9</sup>

We used Gordon (1997) as our point of departure. Gordon's model includes very long lags on past inflation — of twenty-four quarters. His model also includes six lags of the change in relative import prices and food/energy prices. Based on our past experience with such models, we included the contemporaneous food/energy variable, but only lags of the import-price variable. Because we want to leave the unemployment rate aside at this stage, we included the current level of the unemployment rate and twelve unconstrained lags of unemployment.

We looked at the lags of inflation first. We assumed that the coefficients on lagged inflation changes lay on a smooth polynomial — the so-called polynomially distributed lag (PDL) assumption.<sup>10</sup> We searched over both the lag length and the polynomial order and chose the

---

<sup>9</sup>In a regression with  $T$  observations and  $K$  parameters,  $\log(SC) = \log(SSR/T) + K \log(T)/T$ , where  $SSR$  is the sum of squared residuals from the regression. The optimal number of lags minimizes the Schwartz criterion: Adding another regressor reduces  $SSR$  but raises  $K$ , and therefore the second term of the formula. See the references cited in Judge, et al., (1985, p. 688) for a review of simulation studies on the properties of the Schwartz criterion.

<sup>10</sup>In his recent work, Gordon uses four-quarter averages of lagged inflation to reduce the number of degrees of freedom introduced by the long inflation lags he includes. The PDL assumption is another way to reduce the number of degrees of freedom used by the inflation lags; Gordon used the PDL approach in earlier versions of his models.

specification that minimized the Schwartz criterion. We next examined the number of lags of the import price and food/energy variable jointly. We looked at these variables together because there is good reason to think that they may be closely related. For example, oil prices will have an important effect on both import prices and energy prices. Because the lag lengths we're considering here are short, we didn't bother with constraining them to follow a PDL.

Finally, we did a cross-check that the choice of the inflation change PDL was not sensitive to the results for the import price and food/energy lags. We only report the final choice.

The top panel of table 3 shows the chosen models for each of the three measures of inflation. The lagged inflation specification is very similar across the three measures of inflation, and is consistent with Gordon's choice of very long lags. There is some variation in the preferred specification for the other variables. Notably, the product price models (PGDP and PNFB) preferred several lags of import prices while the CPI model omitted import prices.

We find the very long lags on inflation changes chosen by the model selection criterion to be somewhat implausible, since it is not clear why such long lags should matter for current inflation. We therefore also explore a specification with just twelve lags on the change in inflation. As with the longer-lag specifications, the model selection criteria preferred to fit the coefficients to a second-order polynomial. The choice of import and food/energy price lags was somewhat sensitive to this change in specification; the lower panel of table 3 show the results in this case.

### 3.3 — *Estimating the effects of unemployment lags*

Having established our base specifications, we now turn to our main question of interest, namely, how lagged unemployment enters the model.

We begin in table 4 using the specification for inflation lags suggested by the model selection criterion, which were around six years. We will explore the sensitivity to this assumption later on. In the top panel, we show the results of searching over PDL specifications for lagged unemployment, allowing for up to twenty-four lags. The preferred specifications included one to seven quarters on lagged unemployment, depending on the measure of inflation that is used in the model. In all three cases, the order chosen for the coefficient lag polynomial was zero, indicating that the preferred specification was a simple moving average. These specifications implied negative estimates of  $\alpha$ . As we discussed in section one, this result is

symptomatic that the shorter lags in these models are not picking up temporary hysteresis, which would imply that they have positive coefficients, but rather that they reflect the lagged effects of unemployment in the Phillips curve, as in the staggered contracts model, and so have negative coefficients.

To allow for near-term lags to enter the model for reasons other than hysteresis, we repeated the preceding exercise, this time including four unconstrained lags of the unemployment rate as a separate near-term effect.<sup>11</sup> The results are shown in the bottom panel of table 4. For the CPI, the results are qualitatively similar to those in the top panel: The lag length is short, and the sum of the coefficients is negative, implying a negative value for  $\alpha$ . For the two product prices, however, the results now suggest long lags on past unemployment.<sup>12</sup> Once again, the preferred polynomial order is zero, so that the unemployment lags effectively enter the model as a long moving average. While the long moving average of past unemployment is not statistically significant at traditional confidence levels, it implies an economically important effect, with an implied weight of lagged unemployment in the temporary NAIRU of between 0.25 and 0.4.

Table 5 shows results when we use a relatively short lag on past changes in inflation, of just twelve quarters. Looking first at the top panel, the results for the CPI and the nonfarm business sector deflator are now qualitatively different from those in the top panel of table 4: The preferred lag length on past unemployment is now very long, five years for the CPI and six for the NFB deflator. The estimated weight on past unemployment in the NAIRU is also large — between 1/4 and 1/2 — and for the NFB deflator, the effect is statistically significant at the 5 percent level. For the GDP price index, the lag length is also greater than before. However, in this case, the order of the PDL is one, rather than zero, as it had been for the other results. The effect of this slope to the lag coefficients is that the near-term coefficients are negative while only

---

<sup>11</sup>Consistent with the specification in equation 1.5, the total weight on the long moving average of unemployment includes an implied weight on the first four lags.

<sup>12</sup>For the model using the GDP chain price index, the specification search actually preferred a model with no lags of unemployment beyond the four lags already included; the twenty-four lag model is the preferred model from those with nonzero lags. However, because the effect of the long lags in this case was potentially of economic significance, we decided to consider it further.

the longer lags are positive. On balance, the sum of these lagged coefficients is negative, inconsistent with an interpretation in terms of hysteresis.

The near-term negative coefficients in the case of the GDP chain price index once again suggests that near-term unemployment dynamics should be considered separately from longer lags, and the bottom panel of table 5 does this. Here, the unemployment lags for all three inflation measures have the correct sign, and enter the model with an intermediate lag length, implying that the NAIRU is affected by a lagged unemployment of from two-and-a-half to almost four years. The sum of the long unemployment lags is now significant at the 5 percent level for all three measures of inflation. The implied weight on lagged unemployment in the NAIRU is higher than before, at between 40 and 60 percent.

We explored the sensitivity of our results along a number of dimensions. First, in all cases where the effect of lagged unemployment on the NAIRU was estimated to be positive, the effect entered as a zero-degree polynomial — that is, as a simple moving average of past unemployment rates. But it may be reasonable to suppose that recent unemployment may have a larger effect than more-distant unemployment rates. To explore this possibility, we allowed for a first-order polynomial — that is, we allowed the coefficients to lie on a line with a non-zero slope. This exercise did not have an important effect on the the results: For the models without near-term unemployment lags, we found that allowing the coefficients to vary implied that the near-term coefficients were negative, suggesting that these coefficients are capturing near-term lagged unemployment effects not related to hysteresis. For the models that include separate near-term unemployment lags, we found that the slope was very small, and had little effect on the results. As consequence, we do not report them.

We also examined the sensitivity of the results to the use of a logistic transformation of the unemployment rate rather than the raw level of the unemployment rate, as suggested in section 2. We found that the results didn't change much: The  $R^2$  of the models changed very little (the raw level model fit marginally better), and directly comparable coefficients were very similar: For example, in one case, the weight placed on the long moving average rose by 0.01, while the long-run NAIRU fell from 5.8 percent to 5.7 percent. In general, the NAIRU was between 0.1 and 0.2 percentage point lower in each period when calculated from the model that includes the logistic transformation of the unemployment rate. Except for this slight difference in the estimate

of the NAIRU, there appears to be little practical difference in the results using the logistic transformation.

To verify that we were correct to impose that the sum of the coefficients on lagged inflation was indeed one, we added the lagged level of inflation to each of our equations. If the lagged level is statistically significant with a negative coefficient, that constitutes evidence that the sum of the lagged coefficients is less than one. The largest t-ratio we found was 1.2, in the case of the two short-lag models using the nonfarm business deflator. That t-ratio is far from statistically significant: Because this test is closely related to tests of the stationarity of inflation, the cut-off for the 95 percent confidence level is likely closer to three than to two. In this case, the implicit sum of the lagged coefficients was 0.94, which is economically close to one. In all other cases, the sum of the coefficients was larger and the t-ratio was smaller. In fact, in five of the six specifications with long inflation lags, the implicit sum of the coefficients was greater than one, although the largest t-ratio was only 1.2.

To summarize, the results on temporary hysteresis are not very robust. In five of twelve cases in tables 4 and 5, the implied weight on lagged unemployment in the NAIRU was negative, implying that hysteresis was an inappropriate model. In three more cases, the effect of long unemployment lags was not statistically significant at the 5 percent level, although the size of the effect was economically significant. Also, the results were sensitive to the specification along several dimensions: They are affected by the lag on past inflation, with the evidence in favor of temporary hysteresis stronger when the inflation lag is limited to twelve quarters. The results are also sensitive to the measure of inflation used, with the results in favor of temporary hysteresis strongest when the nonfarm business deflator is the measure of inflation, and weakest for the CPI. Finally, including near-term unemployment dynamics appears to sharpen the evidence in favor of temporary hysteresis.

### 3.4 — *The implicit temporary NAIRU*

With estimates of the lag polynomial,  $\alpha(L)$ , and the long-run NAIRU in hand, we can calculate an implicit temporary NAIRU, using equation 1.1. Given that in all cases where the lag polynomial entered with the correct sign, it was a simple moving average of past unemployment rates, the formula for the NAIRU simplifies to:

$$NR_t = \alpha \sum_{i=1}^n \frac{RU_{t-i}}{n} + (1-\alpha)NR^* \quad . \quad (3.1)$$

Figures 1 and 2 show some of the results. In figure 1, we plot the short-term NAIRU for models where the moving-average lag length is twenty-four, as in four of the twelve cases shown in tables 4 and 5. We show estimates when  $\alpha$  is at the two extremes (0.24 and 0.49) and for an intermediate value of  $\alpha$  (0.33), which corresponds to the estimate for the nonfarm business deflator in the bottom panel of table 4. In addition, figures 1 and 2 show the demographically adjusted unemployment rate. We begin with this specification because it isolates the effects of temporary hysteresis on the NAIRU. We will show the implications for the standard civilian unemployment rate, which include the effects of the demographic adjustment, shortly.

Because all three measures of the NAIRU are weighted averages of a six-year moving average of past unemployment and a constant, they move together. The NAIRU estimate that varies the most is the one that puts the biggest weight — 49 percent — on past unemployment. This estimate of the NAIRU varies between a low of 5 percent in the early 1970s — following the low unemployment of the 1960s — to a high of 6-3/4 percent in the mid-1980s — following the high unemployment of the early 1980s. On the other hand, the model with the smallest variation suggests that the NAIRU stayed between 5.4 and 6.3 percent.

For the third quarter of 1998, the last data point plotted, all three models suggest that the NAIRU was not far from the long-run NAIRU estimates of these models. After making the conversion from the demographically adjusted to the published civilian unemployment rate, which trims 0.03 percentage point from the NAIRU estimate in the third quarter of 1998, all three models imply NAIRUs of about 5-3/4 percent.

For a given experience of low unemployment, the models imply estimates of the benefits in terms of a temporarily lower NAIRU. Using the lag length on unemployment — twenty-four quarters — and the weight on lagged unemployment — which, for the nonfarm business deflator model in the top panel of table 5, is 0.49 — the calculation is straightforward: If the unemployment rate spends one year at 4.5 percent rather than at the model's NAIRU of 5.7 percent, the six-year moving average of past unemployment will drop by 0.2 percentage point. As a consequence, the NAIRU will drop by about half that much, or 0.1 percentage point. This benefit to the NAIRU will last for five years but then taper off over the course of the sixth year as the period of low unemployment drops out of the moving average. Of course, in the other two models, the effect would be smaller than this.

To give some notion of the maximum possible benefit of the recent low unemployment, the estimates from the nonfarm business sector model suggest that if the unemployment rate were to stay at 4-1/2 percent for the next six years, the NAIRU would drop to 5-1/4 percent.

The implications of such unemployment paths for inflation would depend on the source of the lower unemployment. If unemployment is low because of high aggregate demand, then during the period when the unemployment rate is below the NAIRU, inflation would rise, and at least a portion of the increase in inflation would be permanent. On the other hand, if unemployment were low because positive shocks to inflation have allowed unemployment to drop below the NAIRU without pushing up inflation, then the benefits of the temporarily lower NAIRU could come without higher inflation.

Figure 2 shows NAIRU estimates based on the set of model results shown in the bottom panel of table 5. These models had the shorter inflation lags and near-term unemployment dynamics. The results for these models had in common an “intermediate” lag length on the moving-average for past unemployment, of between ten and fifteen quarters, and large weights on the moving average in forming the temporary NAIRU.

The estimate of the NAIRU based on the nonfarm business deflator model has the biggest range of variation. The variation is greater than any of the estimates in figure 1, both because the weight on lagged unemployment is higher and because the lag length in the moving average was shorter. The estimate of the NAIRU from this model hit a low of 4-1/2 percent in 1970 and a high of 7-1/2 percent in 1984. The model using the GDP price index has almost as much variation, with similar timing. The estimate from the CPI model varies less, with a range of 5 to 7 percent, both because it uses a longer lag of past unemployment and because it puts a lower weight on the unemployment lag.

In contrast to the models in figure 2, these estimates suggest that the NAIRU has recently dropped below its long-run average. Again, that’s both because they use a shorter lag of past unemployment in forming the NAIRU, and so place a larger weight on the recent unemployment experience, and because the weight on lagged unemployment is larger in these models. After adjusting for the difference between the demographically adjusted and the published unemployment rates, the two product-price models both suggest similar NAIRU estimates — of

5.3 percent for the model with the NFB deflator and 5.4 percent for the model with the GDP price index. The CPI model puts the NAIRU estimate a bit higher, at 5-3/4 percent.

To get a notion of the benefits of recent low unemployment, we can repeat our earlier calculation using the nonfarm business sector model plotted in figure 2. Spending one year at 4.5 percent unemployment rather than at the model's NAIRU of 5.7 percent will lower the eleven-quarter moving average of past unemployment that enters the model by 0.44 percentage point. With a coefficient of 0.58 on the moving average, that implies a reduction in the NAIRU of 0.25 percentage point. The full benefits to the NAIRU will last for the following two years, and will gradually fade over the following year, as the period of low unemployment drops out of the eleven-quarter moving average of lagged unemployment.

Figures 1 and 2 plot the demographically adjusted unemployment rate and the "demographically adjusted NAIRU." In figures 3 and 4, we add the difference between the civilian unemployment rate and the demographically adjusted rate to the temporary NAIRU estimates on a demographically adjusted basis to come up with a NAIRU that can be compared to the civilian unemployment rate. Figure 3 shows the two most extreme NAIRU estimates from figure 1. In addition, to give some notion of the effects of the demographic adjustment, we also add the demographic adjustment to a constant NAIRU of 5.8 percent. In figure 4, we convert each of the NAIRU series in figure 2 to a civilian unemployment basis. As can be seen, adding back the demographic adjustment doesn't change the overall impression that the NAIRU fell in the late 1960s; rose between the early 1970s and the early 1980s; and fell between the mid-1980s and the early 1990s. The NAIRUs based on the shorter lags on unemployment in figures 2 and 4 vary somewhat more, but nonetheless, the overall impression is not affected by the demographic adjustment.

The figures suggest one potentially important policy implication of temporary hysteresis: Following a deep recession, policymakers may need to exercise caution in bringing unemployment back down, since the NAIRU may be higher than before the recession. A good example is the mid-1980s, when the recessions earlier in the decade had pushed the NAIRU above its long-run level.

#### **4. Conclusions**

Our results suggest that hysteresis is not a first-order concern for the evolution of

U.S. unemployment: First, we find that unemployment tends to come back to its long-run average value, ruling out the possibility of permanent hysteresis. Second, we find that evidence for temporary hysteresis is weak, with the effect most often either insignificant or entering with the wrong sign; even when it is statistically significant, it is only marginally so. Third, when the temporary hysteresis effect has the correct sign, the implied swings in the NAIRU are moderate in size, with the NAIRU varying between 5 and 7 percent for most models.

While the model predicts that the recent experience of low unemployment will leave the NAIRU lower than it would otherwise have been, the effects are not likely to be very large. For example, the results suggest that spending one year with the unemployment rate at 4-1/2 percent rather than a long-run NAIRU of 5-3/4 percent will reduce the NAIRU for the next several years by at most 1/4 percentage point.

***Appendix: Do the Long-Term Unemployed Exert Less Pressure on Inflation?***

One possible implication of hysteresis in the unemployment rate is an increase in the duration of unemployment spells. For example, in the human capital story, those who lose skills and find it more difficult to find a job will likely remain unemployed for longer than they otherwise would have.

Longer unemployment spells play a more prominent role in the “insider-outsider” model. In that model, an important source of hysteresis is the reduced pressure that the long-term unemployed — the “outsiders” — place on inflation. Hence, in the insider-outsider model, the long-term unemployed should have a smaller effect on inflation than the short-term unemployed. On the other hand, while unemployment spells would likely lengthen following high unemployment under the human capital hypothesis, it is not clear that the coefficient on the long-term unemployment rate would be different than the coefficient on short-term unemployment. As a consequence, the role of long-term unemployment in the Phillips curve may be useful in distinguishing among different sources of hysteresis.

However, a smaller coefficient on the long-term unemployment rate than on the short-term unemployment rate may not be definitive evidence in favor of the insider-outsider hypothesis. As pointed out in Røed (1997), the argument relies on an implicit assumption of homogeneous workers, so that the only difference between the long-term unemployed and the short-term unemployed is bad luck. But if the long-term unemployed have more trouble finding a job because they are inherently less employable, there would be a lower coefficient on the long-term unemployment rate even if there were no hysteresis. So even if there were a notably smaller effect of the long-term unemployment rate on inflation, it might not tell us anything about hysteresis or insider-outsider effects.

Figure A.1 shows the overall civilian unemployment rate and two subcomponents: The ratio of those unemployed 26 weeks or less to the civilian labor force, and the ratio of those unemployed more than 26 weeks to the civilian labor force.<sup>13</sup> Twenty-six weeks is a natural

---

<sup>13</sup>The unemployment rates plotted in figure A.1 have been adjusted for the revision to the household survey in 1994, using estimates of BLS analysts (Polivka and Miller, 1994) to show the unemployment rates on a current basis. The adjustment for the overall unemployment rate is trivial (0.08 percentage point higher after 1994). However, the adjustment for the long-term unemployment rate is larger — we’ve assumed that it is 1/3 percentage point higher after 1994,

breakpoint between long- and short-term unemployment, since standard unemployment insurance benefits in most states have that duration. Several facts are immediately apparent: First, the vast majority of the unemployed are unemployed for 26 weeks or less: Over the 1956-to-1997 period, the long-term unemployed were only 18 percent of overall unemployment, and even when long-term unemployment runs up in recessions, they are only about one-fourth of the unemployed. Second, the long-term unemployment rate is highly correlated with the short-term unemployment rate: It rises strongly in recessions and returns to a normal value not long afterwards. Finally, it seems unlikely that long-term unemployment could account for low frequency movements in the unemployment rate: For example, the cyclical low in the overall unemployment rate recorded in 1978 was 2.4 percentage point higher than the cyclical low in 1969, but the difference in long-term unemployment rates between those years is only 0.4 percentage point, so about 80 percent of the increase in unemployment was in the short-term component.

Given these observations, it seems unlikely that the long-term unemployment rate will add much to a U.S. Phillips curve: The long-term rate is highly collinear with the short-term rate, and it does not account for much of the longer-term movements in the unemployment rate. To verify this hunch, we tried adding the long-term unemployment rate to our six baseline models, in addition to the demographically adjusted unemployment rate. If the insider-outsider hypothesis is correct, then the long-term unemployment rate should enter with a significantly positive coefficient, indicating that the long-term unemployed have a significantly smaller effect than the short-term unemployed.

We included the current level and four lags of each unemployment rate. Because the insider-outsider hypothesis is a competing hypothesis about hysteresis, we did not look at long lags as well.

Table A.1 shows results. In all but one of the six cases shown, the coefficient on long-term unemployment is economically large, but negative, while the coefficient on the overall civilian unemployment rate is reduced relative to the specification without long-term unemployment. These results are the opposite of the prediction of the insider-outsider model. In the remaining case — the model for the nonfarm business deflator with “short” inflation-change

---

based on figures for unemployment fifteen weeks or greater in Polivka and Miller (1995, table 3); the short-term unemployment rate is therefore about 1/4 percentage point lower after 1994 .

lags — the result is in accord with the predictions of the insider-outsider model, with the coefficient on long-term unemployment positive and offsetting the coefficient on overall unemployment. However, in all cases, the coefficients on long-term unemployment are not statistically significant, suggesting that we should not make much of the large negative, or positive, coefficients. Furthermore, when the long-term unemployment rate is included, the coefficient on civilian unemployment also loses its statistical significance. These results suggest that because of the high collinearity between long- and short-term unemployment, it is not possible to estimate separately their individual effects.

We take these results to be negative for the insider-outsider story. However, other models of hysteresis, notably the human-capital story, don't have any particular predictions for the coefficient on the long-term unemployment rate. Hence, they leave open the possibility that the more general models of hysteresis that we emphasize in the main text of the paper may still be important.

## **References**

- Andrews, D.W.K. (1991): "Heteroskedasticity and Autocorrelation Covariance Matrix Estimation," *Econometrica* 59: 817-858.
- Andrews, D.W.K., and J.C. Monahan (1992): "An Improved and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica* 60: 953-966.
- Banerjee, A., R. Lumsdaine, and J.H. Stock (1992) "Recursive and Sequential Tests of the Unit Root and Trend Break Hypotheses: Theory and International Evidence", *Journal of Business and Economic Statistics*, 10, 271-288.
- Barsky, R.B. (1987) "The Fisher Effect and the Forecastability and Persistence of Inflation," *Journal of Monetary Economics*, 19, 3-24.
- Blanchard, O.J., and L. Summers (1986) "Hysteresis and the European Unemployment Problem," *NBER Macroeconomics Annual 1986*, MIT Press, 15-90.
- Breitung, J. (1994), "Some Simple Tests of the Moving Average Hypothesis", *Journal of Time Series Analysis*, 15, 31-359.
- Campbell, J.Y., and P. Perron (1991) "Pitfalls and Opportunities: What Macroeconomists Should Know about Unit Roots", *NBER Macroeconomics Annual 1991*, MIT Press, 141-200.
- Cheung, Y.-W., and M. Chinn (1997) "Further Investigation of the Uncertain Unit Root in GNP", *Journal of Business and Economic Statistics*, 15, 68-73
- Christiano, L. (1992) "Searching for a Break in GNP", *Journal of Business and Economic Statistics*, 10, 237-250.
- den Haan, W., and A. Levin (1994): "Inferences from Parametric and Covariance Matrix Estimation Procedures," NBER Technical Paper: 195.
- den Haan, W., and A. Levin (1996): "A Practioner's Guide to Robust Matrix Estimation," *Handbook of Statistics 15* Chapter 12, 291-341) and UCSD Discussion Paper 96-17.
- Dickey, D.A. and W.A. Fuller (1979) "Distribution of the Estimators for Autoregressive Time Series with a Unit Root" *Journal of the American Statistical Association*, 74, 427-431.
- Dickey, D.A. and W.A. Fuller (1981) "Likelihood Ratio Tests for Autoregressive Time Series with a Unit Root" *Econometrica*, 49, 1057-1072.
- Elliott G., T.J. Rothenberg, and J. H. Stock (1992) "Efficient Tests for an Autoregressive Unit Root", *NBER Technical Working Paper* 130.

Elliott G., T.J. Rothenberg, and J.H. Stock (1996) "Efficient Tests for an Autoregressive Unit Root", *Econometrica*, 64, 813-836.

Fuller, W.A. (1976) *Introduction to Statistical Time Series*, New York: Wiley.

Gordon, R.J. (1997) "The Time-Varying NAIRU and its Implications for Economic Policy," *Journal of Economic Perspectives*, 11(1), 11-32.

Granger, C.W.J., and J. Hallman (1991) "Nonlinear Transformations of Integrated Time Series" *Journal of Time Series Analysis*, 12, 207-223.

Hall, A. (1992) "Unit Root Tests after Data Based Model Selection", *Proceedings of the 1992 Summer Meetings of the American Statistical Association: Business and Economics Section*.

Hall, A. (1994) "Testing for a Unit Root in Time Series with Pretest Data-Based Model Selection," *Journal of Business and Economic Statistics*, 12, 461-70.

Hamilton, J. (1994) *Time Series Analysis*, New Jersey: Princeton University Press.

Hatanaka, M. (1996) *Time-Series-Based Econometrics*, Oxford: Oxford University Press.

Judge, G.G., W.E. Griffiths, R.C. Hill, H. Lutkepohl, and T.-C. Lee (1995) *The Theory and Practice of Econometrics*, New York: John Wiley and Sons.

Koop, G. and S.M. Potter (1998) "Dynamic Asymmetries in U.S. Unemployment" Manuscript, University of California, Los Angeles.

Kwiatkowski, D., P.C.B. Phillips, P. Schmidt, and Y. Shin (1992) "Testing the Null of Stationarity: How Sure Are We That Economic Time Series Have a Unit Root?" *Journal of Econometrics*, 54, 159-178.

Lee, T.H., H. White, and C.W.J. Granger (1993) "Testing for Neglected Nonlinearity in Time Series Models", *Journal of Econometrics*, 56, 269-290.

Levin, A. and C.F. Lin (1992) "Unit Root Tests in Panel Data: Asymptotic and Finite Sample Properties" manuscript, University of California, San Diego.

Leybourne, S.J., and B.P.M. McCabe (1994): "A Consistent Test for a Unit Root," *Journal of Business and Economic Statistics*, 12, 157-166.

Leybourne, S.J., and B.P.M. McCabe (1999): "Modified Tests with Data-Dependent Model Selection Rules," *Journal of Business and Economic Statistics*, 264-270.

Lindbeck, A., and D.J. Snower (1988) *The Insider-Outsider Theory of Employment and Unemployment*, MIT Press.

Mitchell, W.F. (1993) "Testing for Unit Roots and Persistence in OECD Unemployment Rates" *Applied Econometrics*, 1489-1501.

Nelson, C.R. and C.I. Plosser (1982) "Trends and Random Walks in Macro-economic Time Series: Some Evidence and Implications", *Journal of Monetary Economics*, 10, 139-162.

Ng, S., and P. Perron (1995) "Unit Root Tests in ARMA Models with Data-Dependent Methods for the Selection of the Truncation Lag," *Journal of the American Statistical Association*, 90, 268-81.

Peel, D.A. and A.E.H. Speight (1995) "Non-linear Dependence in Unemployment, Output, and Inflation" In R. Cross (ed.): *The Natural Rate of Unemployment*, Cambridge: Cambridge University Press.

Perron, P. (1988) "Trends and Random Walks in Marcoeconomic Time Series: Further Evidence from a New Approach", *Journal of Economic Dynamics and Control*, 12, 297-332.

Perron, P. (1989) "The Great Crash, the Oil Price Shock and the Unit Root Hypothesis", *Econometrica*, 57, 1361-1401.

Phillips, P.C.B. (1995) "Unit Root Tests", Cowles Foundation Working Paper 1104.

Phillips, P.C.B. and P. Perron (1988) "Testing for Unit Roots in a Time Series Regression", *Biometrika*, 75, 335-346.

Polivka, A.E., and S.M. Miller (1995) "The CPS after the Redesign: Refocussing the Economic Lens," photocopy, Bureau of Labor Statistics (March).

Ramsey, J.B. (1969) "Tests for Specification Errors in Classical Least Squares Regression Analysis", *Journal of the Royal Statistical Society B*, 31, 350-371

Røed, K. (1996) "Unemployment Hysteresis — Macro Evidence from 16 OECD Countries", *Empirical Economics*, 21, 589-600.

Røed, K. (1997) "Hysteresis in Unemployment", *Journal of Economic Surveys*, 11, 389-418.

Rudebusch, G. (1992) "Trends and Random Walks in Marcoeconomic Time Series: A Reexamination", *International Economic Review*, 33, 661-680.

Schwert, G. W. (1989) "Tests for Unit Roots: A Monte Carlo Investigation", *Journal of Business and Economic Statistics*, 7, 147-159.

Sichel, D. (1993) "Business Cycle Asymmetries: A Deeper Look" *Economic Inquiry*, 31, 224-236.

Song, F. and Y. Wu (1997) "Hysteresis in Unemployment: Evidence from 48 U.S. States", *Economic Inquiry*, 35, 235-243.

Stock, J.H. (1994) "Unit Roots, Structural Breaks, and Trends". in R.F. Engle and D.L. McFadden (eds.) *Handbook of Econometrics, Volume IV*, Amsterdam: Elsevier Science.

Taylor, J.B. (1979) "Staggered Wage Setting in a Macro Model" *American Economic Review* 69, 108-113.

Taylor, J.B. (1980) "Aggregate Dynamics and Staggered Contracts" *Journal of Political Economy* 88, 1-23.

White, H. (1987), "Specification Testing in Dynamic Models", in T.F. Bewley (ed.) *Advances in Econometrics, Fifth World Congress — 1*, New York: Cambridge University Press.

White, H. (1992), "Estimation, Inference, and Specification Analysis", New York, Cambridge University Press.

Xiao, Z. and P.C.B. Phillips (1997) "An ADF Coefficient Test for a Unit Root in ARMA Models of Unknown Order with Empirical Applications to the U.S. Economy", manuscript, Cowles Foundation.

Zivot, E. and D.W.K. Andrews (1992) "Further Evidence on the Great Crash, the Oil Price Shock, and the Unit Root Hypothesis", *Journal of Business and Economic Statistics*, 10, 251-270.

**Table 1**  
**Unit Root Tests Using Quarterly Data**  
Sample: 1949:Q1 to 1997:Q4

***Lag lengths chosen with the Schwartz criterion***

	<i>unem</i>	<i>udem</i>	<i>unem_L</i>	<i>udem_L</i>
Number of lags	1	1	1	1
ADF <i>t</i>	-4.16 ***	-3.44 **	-4.14 ***	-4.38 ***
<i>coefficient</i>	0.942	0.947	0.940	0.935
DF-GLS <i>t</i>	-2.21 **	-2.34 **	-2.65 **	-2.31 **
<i>coefficient</i>	0.978	0.975	0.974	0.975
RESET	8.99 **	10.07 ***	1.63	3.86

***Lag lengths chosen by general-to-simple likelihood ratio tests***

	<i>unem</i>	<i>udem</i>	<i>unem_L</i>	<i>udem_L</i>
Number of lags	12	12	12	12
ADF <i>t</i>	-2.02	-2.09	-2.01	-2.20
<i>coefficient</i>	0.967	0.964	0.967	0.962
DF-GLS <i>t</i>	-1.04	-1.14	-0.90	-1.02
<i>coefficient</i>	0.989	0.987	0.991	0.989

Entries in the table are the test statistics; *coefficient* refers to the coefficient on the lagged level of the unemployment rate.

\*, \*\*, \*\*\* Reject the null hypothesis at the 10%, 5%, and 1% levels respectively

- unem* Civilian unemployment rate
- udem* Demographically-adjusted civilian unemployment rate
- unem\_L* Logistically-transformed civilian unemployment rate
- udem\_L* Logistically-transformed, demographically-adjusted civilian unemployment rate
- ADF *t* Augmented Dickey-Fuller *t*-test for a unit root
- DF GLS *t* Elliott, Rothenberg, and Stock GLS extension of the ADF *t*-test
- RESET Regression error specification test for neglected nonlinearity

**Table 2**  
**Results of Unit Root Tests with Annual Data**  
 Civilian unemployment rate

<b>Sample:</b>	<b>1934-1997</b>	<b>1949-1997</b>
<i>Number of lags:</i>	1	0
<i>Chosen by:</i>	Schwartz, general-to-simple	Schwartz, general-to-simple
ADF <i>t</i>	-3.06 **	-2.76 *
<i>Coefficient</i>	0.873	0.734
DF-GLS <i>t</i>	-2.90 ***	-2.50 **
<i>Coefficient</i>	0.892	0.761

*Coefficient* refers to the coefficient on the lagged level of unemployment.

\*, \*\*, \*\*\* Reject the null hypothesis at the 10%, 5%, and 1% levels respectively.

ADF *t* Augmented Dickey-Fuller *t*-test for a unit root.

DF GLS *t* Elliott, Rothenberg, and Stock GLS extension of the ADF *t*-test.

**Table 3**  
**Specification Search for Various Price Measures**  
 Sample: 1956:Q2 to 1997:Q4

	GDP chain price index	Nonfarm business deflator	"Experimental" CPI
<b><i>Unconstrained inflation-change lags</i></b>			
$\Delta$ inflation PDL	Lag length: 23 Polynomial: 2	Lag length: 23 Polynomial: 2	Lag length: 22 Polynomial: 2
Food/energy lags	0	0	0 to 1
Import price lags	1 to 3	1 to 4	None
<b><i>Inflation-change lags limited to twelve</i></b>			
Food/energy lags	0	0	0 to 1
Import price lags	1 to 2	1	None

**Table 4**  
**Results of Unemployment Lags Specification Search**  
*Assuming inflation lags from specification search*

	GDP chain price index	Nonfarm business deflator	“Experimental” CPI
<b><i>Without Near-term Unemployment Dynamics</i></b>			
Preferred unemployment lags <sup>a</sup>	5	1	7
T-statistic on sum of lagged unemployment coefficients	3.3	2.4	1.4
Implied weight on lagged unemployment	< 0	< 0	< 0
Long-run natural rate	5.8	5.8	6.0
<b><i>With Near-term Unemployment Dynamics</i></b>			
Preferred unemployment lags <sup>a</sup>	24 <sup>b</sup>	24	7 <sup>b</sup>
T-statistic on sum of lagged unemployment coefficients	1.4	1.6	.4
Implied weight on lagged unemployment	.24	.33	< 0
Long-run natural rate	5.9	5.8	6.0

Notes:

- a. Unemployment PDL order is zero in all cases. That is, unemployment enters as a moving average. However, unemployment specification search allowed for higher PDL order.
- b. Specification search preferred no additional lags. Result shown is preferred lag length ruling out zero.

**Table 5**  
**Results of Unemployment Lags Specification Search**  
*Assuming twelve inflation lags*

	GDP chain price index	Nonfarm business deflator	“Experimental” CPI
<b><i>Without Near-term Unemployment Dynamics</i></b>			
Preferred unemployment lags <sup>a</sup>	10 <sup>b,c</sup>	24 <sup>b</sup>	20 <sup>b</sup>
T-statistic on sum of lagged unemployment coefficients	1.6	2.1	1.6
Implied weight on lagged unemployment	< 0	.49	.27
Long-run natural rate	5.8	5.7	6.0
<b><i>With Near-term Unemployment Dynamics</i></b>			
Preferred unemployment lags <sup>a</sup>	10 <sup>b</sup>	11	15 <sup>b</sup>
T-statistic on sum of lagged unemployment coefficients	2.2	2.1	2.0
Implied weight on lagged unemployment	.47	.58	.38
Long-run natural rate	5.8	5.7	6.0

Notes:

- a. Except as noted, unemployment PDL order is zero. That is, unemployment enters as a moving average. However, unemployment specification search allowed for higher PDL order.
- b. Specification search preferred no additional lags. Result shown is preferred lag length ruling out zero.
- c. In this case, the PDL order is one.

**Table A.1**

**Evidence on the Role of the Long-term Unemployment Rate in the Phillips Curve**

Sample: 1956:Q2 to 1997:Q4

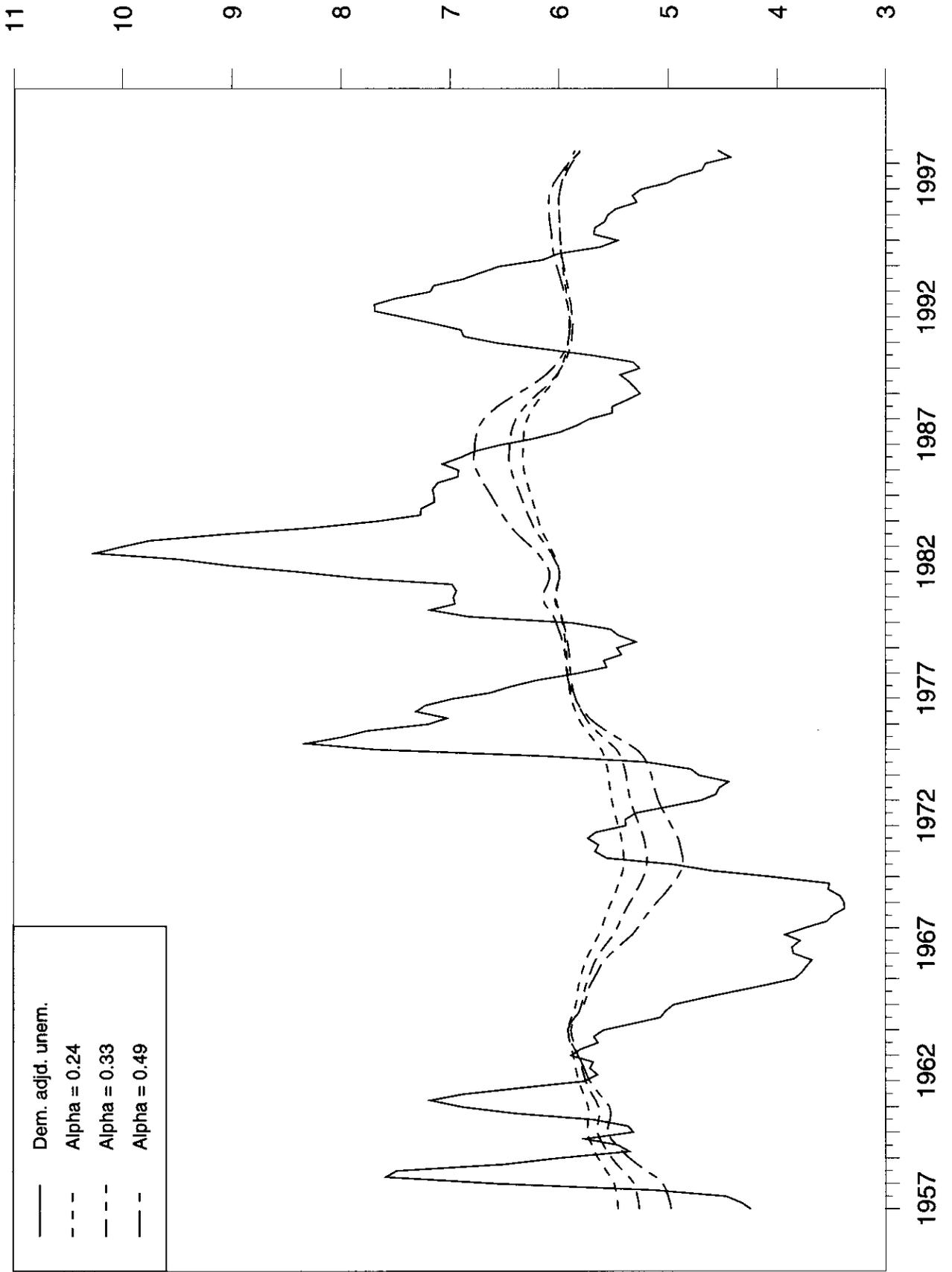
	GDP chain price index		Nonfarm business deflator		"Experimental" CPI	
<b><i>Model with long inflation change lags</i></b>						
Sum of coefficients on civilian unemployment	-.39 (.07)	-.13 (.19)	-.38 (.08)	-.28 (.23)	-.41 (.06)	-.23 (.18)
Sum of coefficients on "long-term" unemployment	--	-.95 (.72)	--	-.32 (.85)	--	-.71 (.64)
<b><i>Model with "short" (twelve-quarter) inflation change lags</i></b>						
Sum of coefficients on civilian unemployment	-.35 (.07)	-.17 (.20)	-.32 (.08)	-.38 (.24)	-.36 (.07)	-.22 (.19)
Sum of coefficients on "long-term" unemployment	--	-.62 (.76)	--	.37 (.90)	--	-.53 (.71)

Notes:

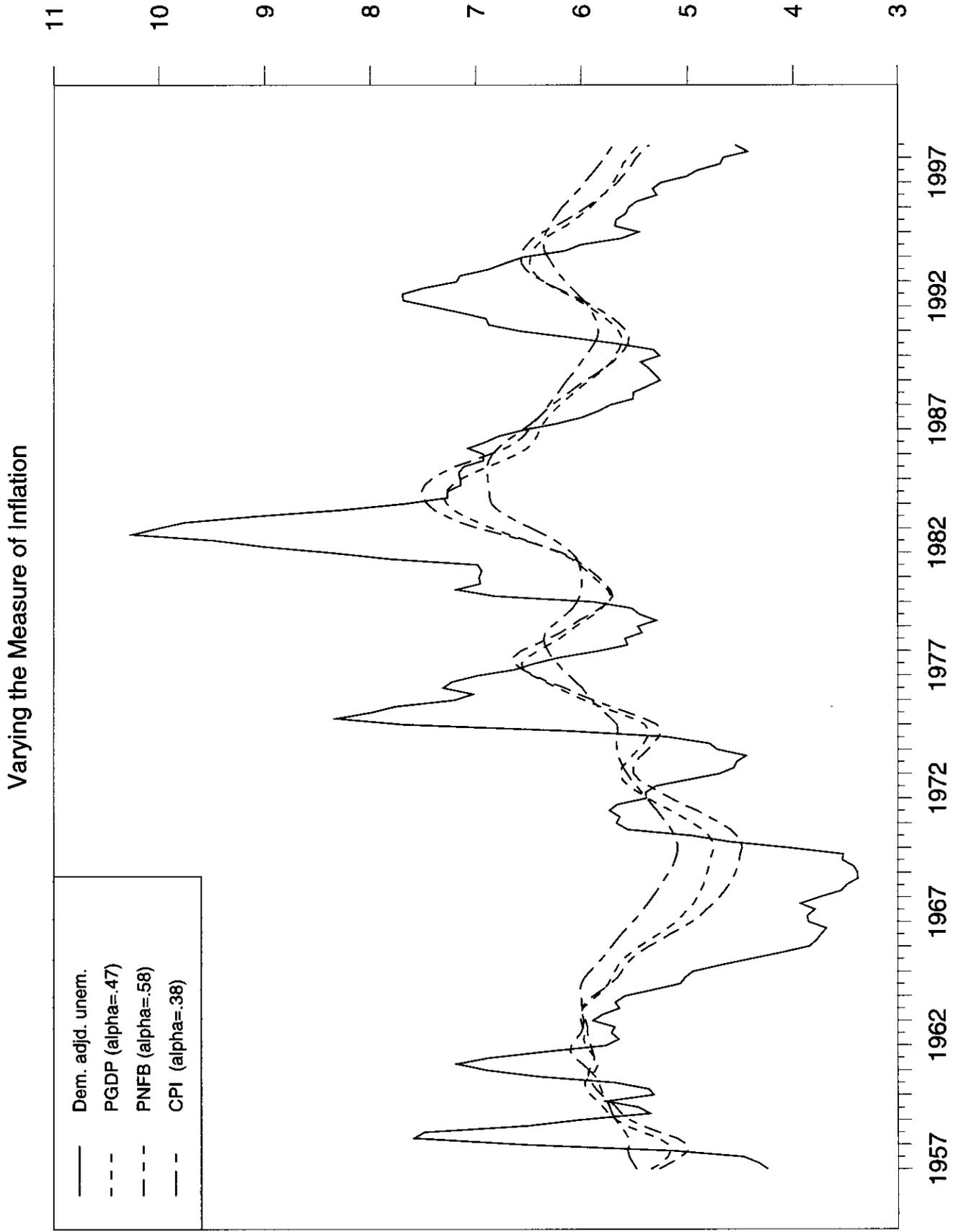
- Numbers in parentheses are standard errors.
- Each regression includes the current level and four lags of the indicated unemployment rate. Other details of the specifications are the same as those in tables 4 and 5.
- Long-term unemployment is the ratio of those unemployed more than 26 to the civilian labor force, expressed as a percent.
- Unemployment rates have been adjusted for the effects of the 1994 household survey revisions, as described in the text.

**Figure 1: NAIRUs That Reflect Temporary Hysteresis**

Using a 24-quarter moving average of lagged unemployment.

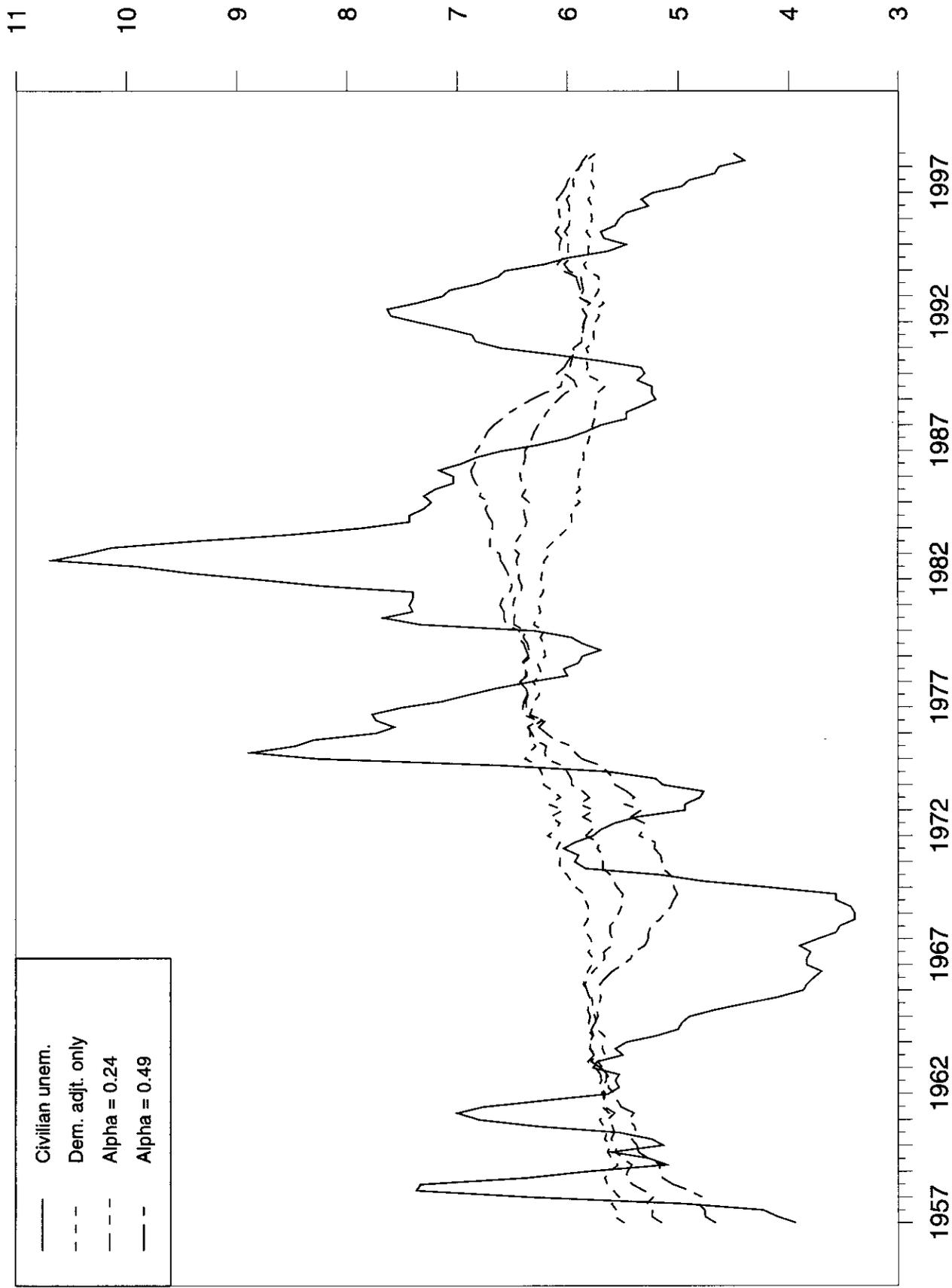


**Figure 2: Results with Short Inflation Lags and Near-term Unemployment Dynamics**



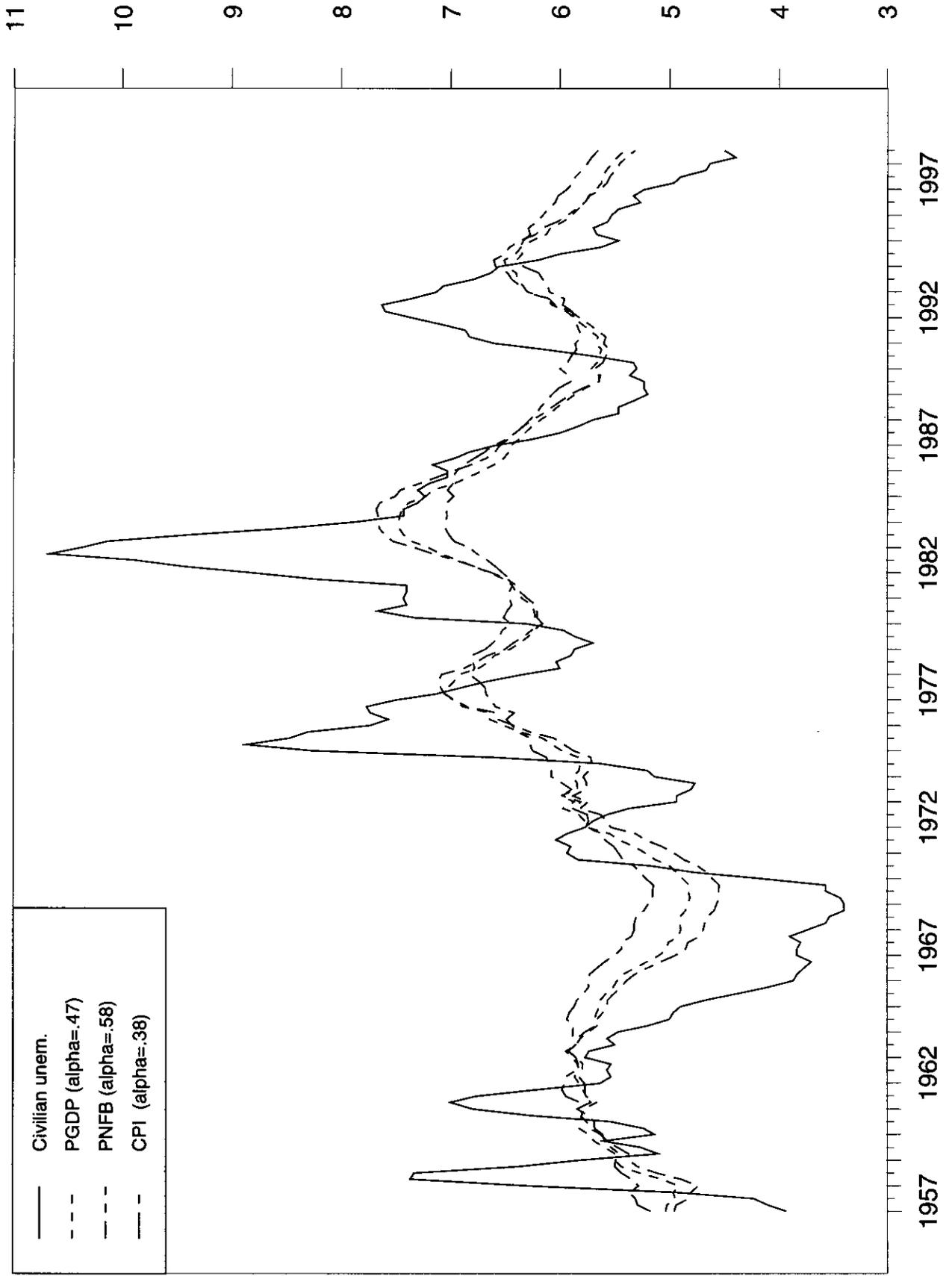
**Figure 3: Switching to Civilian Unemployment Basis**

Estimates using a 24-quarter moving average of lagged unemployment.



**Figure 4: Switching to Civilian Unemployment Basis**

With models using short inflation lags and near-term unemployment dynamics.



**Figure A.1: Long-Term Unemployment Is Not Important in the United States**

Unemployment as a percent of total civilian labor force, adjusted for 1994 survey revisions.

