An Empirical Assessment of "New Theories" of Inflation and Unemployment

Stephen K. McNees

Introduction

It is generally considered impolite for a host to criticize his guests and tell them what they should not discuss. Nevertheless, that is exactly what I propose to do this morning.

I want to start by saying what I think this conference is *not* about. It is not about either the Keynesian (or aggregate demand) explanation of unemployment or the monetarist explanation of inflation. There are mounds of both theoretical and empirical work on each of these propositions. We have already formulated strong prior opinions on each so that it would be too much to hope this conference could resolve our views on these time-honored propositions.

The role of this conference *is*, instead, to advance non-Keynesian views of the determinants of unemployment and nonmonetary views of the inflation process. The role of this paper is to summarize some of the empirical evidence on these "new theories" of inflation and unemployment. Let me warn you now, the preliminary verdict is not good. (I must confess, however, this judgment also springs mainly from my prior opinions – otherwise, we would have no excuse for holding this conference.)

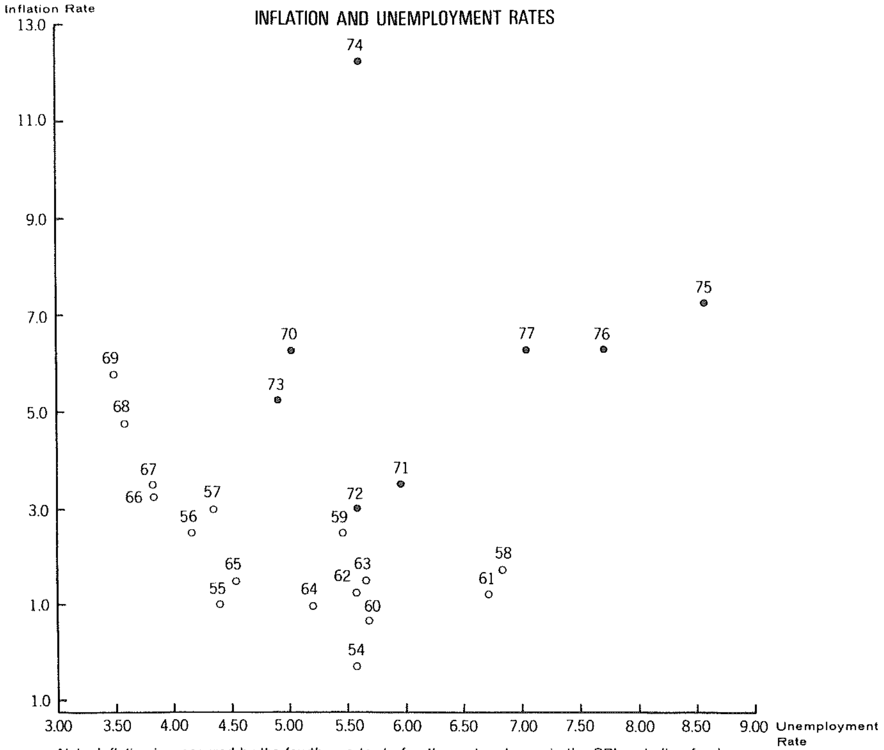
The Keynesian and monetarist propositions can be combined and restated to imply that the rate of inflation is *directly* related and the rate of unemployment is *inversely* related to the strength of aggregate demand (which may wholly or partly reflect the rate of monetary growth). In other words, these two time-honored propositions are consistent with a simple short-run Phillips curve, depicting an inverse relationship between inflation and unemployment rates.

The simple inverse or "Phillips curve" relationship provides a fairly accurate description of the inflation and unemployment rate data for the United States in the 1950s and 1960s, as illustrated by much of the economics literature of that period and by the open circles in Figure 1. In contrast, a positive relationship indicated by the filled circles has often been observed so far in the 1970s, particularly in 1970, 1972, 1974, and 1976. Many of these deviations from the "normal" negative relationship could be accounted for by appealing to "external" or "special" factors, such as extreme "wage distortion" in 1970, wage and price controls in 1972, and the oil price shock in 1974. In short, the inflationary

Stephen K. McNees is Assistant Vice President and Economist, Federal Reserve Bank of Boston. The author is grateful to his colleagues Richard Kopcke and Geoffrey Woglom for helpful comments. He also thanks Neil Berkman for instructing him on time series modeling and Elizabeth Berman for research assistance.

Figure 1

INFLATION AND UNEMPLOYMENT RATES



Note: Inflation is measured by the fourth-quarter to fourth-quarter change in the CPI excluding food.

experience of the 1970s has led empirically oriented nonmonetarists, primarily the builders of large-scale structural econometric models, to attempt to refine the role of supply factors, government policies, and international economic considerations in price determination. (See, for example, Klein [1978].)

Even before the turbulence of the 1970s, the Phillips curve was not without its critics. In the late 1960s, Friedman (1968) and Phelps (1967) criticized the inverse or "trade-off" notion and proposed instead the concept of a "natural" unemployment rate (NUR). The NUR idea gained currency as it was capable of accounting for the experience in 1970 and 1972 without resorting to "special factors." However, in the early 1970s, the original specification of the NUR theory also came under attack. On a theoretical level, Lucas (1972) criticized the "adaptive expectations" mechanisms which characterized early attempts to implement the NUR theory empirically. The early adaptive expectations version of the NUR theory could not simply reconcile the paths of inflation and unemployment in 1973-75 and the failure of inflation to decelerate during the subsequent recovery period.[1] Just as the failure of the Phillips curve has sent nonmonetarists back to their drawing boards to explain price behavior in the 1970s, acceptance of the NUR concept shifts the focus of non-Keynesians' attention to the formulation and measurement of expectations and to explanations of the cyclical behavior of the unemployment rate. One of the earliest and clearest empirical implementations is Sargent's (1973) combination of the NUR hypothesis with the assumption of rational expectations (NUR-RE).

The original objective of this paper was to take a preliminary look at the empirical success of these "new theories." The first part of the paper considers the NUR-RE model and the second part deals with various nonmonetarists' attempts to refine or replace the Phillips curve. As the research progressed a secondary objective developed – to explain the difficulties in determining the "empirical success" of these (or any) theories. Some of the evidence is taken from ex ante (before the fact) forecasting situations, in which no information about the future can be known with certainty. Some of the evidence is from ex post (after the fact) model simulations in which the actual historical values of the exogenous variables are used to solve the model. Some of the ex post simulations (and, of course, all ex ante forecasts) are post-sample – i.e., they pertain to a period subsequent to the period to which the model was fit. Some of the ex post simulations are in-sample, i.e., they show how well the model tracks the period from which the model coefficients were estimated. No single type of evidence will be regarded as conclusive by everyone. This paper, therefore, presents a variety of different types of evidence and employs several

[1] The simplest version of the NUR hypothesis also had some problems in the early 1960s when a stable inflation rate was associated with unemployment rates higher than anyone then (and almost anyone now) would have measured the "natural rate." This problem could be remedied by raising one's estimate of the NUR in the early 1960s but this solution would only come at the cost of rendering earlier periods like 1954 and 1959 inexplicable. The point is that without some modicum of agreement about how to measure the NUR, how to describe and measure the formulation of expectations, and some attention to the appropriate (stable?) lag structure, the NUR theory is without empirical content.

different imperfect (i.e., not definitive) tests of these theories. The reader must determine whether it is the theories that succeed or fail or whether it is the tests themselves that fail. The hope is that the paper will contribute to a greater appreciation of both the importance and the difficulty of attempting to evaluate theories on an empirical basis.

Tests of the Rational Expectations Version of the Natural Unemployment Rate Hypothesis

Sargent (1973) tested the natural unemployment rate hypothesis under the assumption of rational expectations (NUR-RE). His test exploits this theory's strong implication "that the 'innovation,' or new random part of the unemployment rate, cannot be predicted from past values of any variables, and that it cannot be affected by movements in past values of government policy variables." (p. 451). Sargent's model is a simple third-order autoregression for the unemployment rate, following the implication of the theory "that there is no better way to predict subsequent rates of unemployment than fitting and extrapolating a mixed autoregressive, moving-average process in the unemployment rate itself." Adding lagged values of wages and prices did not improve the fit of the basic Sargent model so that the NUR-RE model could not be rejected on the basis of that test. However, when a larger set of information, including the money supply and government deficits, was included, there was a statistically significant improvement in the fit, requiring "rejection of the version of the natural rate hypothesis that assumes rational expectations formed on the basis of this expanded set of information." (p. 453). While this rejection can hardly be taken as support for Sargent's hypothesis, there are good reasons to reject the test itself rather than the model. First, Sargent cites several econometric reasons for interpreting the rejections with caution. In addition, he correctly notes that his tests

> have not been shown to be of comfort to advocates of any particular alternatives to the natural rate hypothesis. That is, it has not been shown that an autoregression for unemployment yields ex ante predictions of unemployment inferior to those of a particular structural macroeconometric model that embodies a particular aggregate supply theory other than the natural rate hypothesis. A particular alternative aggregate supply hypothesis might well be able to predict unemployment better than an autoregressive moving-average process, but there is no way of knowing for sure until a horse race is held.

Sargent cites Nelson (1972) on the performance of the FRB-MIT-PENN model as evidence for his assertion that he was "aware of no evidence that shows that any particular existing structural model embodying a specific alternative to the natural rate hypothesis can outperform it in predicting the course of the unemployment rate." (p. 464). He urges "that the natural unemployment rate hypothesis [with rational expectations], . . . be tested against specific competing hypotheses by setting up statistical prediction 'horse races.' " (p.451).

To the best of my knowledge, no one has accepted Sargent's challenge. Below, I present one test like Sargent's along with several types of "horse races" between the Sargent model and various alternative models and predictive procedures.[2] Although no single statistical test is sufficient to declare "a winner," it is hoped that the battery of tests will provide some indication of the empirical success of the competing hypotheses.

The first test is much like Sargent's – an examination of whether the addition of the other economic variables significantly improves the *within sample fit* of the Sargent model. It would be of little interest to find, after an exhaustive search of economic time series, *some* variable that is correlated with the residuals of the Sargent model and thus could improve its in-sample fit.[3] I have chosen, therefore, to test only the explanatory variable that would probably first occur to a practical forecaster conversant with "Okun's law" – the GNP gap (see Okun [1962]). The result of adding the gap, lagged one-period, to the Sargent model is given below:

$$\underset{}{UR_t} = \underset{(.28)}{1.70} + \underset{(.14)}{1.14}\, UR_{t-1} - \underset{(.16)}{0.72}\, UR_{t-2} + \underset{(.09)}{0.19}\, UR_{t-3} + \underset{(.03)}{0.15}\, GAP_{t-1}$$

$\bar{R}^2 = 0.9569$; S.E. = .292; D.W. = 1.92

Period of fit: 1952:2–1977:4. GAP is based on the Council of Economic Advisers' definition of potential GNP.

The t-statistic on the lagged value of the GNP gap is 5.15, highly significant statistically. Consequently, this application of Sargent's test, like his own second application, requires rejection of Sargent's version of the natural rate hypothesis with rational expectations.

For the reasons Sargent has noted (p. 453), the result of this in-sample test, while certainly not favorable, cannot be regarded as conclusive grounds for rejecting the model. He rightly encourages post-sample "horse races" between his model and alternative competitors.

Table 1 presents three "horse races" between the Sargent equation and alternative predictive techniques. All of the predictions are outside of the sample – the Sargent equation was reestimated each quarter up to the start of the prediction period (using the latest version of the actual data) and extrapolated forward dynamically.

[2] The Sargent model is defined as the third-order autoregression he used in the 1973 tests. Sargent's period of fit was 1952:1 through 1970:4; when the equation is reestimated through 1977:4 the fit improves somewhat, the standard error holds constant, and the coefficients, on the basis of a Chow test, are not significantly different. There is presumably, therefore, no reason to believe this specification is not still representative of the natural rate *cum* rational expectations "new theory" of the unemployment rate.

[3] This is undoubtedly the major reason why Sargent so heavily discounts the results of his second test (pp. 452–53) which is based on the addition of three lagged values of eight economic variables the selection of which was unmotivated and therefore apparently unabashedly ad hoc.

TABLE 1

Post-Sample Test of the Sargent Equation
Root Mean Square Error
(cumulative changes, percentage points)

A. vs. Ex Post Dynamic Simulation of an Econometric Model

Simulation period: 1969:2–1977:4

	Forecast Horizon (quarters)					
	1	2	3	4	5	6
Sargent	.3	.7	1.0	1.3	1.5	1.7
Fair, EM	.4	.6	.8	.8	.8	.8

B. vs. Subjectively Adjusted, Ex Ante Forecasts

Forecast period: 1970:3–1977:2

	Forecast Horizon (quarters)					
	1	2	3	4	5	6
Sargent	.3	.8	1.1	1.4	1.6	1.8
ASA	.2	.4	.7	.9	1.0	–
Chase	.3	.6	.8	1.0	1.2	1.4
DRI	.3	.5	.7	.9	1.1	1.3
Wharton	.3	.6	.8	1.0	1.1	1.1

C. vs. Mechanically Generated Ex Ante Forecasts

Forecast period: 1970:3–1975:2

	Forecast Horizon (quarters)			
	1	2	3	4
Sargent	.3	.8	1.1	1.3
Fair, FM	.3	.7	1.0	1.1

SOURCES: The Fair econometric model (EM) data are from Fair (1978) Table 4. The subjectively adjusted ex ante forecast data are from McNees (1977). The Fair forecasting model (FM) data are from McNees (1975). For each test, the Sargent equation (1973) was reestimated with the latest actual data from 1952:1 through the quarter before the extrapolation period. The Fair econometric model was also reestimated repeatedly through two quarters before the simulation period.

The first test (Panel A) is a comparison of the Sargent equation and ex post, post-sample dynamic simulations of the Fair econometric model (1974). Ex post simulations, in which the actual past and future values of the exogenous variables are used to generate the predictions, are traditionally used to test a model's validity. A structural econometric model is based on the proposition that there is important information in the (future) values of the exogenous variables. The Sargent model contains no exogenous variables. A defender of the Sargent approach could argue that this comparison is biased in favor of the econometric model whose ex post errors reflect information on the actual, future values of the exogenous variables in the model.

Panel B presents a comparison with the ex ante (or before the fact) forecasts of three of the major econometrically based forecasting services as well as the median forecast from the American Statistical Association/National Bureau of Economic Research survey. The forecasts were formulated before the fact and clearly, therefore, do not benefit from any certain information about the future. Although these forecasts were based on an econometric model, they are not strictly "scientific" (in the sense of being mechanically replicable) because the model forecasts are subjectively adjusted by the model proprietor. These forecasts can benefit (or suffer!) from the forecasters' subjective opinions about the future.

The last test (Panel C in the table) is a comparison of the Sargent equation and the ex ante forecasts which were mechanically generated with the Fair forecasting model (1970). In order to solve a model some estimate of the future values of the exogenous variables must be made. The future values of many of the exogenous variables were taken from external sources available at the time the forecast was made. The values of the other variables appear to have been chosen on the basis of fairly simple, mechanical rules involving a minimal amount of judgment. Once the exogenous variables were chosen, no subjective adjustments were made to the "pure model" results to account for events such as wage and price controls or increases in the price of imported oil. This test, which excludes both subjective adjustments and exogenous variable certainty, does not appear to contain any bias in favor of the structural model.

The results of these three tests are similar and, hence, easily summarized: The Sargent equation's one-quarter-ahead post-sample predictions are about the same as those based on alternative techniques. However, the Sargent equation does exhibit a distinctly stronger tendency toward error accumulation when extrapolated dynamically over a longer horizon.

Interpretation of this result is not as straightforward – the glass can be viewed as half empty or half full. A defender of the Sargent approach would stress the similarity of the one-period result and would note that the multi-period results for the alternative approaches incorporate external information – subjective (in Panel B) or objective (in Panel A) – which, it could be argued, biases the multi-period test against the Sargent equation. The results in Panel C, where there are no apparent biases, are probably too similar to draw a statistically rigorous verdict.

A critic of the Sargent equation could argue that the comparisons in panels B and C are biased in favor of the Sargent equation because it was estimated with and judged against the latest revision of the data whereas in a realistic ex ante forecasting situation even recent history is uncertain. As for the incorporation of external information, this is an inherent difference between econometric and time series modeling. Placing the econometric models in an ex ante forecasting situation puts each approach on an equal footing with respect to using only "historical" rather than "future" information for the forecast. Time series models, by their very nature, are restricted to using a limited amount of information in arriving at their forecasts.[4]

Summary and Assessment of the Evidence on the NUR-RE Model

In the strictest sense, a time series model, such as Sargent's NUR-RE model, and a structural, econometric model are not comparable. The former contains no exogenous variables while the latter inherently must. This standard is too strict for most who strive to have some informed opinion on the relative importance of Sargent's NUR-RE model and its alternatives. On the basis of one of Sargent's tests and the similar one conducted above, Sargent's model can be rejected on rigorous statistical grounds. While this evidence ought not to be ignored, Sargent's verdict that these tests must be interpreted with caution is sound. The results of in-sample tests cannot be regarded as conclusive. For this reason, three post-sample tests were conducted. The post-sample results show a disparity between the single-period and the multi-period results.[5] One-period-ahead, the NUR-RE model performs about as well as the alternative approaches. On the basis of this evidence *alone,* the results are inconclusive – whatever differences that would emerge by presenting the data to more decimal places could surely not be regarded as significant in a statistical sense. In the multi-period results, the NUR-RE model exhibits a greater tendency toward error accumulation as the horizon extends further into the future. This may be due to the linear specification of the NUR-RE model. It may also be due to the enhanced value of the information in the exogenous variables over longer horizons. This result appeared in three different tests, each containing a different type of information: a) ex post simulations (using actual values of the exogenous variables), b) subjectively adjusted ex ante forecasts, and c) ex ante forecasts with no subjective adjustments and mechanical selection of the values of the exogenous variables. In light of the small number of post-sample observations and the small differences in the summary error statistics, the multi-period results may be insufficient for making a statistically rigorous rejection of the NUR-RE model. Nevertheless, if *any* importance is assigned to either the in-sample results or the multi-period results, the case for Sargent's NUR-RE model stands unproven.

[4] This matter is discussed more fully below. For an alternative method of accounting for exogenous variable uncertainty, see Ray Fair, "Estimating the Expected Predictive Accuracy of Econometric Models," Cowles Foundation Discussion Paper No. 480, January 1978, where he develops and estimates standard errors for econometric and autoregressive models. His results for the unemployment rate are similar to those reported in Panel A of Table 1.

[5] A more complete discussion of the problems of interpretation in comparing single-period and multi-period results of time series models and structural models appears in McNees (1978).

Tests of Nonmonetarist Wage and Price Models

The term "nonmonetarist wage and price models" clearly covers a variety of different approaches which probably should not be lumped into one amorphous phrase. The term may have been more applicable to the circa-1970 vintage of wage and price models but modelers have reacted differently to the dramatic events of the 1970s. Most have chosen to refine the Phillips curve approach by incorporating additional equations representing supply phenomena, while some have taken new (e.g., "stage-of-processing") approaches. It is one of the goals of this conference, but beyond the scope of this paper, to describe and catalogue these efforts.

The best test of a model is its post-sample performance. The opening section presents recently published post-sample assessments by two model builders. Post-sample assessments of other wage and price models are not readily available and it is perilous for an outsider to attempt to reestimate others' models because special data and estimation techniques are often used. On the other hand, it is fairly easy to perform simulation experiments with current versions of structural models and these are also presented below. Although these in-sample simulations are not sufficient to establish the validity of the models, fitting the historical data relatively well is the logical first check of a model's performance.

Post-sample Results

Robert J. Gordon (1977) analyzed the post-sample performance of his wage-price model originally fit through mid-1971.[6] When the price equation is refit over the same period using the latest revised data, several of the coefficients change substantially and the fit deteriorates somewhat (the standard error increased 18 percent). (See Table 2, A1 and A2.) More importantly, the refitted equation does not work well outside the sample period. The post-sample root mean square error (RMSE) was 2.4 percent (at a simple annual rate), nearly two and one-half times larger than the in-sample standard error. This increase is large enough to support the conclusion that the model fitted to the sample period was a poor representation of the post-sample events which were to follow.[7] In addition, the post-sample RMSE for Gordon's price measure (the deflator for nonfood business product net of energy) is 50 percent larger than the RMSE of ex ante forecasts of the more volatile implicit GNP price deflator (IPD) over the same period. In addition, the post-sample errors accumulate

[6] It is important to note that although the equation is similar to the one Gordon originally proposed in 1971, the equation "was altered somewhat in 1975 and thus incorporates knowledge of events to that point." (p. 264).

[7] Under the null hypothesis that the estimated coefficients and the standard error of estimate computed from the sample period accurately represent the post-sample structure of the mechanism generating the variable of interest, the ratio of the mean squared error and the square of the standard error of estimate is distributed as an F statistic. If the null hypothesis were true in this case, the F value is highly unlikely to exceed a critical value of about two. In this and the following instances the ratio exceeds two, indicating an inappropriate model and/or a particularly misleading sample-period draw.

TABLE 2

Wage-Price Models:
Post-Sample Performance
(Simple annual growth rates)*

A. Gordon price equation:	
1) Standard error, original data	.8
2) Standard error, revised data	1.0
3) RMSE post-sample (1971–76)	2.4
4) RMSE Ex Ante IPD forecast error (1971–76)	1.6
B. Gordon wage equation:	
1) Standard error, original data	.5
2) Standard error, revised data	.6
3) RMSE, post-sample	
Price deflator	
a) private nonfood business product net of energy	1.5
b) private nonfarm business	2.1
4) RMSE Ex Ante forecast error	1.7

C. Fair price (IPD) equation, RMSE of cumulative percent changes:

	1969:II–1977:IV Forecast Horizon (quarters)							
	1	2	3	4	5	6	7	8
1) Fair Model	2.00	1.86	1.91	1.97	1.99	1.97	1.96	1.92
2) Naive Model	1.88	1.96	2.12	2.36	2.61	2.82	3.06	3.26
3) Ex Ante Forecast	1.58	1.78	2.05	2.26	N.A.	N.A.	N.A.	N.A.

D. Fair wage equation, RMSE of cumulative percent changes.

	1969:II–1977:IV Forecast Horizon (quarters)							
	1	2	3	4	5	6	7	8
1) Fair Model	3.12	2.50	2.28	2.31	2.31	2.33	2.32	2.32
2) Naive Model	2.68	1.88	1.61	1.47	1.37	1.34	1.37	1.42
3) Ex Ante Forecast	2.42	1.74	1.55	1.52	1.60	1.66	1.68	1.63

*To facilitate comparisons, all data were converted to simple annual rates.

SOURCES: The model data are from Gordon (1977) and Fair (1978). Gordon's price variable is the deflator for nonfood business product net of energy. The ex ante price forecasts are of the implicit GNP price deflator (IPD) and are the median from the ASA/NBER survey. Gordon's wage variable is his own measure of the wage rate. The ex ante wage forecasts are Wharton EFA's forecasts of its own compensation measure and are not, therefore, strictly comparable to either Gordon or Fair. The Fair model (naive model) was reestimated repeatedly through two quarters prior to (one quarter of) the simulation period.

leading the equation to "overpredict inflation during 1971–76 very substantially." (p. 258). The error accumulation problem can be remedied by using a different proxy for excess demand and by constraining the sum of the coefficients on labor cost to equal 1.0 but even this altered version of "the best equation" exhibits a post-sample RMSE of 2.7 percent, nearly three times larger than its in-sample standard error and an even larger multiple of the RMSEs of the ex ante forecasts of that period. When the price equation is refit through 1976, however, the in-sample standard error falls to a little more than 1.0 percent at an annual rate.

A parallel story can be told for Gordon's wage equation. When the original 1971 wage equation was reestimated with revised data, the fit deteriorated only slightly but the coefficients were unstable. The data revisions rendered one of the proxies for labor market tightness, unemployment-dispersion, insignificant and vindicated the natural rate hypothesis in that wage changes fully incorporate changes in [product] price inflation. "As in the case of the structural price equation, the post-sample extrapolation errors of the wage equation are vastly larger than the in-sample standard error." (p. 268) More precisely, the post-sample RMSE increased to about 1.5 percent at a simple annual rate, nearly three times the standard error, using Gordon's preferred price measure (excluding food and energy) and to more than 2.0 percent, four times the standard error, using a broader alternative index. These compare with the RMSEs of ex ante forecasts which are about 1.7 percent over a two-quarter horizon.

Constraining the equation to conform to the adaptive expectations version of the NUR hypothesis cut down the post-sample error accumulation but even this constrained equation had a RMSE nearly two and a half times (using the deflator for private nonfood business product net of energy as the price measure) or more than three times (using the deflator for private nonfarm business) the in-sample standard error. When the equation was refit through 1976, the standard error declined to about .6 percent at an annual rate.

This evidence shows that Gordon's original price and wage model did not perform well outside the fit period. Even with the "best" respecification, the post-sample errors are larger than those made by ex ante forecasts at the time. When the equations are refit through 1976, their standard errors fall back to near those of the original specification. However, in light of the poor post-sample performance of the original specifications, there is no assurance that the later specifications will perform successfully outside the period of fit.

Ray Fair (1978) analyzed the post-sample properties of his econometric model. Some of his results for the wage and price variables in his model are presented in the bottom part of Table 2.[8] Panel C of the table presents the

[8] Fair also conducts stochastic simulations to compute the estimated standard errors of his model and a naive autoregressive model. Fair's ingenious method permits decomposition of the standard error into four alternative sources – stochastic error terms, coefficient estimates, exogenous-variable forecasts, and the degree of misspecification of the model. (See Table 1, pp. 27–28). The RMSEs of ex post simulations presented above do not, of course, reflect errors due to incorrect selection of values for exogenous variables. Nevertheless, Fair's estimated standard errors and the RMSEs are very similar, quantitatively *and* qualitatively, for the deflator and the wage rate.

post-sample RMSEs of the model's price (IPD) predictions along with the corresponding measures for a "naive" autoregressive model. The naive model consists of regressing each variable on a constant, a linear time trend, and its first eight lagged values. For purposes of comparison the RMSEs of the median ex ante forecast of the ASA/NBER survey over the same time period are also shown. The performance of the Fair model price predictions is mixed. Over very short horizons, the Fair model simulations are somewhat inferior to the ex ante forecasts and about the same as the naive model. Over horizons of a year or more, the Fair model simulations are more accurate than the naive model and appear to be more accurate than the ex ante forecasts (although precisely comparable ex ante data were not available). The ex post model simulations clearly benefit from using actual exogenous variable values, the importance of which probably increases with the forecast horizon. Although these are post-sample simulations (so that the coefficients are not estimated with actual data from the simulation period), it should be noted that the model was first specified in 1974, well into the test period, and that some modifications were made as late as 1977. A conclusive assessment of the Fair model's price performance must await the accumulation of more post-specification experience.

Ex post simulations of the wage rate, shown in Panel D, were disappointing. The naive model outperforms the structural model in forecasting the wage rate over all horizons. The same result holds for a comparison with set of ex ante forecasts of compensation, shown in row D3, although these results are not strictly comparable because they pertain to a different variable and a slightly different forecast period. Nevertheless, these results are not encouraging for the structural model's ability to explain wage behavior. Taken at face value, the model contributes no additional explanatory power to that of a naive, purely statistical model.

Thus, there is considerable evidence that *no* 1970–71 vintage wage-price model is capable of explaining the wage-price behavior of the 1970s at all adequately. As a consequence, those earlier specifications are now obsolete – probably *all* serious wage-price models have been respecified during the last few years. Rather than continue to search the entire 1970–71 vintage of wage-price models, it would seem more fruitful to focus attention on the performance of the current stock of wage-price models.

In-sample Results

What have we learned about wage-price determination from the 1970s experience? Were the large errors in the 1970s unavoidable (in the sense of being due to noneconomic events) or could different specifications have tracked the wage-price behavior in the 1970s? In order to address this question, I collected one- through eight-quarter-ahead dynamic simulations of the current versions of some of the most prominent macroeconometric models – DRI, Fair, FMP, Michigan, and Wharton. An assessment of these in-sample results must be conducted with extreme care because *cross-model comparisons are unwarranted and virtually certain to be misleading.*[9]

Even though cross-model comparisons are not warranted, simulation performance cannot be evaluated in a vacuum. Because the absolute standard of perfection is unrealistic, prediction performance must be evaluated relative to some standard of comparison. Two standards of comparison are considered here: (1) each model simulation is contrasted with a companion "naive" or purely statistical time series model (which contains no exogenous variables); (2) each model simulation is also compared with ex ante forecasts.

(1) *Time Series Model as a Standard of Comparison*

A time series model was fit for the implicit GNP price deflator (IPD) over the period 1954-1977. The estimated equation is:[10]

$$\Delta IPD_t = \underset{(6.6)}{.6555}\, \Delta IPD_{t-1} + \underset{(2.8)}{.2827}\, \Delta IPD_{t-2} + \underset{(1.5)}{.0767}$$

$$\hat{\sigma}_u = .327$$

[9] To appreciate this fact one must recall that an econometric model is conditioned upon a set of exogenous variables (the values of the predicted variables are determined by the set of predetermined variables). Models inherently differ with respect to the size and composition of the set of exogenous variables on which they are based. To attempt to standardize for use of exogenous variables would violate the intended structure of the model. (Requiring all models to adopt a common set of assumptions with regard to the *values* of the exogenous variables does not violate the integrity of the model. This is, in fact, what is done in ex post comparisons where all models use the actual values of exogenous variables. Standardizing on a common set of exogenous variable values is a useful exercise which, unfortunately, is seldom followed in an ex ante context.) A simple example is the treatment of the period of wage and price controls (both imposition and relaxation). Some of the models *include* dummy variables in their estimated equations to account for controls. Other models *exclude* any special treatment of the controls period. To standardize the treatment of the controls episode among the models, i.e., to force all the models to conform to a certain procedure, would, at a minimum, require reestimation of several models and quite possibly would call for respecification of their basic structure. This standardized set of models would not be representative of those models in current use.

An additional obstacle to cross-model comparisons is that, due primarily to data problems, different models are capable of being simulated over different historical intervals. Specifically, some models cannot be simulated for the early postwar period due to the lack of data for some key series. Experience has shown that the forecast interval is a major source of variation in prediction errors. To standardize for simulation intervals would require discarding some of the available data, which are all too limited in any case.

[10] A chi square test (X^2_{21} = 12.08) indicates that the hypothesis that the residuals are white noise could not be rejected at a high (94 percent) level of significance. It is of some interest to note that the fit is better than that which Nelson (1972) obtained with a different time series model. Nelson's specification was not satisfactory over the longer sample period. suggesting that some time series models may share with some econometric models the property of structural instability.

Table 3 presents the in-sample ex post dynamic simulation RMSEs of five econometric models relative to those of this time series model. With a single exception, the current vintage of econometric models are able to track history more accurately than the time series model. The margin of superiority is inversely related to the forecast horizon – the one-period RMSEs are generally fairly close, the two-period errors substantially smaller with the margin of superiority increasing steadily through the eight-quarter horizon (where all of the models are at least as good as their companion time series model).

These results need to be interpreted with the following qualifications: (a) the model simulations were taken "as is" with no attempt to standardize for selection of exogenous variables, or simulation intervals; (b) the time series model, by design, does not benefit from information contained in the actual values of the exogenous variables which, by design, are presumed of great importance in the structural models.[11] The evidence is consistent with the hypothesis that this information takes on greater importance as the forecast horizon is extended and/or the alternative hypothesis that the nonlinearity of the econometric models is a greater advantage for longer horizons.

These results are consistent with several previous studies: (a) Nelson (1972) found that the one-period-ahead RMSE of IPD predictions of the FMP model was smaller than that of a time series model both in-sample (.195 vs. .230) and post-sample (.261 vs. .346); (b) Hirsch, Grimm, and Narasimham (1976, p. 245) found substantially smaller RMSEs for various versions of the BEA model than a time series model for one-period forecasts (.45 to .59 vs. .75 index points) and the margin of superiority increased as the forecast horizon was extended.

Even though some earlier studies reached different conclusions, the recent evidence indicates that most of the current stock of econometric models fit the historical data for IPD better than a time series model and better than earlier vintages of structural models. While one can hope that this superior performance is an indication that current models have captured the "true" wage-price structure more completely, one must await their post-sample performance to be sure.

(2) *Ex Ante Performance as a Standard of Comparison*

Much of the dissatisfaction with the current state of empirical macroeconomics undoubtedly stems from the errors of ex ante forecasts that were issued for the period 1973-1974. Many believe these were "poor" forecasts in the sense that it is now known (or perhaps known even then) that important factors were overlooked or misassessed. Nevertheless, it is obvious that there is a certain irreducible, minimum feasible error inherent in any attempt to prophesy the future. Because this "innovation" undoubtedly varies over time, it would be inappropriate to attach much significance to a large *absolute* error. The ex post simulations provide an interesting standard of comparison for evaluating these errors. They address the hypothetical question: Knowing what we know now

[11] Howrey, Klein, and McCarthy (1976) have argued "that sample-period mean squared error comparisons of the autoregressive and structural models are not powerful tests," particularly when confined to one-period horizons.

TABLE 3

Ex Post Dynamic Simulations of Prices:
Structural and Time Series Models
RMSE (cumulative percent change at simple annual rate)

Model	Simulation Period	Forecast Horizon (quarters)							
		1	2	3	4	5	6	7	8
DRI	1965–77	1.2	.9	.6	.6	.5	.5	.5	.4
ARIMA		1.2	1.5	1.5	1.7	1.7	1.8	2.1	2.4
FAIR	1960–77	1.3	.8	.7	.8	.8	.8	.7	.6
ARIMA		1.3	1.3	1.3	1.4	1.6	1.6	1.8	2.0
FMP	1961–77	1.0	1.0	.9	1.0	1.1	1.1	1.1	1.2
ARIMA		1.4	1.4	1.3	1.5	1.6	1.6	1.8	2.1
MICHIGAN	1956–77	1.0	.6	.5	.5	.6	.5	.5	.5
ARIMA		1.3	1.3	1.2	1.4	1.5	1.5	1.7	1.9
WHARTON	1963–77	3.4	2.6	2.4	2.5	2.3	2.2	2.1	2.2
ARIMA		1.3	1.5	1.4	1.6	1.7	1.8	2.0	2.2

NOTES: The times series model's RMSEs are in-sample dynamic extrapolations of the (2, 1, 0) model described in the text. The in-sample dynamic model simulations were generously supplied by Otto Eckstein and Frank Cooper of Data Resources, Inc. (DRI), Ray Fair of Yale University (Fair), Jared J. Enzler, Board of Governors of the Federal Reserve System, using the Federal Reserve–MIT–University of Pennsylvania model (FMP), Saul H. Hymans of the University of Michigan (Michigan), and Lawrence R. Klein and Richard M. Young, Wharton Econometric Forecasting Associates, Inc. (Wharton). Each simulation started from the fourth quarter of the preceding year so that for DRI, for example, there are 13 one-quarter through four-quarter-ahead observations and 12 five-quarter through eight-quarter-ahead observations.

(including the actual values of the exogenous variables), how accurate might our predictions have been?

Table 4 presents a comparison of ex post and ex ante forecast errors over the 1970s. These data suggest that the current version of four of the five econometric models can, with hindsight knowledge of the actual values of the exogenous variables, simulate the course of inflation in the 1970s considerably more accurately than the ex ante forecasts of those times. However, these ex post simulations benefit both from using actual values of the exogenous variables and from being in-sample, i.e., using actual values to estimate their coefficients. Some indication of the benefits of using actual data to estimate the model can be gained from contrasting the in-sample and post-sample performance of the time series model. (For the post-sample extrapolations, the time series model was repeatedly reestimated up to the forecast period.) Row 1 in the table shows that the in-sample performance of the time series model is about the same as the ex ante forecasts and that the post-sample performance is considerably worse.

TABLE 4

Ex Post Simulations and Ex Ante Forecasts
of Inflation in the 1970s

RMSEs (cumulative percent changes at simple annual rates)

Model/Forecaster	Forecast Horizon (quarters) 1	2	3	4
1) ARIMA, in-sample	1.4	1.7	1.8	2.1
ARIMA, post-sample	1.7	2.3	2.5	2.9
2) DRI, ex post	1.2	.9	.6	.5
DRI, ex ante	1.6	2.0	2.0	2.4
3) Fair, ex post	1.7	1.0	.7	.7
ASA, ex ante	1.5	1.6	1.9	2.3
4) FMP, ex post	1.0	1.2	1.3	1.5
ASA, ex ante	1.5	1.6	1.9	2.3
5) Michigan, ex post	.8	.5	.4	.3
ASA, ex ante	1.5	1.6	1.9	2.3
6) Wharton, ex post	4.8	3.6	3.4	3.6
Wharton, ex ante	1.6	1.5	1.6	1.9

NOTES: The ex ante data are from McNees (1977). The ex post simulations are those described in Table 3. Each simulation was started in the fourth quarter, so that each RMSE is based upon only seven observations.

The ex post simulations benefit similarly and thus give an overly optimistic impression of the maximum accuracy which could have been expected.

Summary and Assessment of the Evidence on Wage and Price Models

Current specifications of macroeconometric models can simulate the in-sample movements of wages and prices well relative to a time series model and well also relative to the forecasts that were released ex ante. However, these ex post, in-sample results may provide an optimistic impression of the models' post-sample performance or their future performance in ex ante situations where the future values of the exogenous variables are unknown. As was noted above, previous wage-price models which also fit the historical data at the time did not perform nearly as well outside the same period *or* relative to autoregressive models *or* relative to the ex ante forecasts. Have we learned something from the 1970s' experience, or do we just think we have learned something? Will the current vintage of econometric models suffer the same post-sample fate as their predecessors?

These are empirical questions that cannot be answered with any degree of certainty with data presently available. The fundamental empirical problem is the size and frequency of institutional changes or "external shocks" which occurred in the 1970s. Among the more obvious are the following: (1) The imposition and relaxation of several different phases of wage and price controls. (2) The switch from fixed to flexible exchange rates and the subsequent experience of learning to live with exchange rate induced variations in prices of traded commodities. (3) The sudden quadrupling of the price of imported oils. (4) The changes in demographics, public policy, and social attitudes and their alleged impact on the "natural" rate of unemployment. (5) The growing importance of governmentally mandated supply-restricting or cost-raising measures. (6) The introduction of a new framework for conducting monetary policy.

All of these "special factors" were intertwined, both temporally and causally, and superimposed upon the most extreme business cycle of the postwar period. Econometric model-builders responded to these events by respecifying (sometimes repeatedly) their empirical statement of the wage-price process. Even if a more complete list of the determinants of wage-price behavior has now been identified, it is not clear that our experience with these new institutions has been sufficiently long to quantify the independent influence of each of these factors.

REFERENCES

Fair, Ray. *A Short-Run Forecasting Model of the United States Economy.* Lexington, Mass: D.C. Heath & Co., 1971.

———. *A Model of Macroeconomic Activity. Volume I: The Theoretical Model.* Cambridge, Mass.: Ballinger Publishing Co., 1974.

———. *A Model of Macroeconomic Activity. Volume II: The Empirical Model.* Cambridge, Mass.: Ballinger Publishing Co., 1976.

———. "Estimating the Expected Predictive Accuracy of Economic Models." Cowles Foundation Discussion Paper, No. 480, January 17, 1978.

Friedman, Milton. "The Role of Monetary Policy." *American Economic Review,* March 1968, Vol. 58, pp. 1-17.

Gordon, Robert J. "Can the Inflation of the 1970s be Explained?" *Brookings Papers on Economic Activity,* 1977:1, pp. 253-279.

Hirsch, Albert A.; Grimm, Bruce T.; and Narasimham, Gorti V.L. "Some Multiplier and Error Characteristics of the BEA Quarterly Model." In *Econometric Model Performance,* edited by Lawrence R. Klein and Edwin Burmeister. Philadelphia, Pa.: University of Pennsylvania Press. 1976, pp. 232-247.

Howrey, E. Philip; Klein, Lawrence R.; and McCarthy, Michael D. "Notes on Testing the Predictive Performance of Econometric Models." In *Econometric Model Performance,* edited by Lawrence R. Klein and Edwin Burmeister. Philadelphia, Pa.: University of Pennsylvania Press, 1976, pp. 108-125.

Klein, Lawrence R. "The Supply Side." *American Economic Review,* March 1978, vol. 68, pp. 1-7.

Lucas, Jr., R. E. "Econometric Testing of the Natural Rate Hypothesis." In *The Econometrics of Price Determination,* Board of Governors of the Federal Reserve System, 1972, pp. 50-59.

McNees, Stephen K. "An Evaluation of Economic Forecasts." *New England Economic Review,* November/December, 1975, pp. 1-39.

———. "The 'Rationality' of Economic Forecasts." Unpublished Working Paper, Federal Reserve Bank of Boston, December 1977.

———. "An Accuracy Analysis of Macroeconomic Models of the U.S. Economy." Proceedings of the London Business School Conference on Economic Modelling, July 1978. New York: Heinemann Press. (forthcoming)

Nelson, Charles R. "The Prediction Performance of the FRB-MIT-PENN Model of the U.S. Economy." *American Economic Review,* December 1972. Vol. 62, pp. 902-917.

Okun, Arthur M. "Potential GNP: Its Measurement and Significance." *Proceedings of the American Statistical Association,* 1962.

Phelps, E.S. "Phillips Curves, Expectations of Inflation, and Optimal Unemployment Over Time." *Economica,* 1967, Vol. 34, pp. 254-281.

Sargent, Thomas J. "Rational Expectations, the Real Rate of Interest, and the Natural Rate of Unemployment." *Brookings Papers on Economic Activity,* 1973:2, pp. 429-472.