In search of a robust inflation forecast

Scott Brave and Jonas D. M. Fisher

Introduction and summary

The sound conduct of monetary policy is the bedrock on which a well-functioning economy rests. In the United States, the conduct of monetary policy is guided by the goals set out in the 1977 amendment to the Federal Reserve Act of 1913. According to this amendment, the Federal Reserve System and the Federal Open Market Committee (FOMC) should conduct monetary policy to promote the goals of “maximum” employment and output and to promote “stable” prices.

Of these goals, the primary focus, many economists believe, should be on achieving price stability. A stable price level means that prices of goods and services are undistorted by inflationary surprises. This enhances the role of prices in providing signals to ensure the efficient allocation of resources and the maximum possible sustainable level of employment. Many also believe that a stable price level encourages saving and capital accumulation, because it prevents asset values from being eroded by unanticipated inflation or debt being amplified by unanticipated deflation. This should also contribute to the goals of attaining maximum employment and output.

For these reasons, monetary policy is heavily influenced by factors thought to affect the rate of change of prices, that is, inflation. Until recently, the dominant concern had been a recurrence of past episodes of high inflation that have been associated with bad macroeconomic outcomes. In recent years, however, concern has shifted to the possibility of deflation. In either case, given the long lags over which policy actions can take effect, it is often necessary for the FOMC to take action before inflation starts to move in an undesired direction. The only way to do this with some confidence is to have effective ways of predicting the future course of inflation. Hence, forecasting inflation is a crucial ingredient in the formulation of monetary policy.

This article is concerned with the ability to forecast inflation. This is a relevant issue since recent work has cast doubt on the reliability of traditional approaches to forecasting inflation. Inflation forecasting is usually conducted with statistical models based on some version of the Phillips curve, the statistical relationship between inflation and overall aggregate economic activity. The recent literature suggests that this approach has not been reliable. In particular, Atkeson and Ohanian (2001) found that over the period 1985–99, one-year-ahead forecasts of inflation based on the Phillips curve do no better than a “naive” forecast where the forecast is set to the inflation rate over the prior year.

Some researchers have come to the defense of traditional forecasting models, arguing that the failure pointed out by Atkeson and Ohanian (2001) is special to the sample period they consider.[1] Still, it is difficult to dismiss their finding out of hand. As is clear from the work of Stock and Watson (1999, 2002, 2003), the forecasting failure in the post-1985 period reflects a more fundamental problem. While particular inflation forecasting models may do well in some periods, more often than not these models perform poorly at other times. It is not enough for a forecasting model to do well in just the recent period, because it is also important to guard against the possibility of structural change. Forecasters need to know that their forecasting strategy is robust to changes in the economic environment that are not noticed until well after they have occurred.

This article, therefore, addresses the question: Is it possible to build a robust inflation forecasting framework that does well in the recent period as well

Scott Brave is an associate economist and Jonas D. M. Fisher is an economic advisor at the Federal Reserve Bank of Chicago. *The authors thank Craig Furfine and Marcelo Veracierto for helpful comments.*

as earlier periods? We find that the answer to our question is "yes," although the gains compared with models based only on past inflation are at times quite modest. However, around periods in which inflation begins to pick up, the best models we consider show clear advantages over inflation-only models.

We address our question by considering the out-of-sample forecasting performance of a large set of models. We study forecast errors for the one-year and two-year forecasting horizons and at the monthly and quarterly frequencies. Our notion of robustness is that the model consistently lies near the top of performance lists of alternative models and is consistently more successful than models based only on past inflation, such as Atkeson and Ohanian's naive model.

Our main findings are as follows. First, consistent with previous studies, we show that different inflation indicators do well at forecasting inflation at different times. This makes the basic point that one should not rely on the "indicator du jour" when assessing the inflation outlook and that forecasters should be looking at many different indicators.

Second, we show that individual forecasting models that combine data in different ways do not consistently outperform the naive model (which turns out to be superior to other inflation-only models) in terms of mean-squared errors. For example, in some periods the naive model is better; at other times there is at least one model that does better than the naive model, but it is never the same one. This is true at both the one-year and the two-year horizon and with monthly and quarterly data. These findings are consistent with those reported by Fisher, Liu, and Zhou (2002).

Third, we show that certain kinds of models based on weighted averages of forecasts from individual models consistently outperform the naive model and other models based only on past inflation. This is true for both monthly and quarterly data and at both forecast horizons. At the one-year horizon, the best model involves weights computed using the within-sample forecasting performance of the individual models. At the two-year horizon, the best model uses a simple average of the individual models. For both forecasting horizons, the best versions of these models use a rolling window of data for the forecast, and these models are typically superior to the individual models for all sub-samples considered. These findings lead us to conclude that the most robust forecasts combine information from several different forecasting models, each of which incorporates the information in the available inflation indicators in different ways.

Another finding is that data available at the quarterly frequency that are not available at the monthly frequency appear to add little additional information to our forecasts. This might seem surprising, given that existing theoretical models suggest that data on real unit labor costs and productivity should be useful for predicting inflation, and these data are only available at the quarterly frequency. Still, we find that the additional data do not improve our forecasts very much, suggesting that most of the information about future inflation in the quarterly data is already incorporated in the monthly series we consider.

Below, we describe the different models we consider. Then, we discuss the methodology for assessing the forecasting performance of these models and present our findings.

Statistical models of inflation

In order to leave no stone unturned in our quest for a robust framework for forecasting inflation, we consider a large number of models. These models involve different ways of incorporating the vast amount of data available to the inflation forecaster. In principle, almost all the available macroeconomic data contain *some* information about future inflation. The challenge is to find a way to incorporate this information into a forecasting model. There are many ways to do this. One way would be to summarize the information useful for forecasting inflation *before* it is put into a model. Another approach would be to summarize the relevant information *after* it has been included in individual models. We employ each of these methods and also combine aspects of both. Finally, we combine the forecasts from several different types of models, each of which involves a different approach to forecasting. In the sub-sections that follow, we describe examples of each of these approaches. Many of these examples are motivated by the work of Stock and Watson (1999, 2002, 2003). For convenience we focus on the monthly frequency case. It should be clear how to extend the models to the quarterly frequency case. Table 1 summarizes the models underlying our analysis.

The basic regression equation

All the models we consider have as their foundation the basic regression equation:

$$1)\quad \pi^{12}_{t+J} - \pi^{12}_{t} = \alpha + \beta(L)(\pi_t - \pi_{t-1}) + \sum_{i=1}^{K} \theta_i(L)x_{it} + \varepsilon_{t+J},\quad J = 12, 24.$$

This equation relates changes in the 12-month inflation rate, defined as the 12-month change in the natural logarithm of the price index p_t,

$$\pi^{12}_t = \ln p_t - \ln p_{t-12},$$

TABLE 1		
Summary of models		
Model	**Estimation equation**	**Indicators used**
Naive	$\pi^{12}_{t+J} - \pi^{12}_t = \varepsilon_{t+J}$	None
Autoregression	$\pi^{12}_{t+J} - \pi^{12}_t = \alpha + \beta(L)(\pi_t - \pi_{t-1}) + \varepsilon_{t+J}$	None
Natural rate	$\pi^{12}_{t+J} - \pi^{12}_t = \alpha + \beta(L)(\pi_t - \pi_{t-1}) + \theta_1(L)x_{1t} + \varepsilon_{t+J}$	Filtered unemployment rate
Output gap	$\pi^{12}_{t+J} - \pi^{12}_t = \alpha + \beta(L)(\pi_t - \pi_{t-1}) + \theta_1(L)x_{1t} + \varepsilon_{t+J}$	Filtered real GDP
Activity	$\pi^{12}_{t+J} - \pi^{12}_t = \alpha + \beta(L)(\pi_t - \pi_{t-1}) + \theta_1(L)x_{1t} + \varepsilon_{t+J}$	Index based on indicators listed in appendix
Indicator	$\pi^{12}_{t+J} - \pi^{12}_t = \alpha + \beta(L)(\pi_t - \pi_{t-1}) + \sum_{i=1}^{3}\theta_i(L)x_{it} + \varepsilon_{t+J}$	Change in fed funds rate, unemployment rate, indicators listed in appendix
Combination	$\pi^{12}_{t+J} - \pi^{12}_t = \alpha + \beta(L)(\pi_t - \pi_{t-1}) + \theta_1(L)x_{1t} + \varepsilon_{t+J}$	Indicators listed in appendix
Diffusion	$\pi^{12}_{t+J} - \pi^{12}_t = \alpha + \beta(L)(\pi_t - \pi_{t-1}) + \sum_{i=1}^{K}\theta_i(L)x_{it} + \varepsilon_{t+J},\ K = 1, 2, \ldots, 6$	Six indexes based on indicators listed in appendix

Notes: See the text for a description of the notation and terminology. NA denotes not applicable; GDP denotes gross domestic product.

to past values of the one-month inflation rate, π_t,

$$\pi_t = \ln p_t - \ln_{pt-1},$$

and past values of other variables deemed useful for forecasting inflation, x_{it}, $i = 1, 2, \ldots, K$. In equation 1, α is a constant and $\beta(L)$ and $\theta_i(L)$, $i = 1, 2, \ldots, K$, specify the number of lags in inflation and other variables included in the equation. The number of other variables included is given by K, which is greater than or equal to zero.[2] We estimate equation 1 by ordinary least squares and use a standard lag selection criteria to choose the number of lags of inflation and other variables.[3] We allow for the possibility that lags could vary from one month to a year.

For given estimates of the coefficients in equation 1 at date $T, \hat{\alpha}_T, \hat{\beta}_T(L)$, and $\hat{\theta}_{iT}(L)$, the date T forecast of 12-month inflation J periods ahead using the basic regression equation is[4]

$$2)\quad \hat{\pi}^{12}_{T+J} = \pi^{12}_T + \hat{\alpha}_T + \hat{\beta}_T(L)(\pi_T - \pi_{T-1}) + \sum_{i=1}^{K}\hat{\theta}_{iT}(L)x_{iT},$$
$$J = 12, 24.$$

Models based only on inflation

We consider two models based only on inflation. The first is the "naive" model described by Atkeson and Ohanian (2001). The naive model can be viewed as a special case of equation 1, where $\alpha_T = \beta_T(L) = K = 0$. That is, the naive model equates the date T forecast of inflation over the next 12 months, $\hat{\pi}^{12}_{T+12}$, with its value over the most recent 12-month period,

$$3)\quad \hat{\pi}^{12}_{T+12} = \pi^{12}_T.$$

Similar to the 12-month forecast, the naive model equates the date T forecast of 12-month inflation 24 months into the future, $\hat{\pi}^{12}_{t+24}$, with its most recent value:

$$4)\quad \hat{\pi}^{12}_{t+24} = \pi^{12}_t.$$

The other model based only on inflation is called the *autoregression model*. This model postulates that changes in 12-month inflation only depend on recent changes in one-month inflation, that is, it sets $K = 0$ in equation 1.

Single equation models with inflation indicators

We consider three models that involve implementing equation 1 with $K = 1$. For the *natural rate model*, x_{1t} is set equal to the difference between a measure of the actual unemployment rate and an estimate of the "natural rate."[5] The *output-gap model*, is similar. In particular, x_{1t} is set equal to the difference between a measure of aggregate output and an estimate of "potential" output, where the latter is estimated using the same approach as with the natural rate.

For the *activity model*, x_{1t} is the Chicago Fed National Activity Index (CFNAI). This index is a weighted average of 85 monthly indicators of real economic activity. The CFNAI provides a single, summary measure of a common factor in these national economic data. As such, historical movements in the CFNAI closely track periods of economic expansion and contraction.[6]

Multiple equation models with inflation indicators

We also consider models that combine forecasts from applying versions of equation 1 with different indicator variables. The *diffusion model* can be viewed as a generalization of the activity model. We use a small number of indexes that explain the movements in 145 macroeconomic time series, including data measuring production, labor market status, the strength of the household sector, inventories, sales, orders, financial markets, money supply, and price data. The procedure that obtains the indexes processes the information in the 145 series, so that each index is a weighted average of the series and each index is statistically independent of the others. We consider six indexes computed in this way, $d_{1t}, d_{2t}, \ldots, d_{6t}$. These are listed in descending order in terms of the amount of information embedded in them.[7] The diffusion model involves first calculating an inflation forecast based upon including x_{1t} equal to the index with the most information, d_{1t}. We repeat this exercise five times, successively including one more index in descending order of importance. For instance, the third forecast created includes the three most important indexes, d_{1t}, d_{2t}, and d_{3t}, as x_{1t}, x_{2t}, and x_{3t}. The forecast from the diffusion model is the median of these six forecasts.[8]

Consider a list of forecasts of 12-month inflation J periods ahead at date T. Index these forecasts by n and denote them $f_{T+J}(n)$. The *combination model* is the median of these forecasts,

$$5)\quad \hat{\pi}^{12}_{T+J} = \text{median}\left\{ f_{T+J}(n) : n \in S \right\},$$

where the set of forecasts, S, is derived from the same 145 variables used to compute the diffusion indexes. In particular, each forecast $f_{T+J}(n)$ is based on equation 1 with $K = 1$ and x_{1t} set equal to one of the 145 variables used in the diffusion model.

The *indicator model* is based on a smaller list of variables grouped into six categories: economic activity, slackness measures, housing and building activity, industrial prices, financial markets, and, for the quarterly case only, productivity and marginal cost. Within each group, we compute a forecast using equation 1 with $K = 3$, x_{1t} set equal to the change in the federal funds interest rate, x_{2t} set equal to the unemployment rate, and x_{3t} to one of the variables in the group of indicators. We average the forecasts within each group. Then the indicator model forecast is based on equation 5 with $f_{T+J}(n)$ corresponding to one of the average forecasts from the five categories and S corresponding to the set of five average forecasts.

The combination and indicator models are useful to consider since they represent two alternatives to index-based methods for summarizing the information in many variables. The combination model is directly comparable to the diffusion model in that it involves the same set of variables. Therefore, it is useful to assess which method is superior for incorporating the information in a large number of variables. We work with the indicator model for two reasons. First, experience has shown it to be a relatively reliable approach to forecasting. Second, since it involves a small list of indicators, it represents a compromise between models that put a lot of weight on a single indicator, such as the natural rate and output gap models, and models that take virtually no stand on which indicators are useful, such as the diffusion and combination models.

Meta models

The preceding discussion introduced six models in addition to the inflation-only naive and autoregression models. To summarize, these models are the natural rate, output gap, activity, diffusion, combination, and indicator models. As we show below, none of these models consistently outperforms the inflation-only models over the various sub-samples we consider. However, for most of the sub-samples, at least one of the models does outperform the inflation-only models. This raises the question of whether it is possible to combine the information in these individual models to arrive at a superior forecast. The final group of models we study are designed to do just this. We call them *meta models*.[9]

Consider a list of forecasts of 12-month inflation J periods ahead at date T generated by the models listed above. Index these forecasts by n and denote them $f_{T+J}(n)$. The forecast of a given meta model is

$$6)\quad \hat{\pi}^{12}_{T+J} = \sum_{n \in M} w_{n,T}\, f_{T+J}(n),$$

where M is the set of models from which the meta model is constructed and $w_{n,T}$ is the weight attached to model n at date T. Equation 6 says that the forecast is set equal to a weighted average of the forecasts of the models comprising the meta model.

The meta models we consider differ according to the set of models from which the forecast is constructed and the manner in which the weights are computed. In the *equally weighted* models, the weights are all set equal to the inverse of the number of models comprising the model. That is, these forecasts are just the average over the forecasts of the individual models. The *optimally weighted* meta models have weights computed for each forecast date. These weights are computed as follows. At each forecast date, there is a prior history of forecasts and a history of actual inflation realizations corresponding to these forecasts. We reset the weights in equation 6 each forecast date to equal the coefficients of a regression of realized inflation on the forecasts using data on these variables available up to the date of the forecast.

Model evaluation methodology

We evaluate the accuracy of the models by comparing them with the naive and autoregression models. A modeling strategy will be deemed to be "robust" if it lies near the top of performance rankings and outperforms models based only on past inflation consistently across the various sub-samples we consider. We assess performance by *simulated out-of-sample forecasting*. This involves constructing inflation forecasts that a model would have produced had it been used historically to generate forecasts of inflation. We study forecasts of personal consumption deflator inflation, excluding food and energy, that is, core personal consumption deflator inflation.[10]

Two drawbacks of this approach are 1) we assume all the data are available up to the forecasting date, and 2) we do not use real-time data in our forecasts.[11] On a given date particular data series may not yet be published. Also many data series are revised after the initial release date. In our forecasting exercises, we compute forecasts and calculate the CFNAI and diffusion indexes assuming all the series underlying the forecasts and the indexes are available up to the forecast date. In practice this is never the case. Since we do not use real-time data, we also abstract from problems associated with data revisions. We suspect 1) and 2) lead us to overstate the effectiveness of our models.[12]

Root mean-squared error criterion

Our performance measure is the standard *root mean-squared error* (RMSE) criterion. The RMSE for any forecast is the square root of the mean squared differences between the actual inflation rate and the predicted inflation rate over the period for which simulated forecasts are constructed. For J = 12, 24

$$7)\quad RMSE = \left(\frac{1}{T-J} \sum_{t=1}^{T-J} \left[\pi^{12}_{t+J} - \hat{\pi}^{12}_{t+J} \right]^2 \right)^{1/2},$$

where $T - J$ denotes the number of forecasts made over the period under consideration.[13]

An advantage of the RMSE measure of performance is that its units are the same as inflation. This means, for example, the magnitude of RMSE for a given model can be directly compared with the average rate of inflation over the sample period. Another advantage is that large forecast errors are given more weight than small errors. Presumably, we care more about large mistakes than small mistakes. At the same time, a potential drawback of the RMSE measure is that it weights positive and negative errors of the same size in the same way. If we are more concerned about inflation increases than decreases, then this is definitely a drawback. Recent debates about the possible perils of deflation suggest that inflation decreases, at least at low levels of inflation, are certainly a concern of policymakers and so they should not be ignored. It would be interesting to consider other measures of forecast performance that weight increases and decreases in inflation differently, depending on the prevailing level of inflation.

Data and sample periods

The data we use in the analysis are described in the data appendix. The sample period of our analysis begins in 1967. We choose this date because it is the beginning date for the data used to construct the CFNAI and the diffusion indexes. We estimate the forecasting equations using all the data available at the time of the forecast and also consider the method of *rolling regressions*. A rolling regression keeps the number of observations in the regression constant across forecasts. Since it excludes observations from the distant past, this approach can in principle accommodate the possibility that there has been structural change in the data-generating

TABLE 2

Top five indicators, various sample periods: Combination and indicator variables

A. One-year ahead forecasts	B. Two-year ahead forecasts
1977–84	**1977–84**
ISM: Mfg: Prices Index	ISM Mfg: PMI Composite Index
Real inventories: Mfg: Durable goods industries	ISM: Mfg: Supplier Delivery Index
Housing starts: Northeast	ISM: Mfg: Inventories Index
ISM: Mfg: Inventories Index	ISM: Mfg: Employment Index
ISM: Mfg: Supplier Delivery Index	Housing starts: Midwest
1985–92	**1985–92**
Housing starts: Midwest	Housing starts: Midwest
NBER XLI2	Civilians unemployed for 15–26 weeks
Gold prices	Gold prices
Silver prices	Silver prices
CRB Futures Index	New home sales
1993–2000	**1993–2000**
Civilians unemployed for 5–14 weeks	Civilians unemployed for 5–14 weeks
Housing starts	Housing starts
3-year/1-year T-bill spread	Civilians unemployed for 15–26 weeks
10-Year Treasury note yield – federal funds rate	Housing starts: South
Civilians unemployed for 15–26 weeks	Building permits
2001–03	**2001–03**
Civilians unemployed for 27 weeks and over	Civilians unemployed for 5–14 weeks
Average duration of unemployment	Civilian unemployment rate: 16yr+
Civilians unemployed for 15 weeks and over	Employment retail and wholesale trade
Civilians unemployed for 5–14 weeks	Industrial Production Index
10-Year Treasury note yield – federal funds rate	Civilians unemployed for 15–26 weeks

process. To implement the rolling regression procedure, we choose a sample length of 15 years.

Finally, we consider four distinct periods over which to evaluate the forecasts of the models: 1977–84, 1985–92, 1993–2000, and 2001–2003. The first three periods are all 96 months long. We also consider the 1985–2003 period. The 1977–84 period is a period of high inflation volatility and general economic turbulence. The 1985–92 period is generally associated with a new monetary policy regime. This period also includes a mild recession. The 1993–2000 period witnessed uninterrupted economic expansion, stable monetary policy, and declining inflation. The 2001–2003 period is interesting because it involves recent forecast performance.

Findings

Next, we describe our findings. We focus on the monthly results and only discuss the findings with quarterly data at the end.

The best indicator keeps changing

Before evaluating our models, it is useful to consider the forecast performance of individual indicators. Each forecast is based on equation 1 with $K = 1$ and x_{1t} set equal to one of the list of indicators that includes the union of the set of variables used in the indicators model and the combination (or diffusion) model. Table 2 shows the top five indicators for the sample periods 1977–84, 1985–92, 1993–2000, and 2001–03. The key thing to notice from this table is that the list keeps changing! In the earliest sub-sample, indicators of manufacturing activity seem to do best at both the one-year and two-year horizons. At other times, employment, housing, or financial indicators do well. Overall, variables that do well at the one-year horizon do not necessarily do well at the two-year horizon. The lesson to be learned here is: beware of the indicator du jour.[14]

The best model keeps changing, too

Table 3 (p. 19) shows the performance of all the models (except for the output-gap model, which we only consider at the quarterly frequency) for the one-year and two-year forecast horizons, respectively. The meta models are in bold type. We discuss these models in the following sub-section. In table 3, we list the models for the four sub-samples as well as the period 1985–2003. We also display some useful summary statistics. For each sample period, we show the RMSE of the best model, the range of RMSE across forecasting models, the absolute value of the difference between the naive model and the best model, and average actual inflation.

The first thing to notice is that for both forecast horizons and across all sample periods the naive model performs better than the autoregression model. That is, there is no more information about future inflation in past inflation than that already contained in the most recent reading of 12-month inflation. This fact motivates our focus on using the naive model as a benchmark for comparison.

Now, consider the one-year ahead forecasts. In the earliest period, 1977–84, the natural rate model performed best. The magnitudes of the errors from this forecast are about one-sixth of the average inflation rate in this period. This is large relative to the amount by which this best model outperforms the naive model; the difference between the best model and the naive model is only about one-thirtieth of the average inflation rate in this period. So, even in this early period, the naive model is difficult to beat.

Since 1985, it has been even harder to beat the naive model. Indeed, over the entire 1985–2003 period the naive model is the best performer of the individual models. Consistent with the findings in Fisher, Liu, and Zhou (2002), the success of the naive model is concentrated in the 1985–92 period. In the latter part of the post-1985 sample, there is a model that beats the naive model, but this model changes and the extent of the victory is quite small. We should not attribute too much to the differences among the models for this forecast horizon; the range of root mean-squared errors is never that large and in the recent period is only about two-tenths of a percentage point.

The two-year ahead forecasts in table 3 present a similar picture. No individual model does well across all the sub-samples, although the diffusion model does perform reasonably well. The naive model does surprisingly well after 1985. Indeed, over the entire 1985–2003 period it is only one-tenth of a percentage point worse than the best individual model for this period, the diffusion model. The range of forecast errors is, as expected, a little larger for the two-year ahead forecasts, but still quite small.

Overall, table 3 indicates that no individual model consistently beats the naive model, and when one model does do better, the gains are small. We conclude that the natural rate, activity, diffusion, combination, and indicator models are not robust inflation forecasting frameworks.

Finally, it is interesting to note the relative performance of the combination, diffusion, and indicator models. Recall that these models involve using many indicators to forecast inflation, but do so in different ways. At the one-year horizon, there is little to choose between the models. Indeed the difference between the models is always less than one-tenth of a percentage point (not shown). At the two-year horizon, the diffusion model consistently outperforms the other two models except for the most recent period. Here the gains are more substantial (also not shown). For example, the diffusion model is superior to the indicator model by over 1 percentage point in the pre-1985 period and superior to the combination model by eight-tenths of a percentage point. In the post-1985 period the gains are about two-tenths and one-tenth of a percentage point, respectively.

The gains to combining forecasts

We now consider what happens when we combine the information in the forecasts from the various models. That is, we add to the list of models compared with the naive model the equally weighted and optimally weighted meta models. For good measure, we throw meta models based on rolling regressions into the mix. These are indicated in the table by the term “rolling.” The meta models are indicated by bold type in table 3. Since the optimally weighted models require a sample of forecasts to compute the weights, we only include these models in the mix after 1985. The meta models consist of the naive, natural rate, indicator, activity, diffusion, and combination models.

Notice that for both forecast horizons, the meta models generally outperform the individual models. Moreover, there is always a meta model that outperforms the naive model no matter which sub-sample we consider. Of special note is that it is possible to beat the naive model in the challenging 1985–92 period. Still, overall, the gains over the naive model are modest. Using the rolling regression approach provides some additional gain. At the one-year horizon, the regression strategy for computing weights seems to do better than just averaging the forecasts, but at the two-year horizon the opposite is true.

Is there evidence of a robust model here? Looking at the different sample periods and forecast horizons, it seems that the rolling optimally weighted model consistently outperforms the naive model and is near the top of the performance lists for the one-year horizons. The rolling equally weighted model is a very good performer at the two-year horizon. In both cases, when the model is not at the top of the performance list, it is within one-tenth of a percentage point of the top model and usually much less than that. The gains relative to the naive model are small in the 1985–92

TABLE 3

Monthly RMSE ranking, including meta and rolling models: One-year and two-year ahead forecasts

A. 1-year ahead forecasts	**1977–84**	**1985–92**	**1993–2000**	**2001–03**	**1985–2003**
	Natural rate	**Optimally weighted**	**Rolling optimally weighted**	**Rolling optimally weighted**	**Rolling optimally weighted**
	Equally weighted	**Rolling equally weighted**	**Rolling equally weighted**	**Optimally weighted**	**Rolling equally weighted**
	Rolling equally weighted	**Rolling optimally weighted**	**Optimally weighted**	Natural rate	**Optimally weighted**
	Naive	Naive	**Equally weighted**	**Rolling equally weighted**	**Equally weighted**
	Activity	**Equally weighted**	Diffusion	**Equally weighted**	Naive
	Indicators	Combination	Naive	Naive	Combination
	Combination	Autoregression	Activity	Combination	Autoregression
	Autoregression	Indicators	Natural rate	Autoregression	Diffusion
	Diffusion	Diffusion	Combination	Diffusion	Natural rate
	Natural	Rate	Autoregression	Indicators	Indicators
		Activity	Indicators	Activity	Activity
Summary statistics					
Best RMSE	1.03	0.50	0.33	0.38	0.42
Worst RMSE – Best RMSE	0.49	0.39	0.23	0.29	0.28
| Naive RMSE – Best RMSE |	0.20	0.02	0.11	0.12	0.06
Average inflation	6.48	3.84	1.87	1.57	2.65

B. 2-year ahead forecasts	**1977–84**	**1985–92**	**1993–2000**	**2001–03**	**1985–2003**
	Rolling equally weighted	**Rolling equally weighted**	**Optimally weighted**	**Equally weighted**	**Rolling equally weighted**
	Equally weighted	**Rolling optimally weighted**	**Rolling optimally weighted**	Natural rate	**Rolling optimally weighted**
	Diffusion	Naive	**Rolling equally weighted**	**Rolling optimally weighted**	**Optimally weighted**
	Naive	Diffusion	**Equally weighted**	**Optimally weighted**	Diffusion
	Natural rate	**Optimally weighted**	Diffusion	Combination	Naive
	Activity	**Equally weighted**	Activity	**Rolling equally weighted**	**Equally weighted**
	Combination	Combination	Naive	Naive	Combination
	Autoregression	Autoregression	Indicators	Autoregression	Autoregression
	Indicators	Indicators	Combination	Indicators	Indicators
		Activity	Natural rate	Diffusion	Activity
		Natural rate	Autoregression	Activity	Natural rate
Summary statistics					
Best RMSE	1.62	0.60	0.39	0.30	0.54
Worst RMSE – Best RMSE	1.32	0.87	0.45	0.47	0.57
| Naive RMSE – Best RMSE |	0.50	0.12	0.35	0.25	0.16
Average inflation	6.48	3.84	1.87	1.57	2.65

Notes: RMSE is root mean-squared error. Meta models are in bold above and include the following individual models: naive, activity, diffusion, combination, natural rate, and indicators.

period, but there are gains. Since 1993, the best meta-models beat the naive model by about one-tenth of a percentage point at the one-year horizon and two-and-a-half-tenths at the two-year horizon. This latter advantage is not insubstantial given that inflation over this period is on average less than 2 percent.

The robust models

Since 1985, the most robust models seem to be the rolling equally weighted and rolling optimally weighted models. It is instructive to study these models a little more.

Cumulative forecast errors

Figures 1 and 2 display cumulative squared forecast errors for the rolling optimally weighted model and the naive model for the one-year and two-year horizons. Figures 3 and 4 (p. 22) are similar, but with the rolling equally weighted and naive models. The vertical lines in these figures indicate the boundaries of the sample periods we consider. To interpret these figures, note that differences in performance are indicated by differences in the slopes of the lines. The model with the flatter line is performing better than the other model over the particular period in which the line is flatter. When one line is below another at a particular date, the model associated with that line has performed better in an RMSE sense up to that date. Note that, due to the need to have data to compute the weights, the figures for the rolling optimally weighted model begin in 1985.

Consider the rolling optimally weighted model first. For the one-year horizon there is little to choose between this model and the naive model in the 1985–92 period. Differences emerge after 1993, but these are concentrated in 1994 and 1995. Additional gains relative to the naive model appear in 2003, though. For the two-year horizon the differences are more substantial, but the overall impression is similar. The location of when the largest gains appear is interesting, since these correspond to periods in which inflation was increasing.

The figures for the rolling equally weighted model present a similar picture for the post-1985 period. The pre-1985 observations are particularly interesting. These illustrate the fact that most of the gains relative to the naive model are in the period before 1985. We can see this in the distance between the two lines in the figures, which does not get much wider after 1985.

Model weights

Figures 5 and 6 (pp. 23–24) display the evolution of the weights underlying the rolling optimally weighted model for the one-year and two-year horizons, respectively. Recall that these weights are based on regressing actual inflation on forecasts from six models, the naive, activity, natural rate, indicator, combination, and diffusion models. The individual models are estimated using rolling regressions, but the weights are based on forecasts for the entire available sample.

Figure 5 shows that for much of the sample all the models get a non-trivial weight for the one-year horizon. Except for the early part of the sample, the weights have not changed that much. Still, their time paths provide some interesting insight into the evolution of the economy. For example, the natural rate model has declined in importance over the sample. Nonetheless, it still gets a large weight. The weight on the naive model has grown over the sample. The activity, diffusion, and combination models get negative weights.[15] Figure 6 indicates that forecasting the two-year horizon involves using the models differently. The natural rate model gets much less weight, and for much of the sample the activity and indicator models get very small weights. Consistent with their individual performances (see table 3), the naive and diffusion models get large weights.

Quarterly data

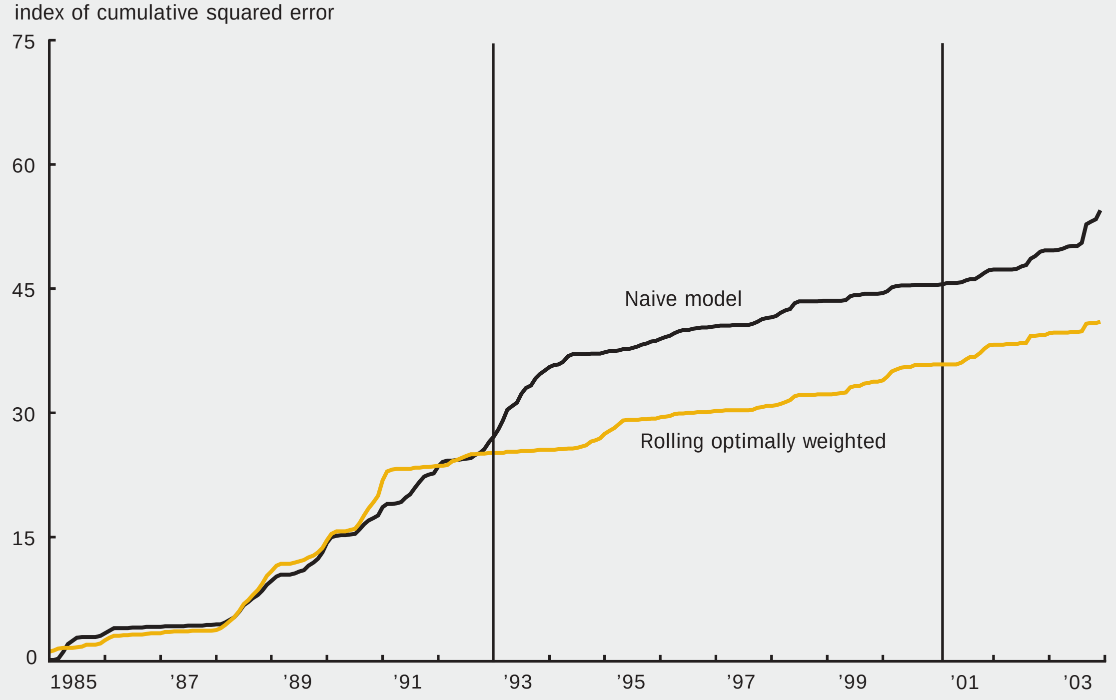
Now, we briefly summarize our findings with quarterly data. To conserve space we do not display our findings. Our purpose here is twofold. We want to know whether averaging the forecasts obtained by different forecasting procedures also improves forecasts at the quarterly frequency. We also want to understand whether adding quarterly data to the analysis that are not available at the monthly frequency improves the quality of the forecasts. The new data include data from the *National Income and Product Accounts*, the output gap, and data on productivity and costs (see the appendix for a list of the specific series).

Regarding the first question, we find that the basic principle of averaging different forecasts also yields forecasting benefits at the quarterly frequency. Indeed the same meta models that show promise at the monthly frequency are also among the most robust at the quarterly frequency when we include the additional quarterly data.[16] With one exception, these models improve on the naive forecast over all sub-samples and both forecast horizons we consider. The exception is in the 1985–92 period for the one-year horizon, in which no model is superior to the naive model.

Incorporating the additional data leads to mixed results. We use the third month in each quarter to compare a given monthly model with its quarterly counterpart. When we do this and compare corresponding monthly and quarterly models, we find little evidence

FIGURE 1

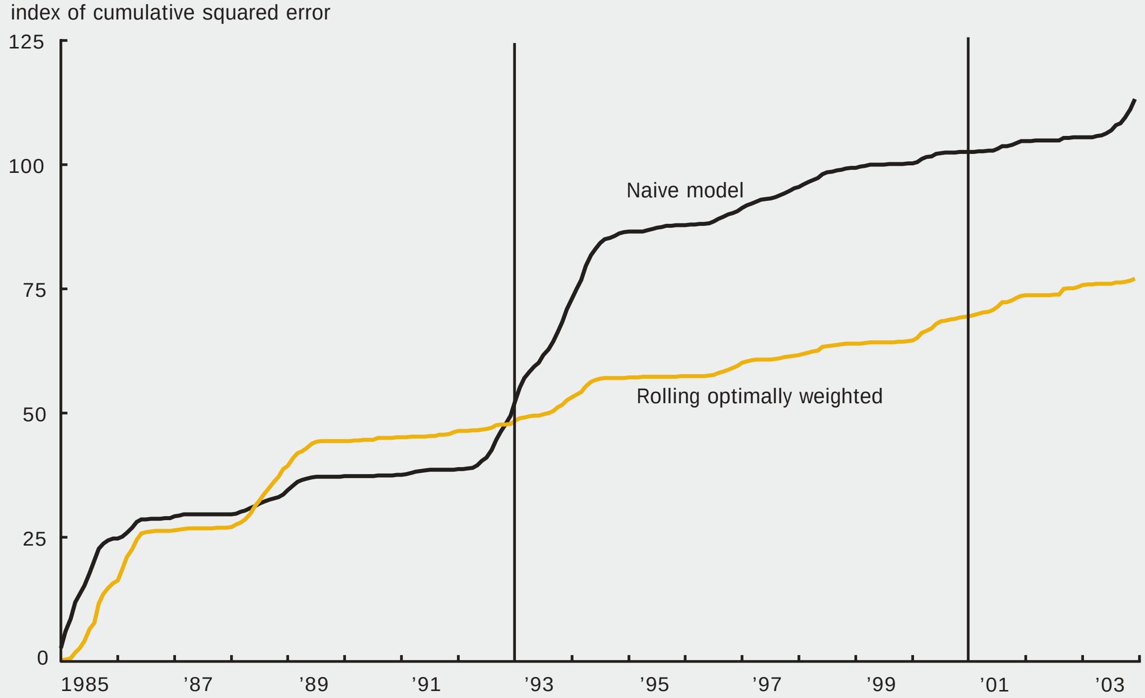
Cumulative squared errors at the 1-year horizon: Naive and rolling optimally weighted models



Note: The vertical lines indicate the bounds of the sample period being considered.

FIGURE 2

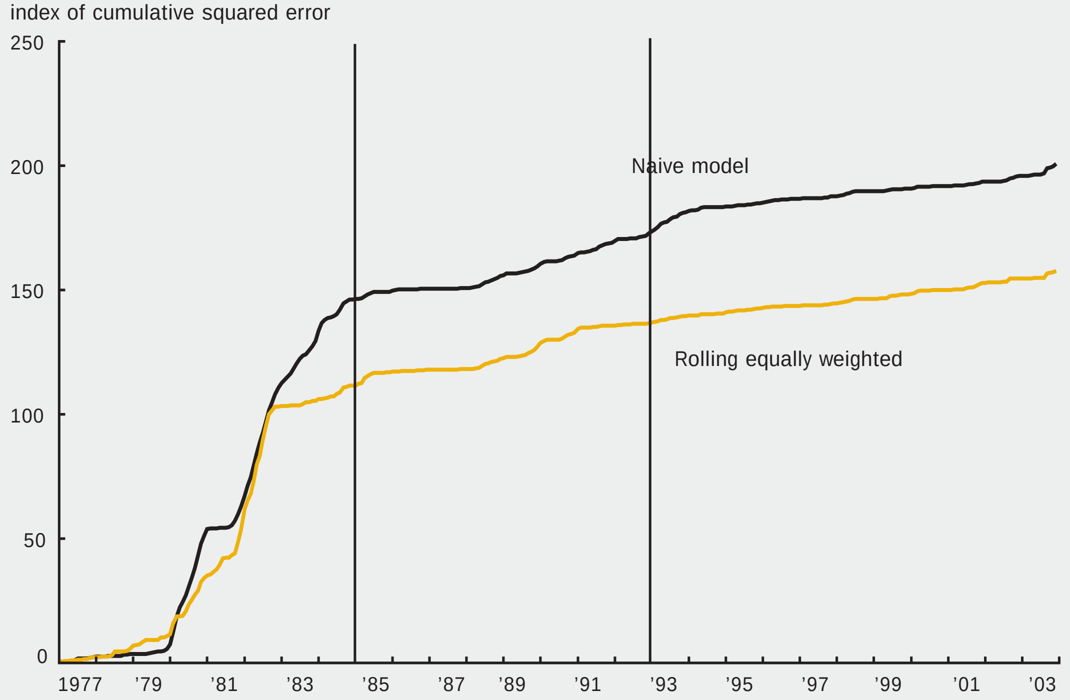
Cumulative squared errors at the 2-year horizon: Naive and rolling optimally weighted models



Note: The vertical lines indicate the bounds of the sample period being considered.

FIGURE 3

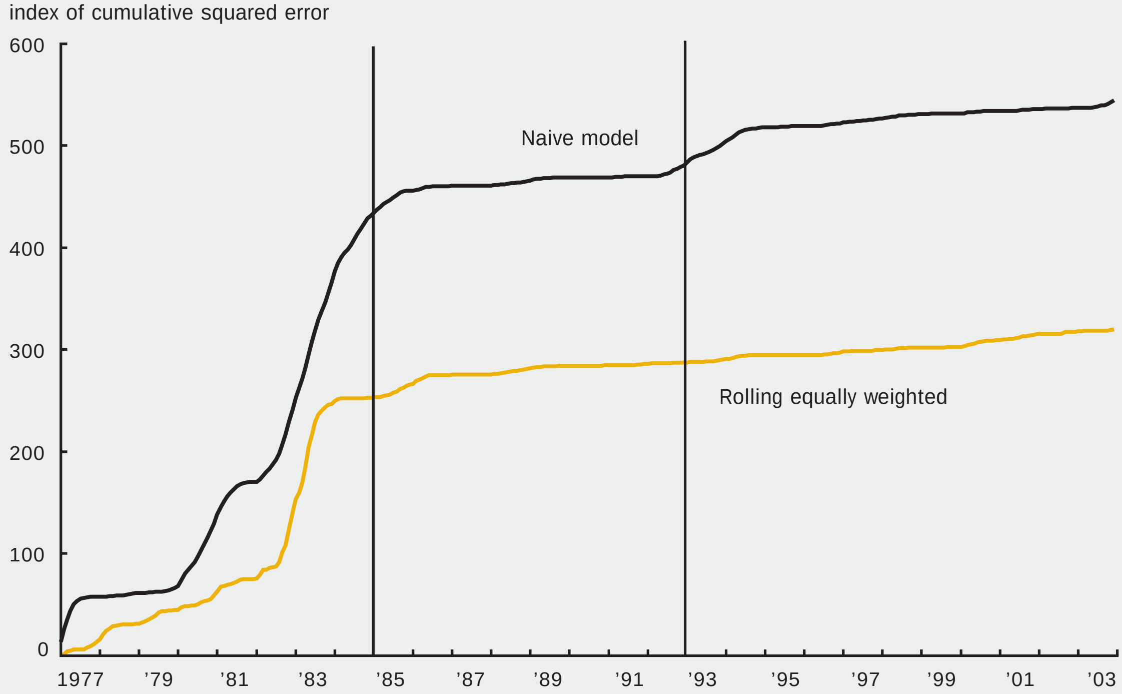
Cumulative squared errors at the 1-year horizon: Naive and rolling equally weighted models



Note: The vertical lines indicate the bounds of the sample period being considered.

FIGURE 4

Cumulative squared errors at the 2-year horizon: Naive and rolling equally weighted models



Note: The vertical lines indicate the bounds of the sample period being considered.

FIGURE 5

Regression weights for rolling forecasts, 1-year horizon

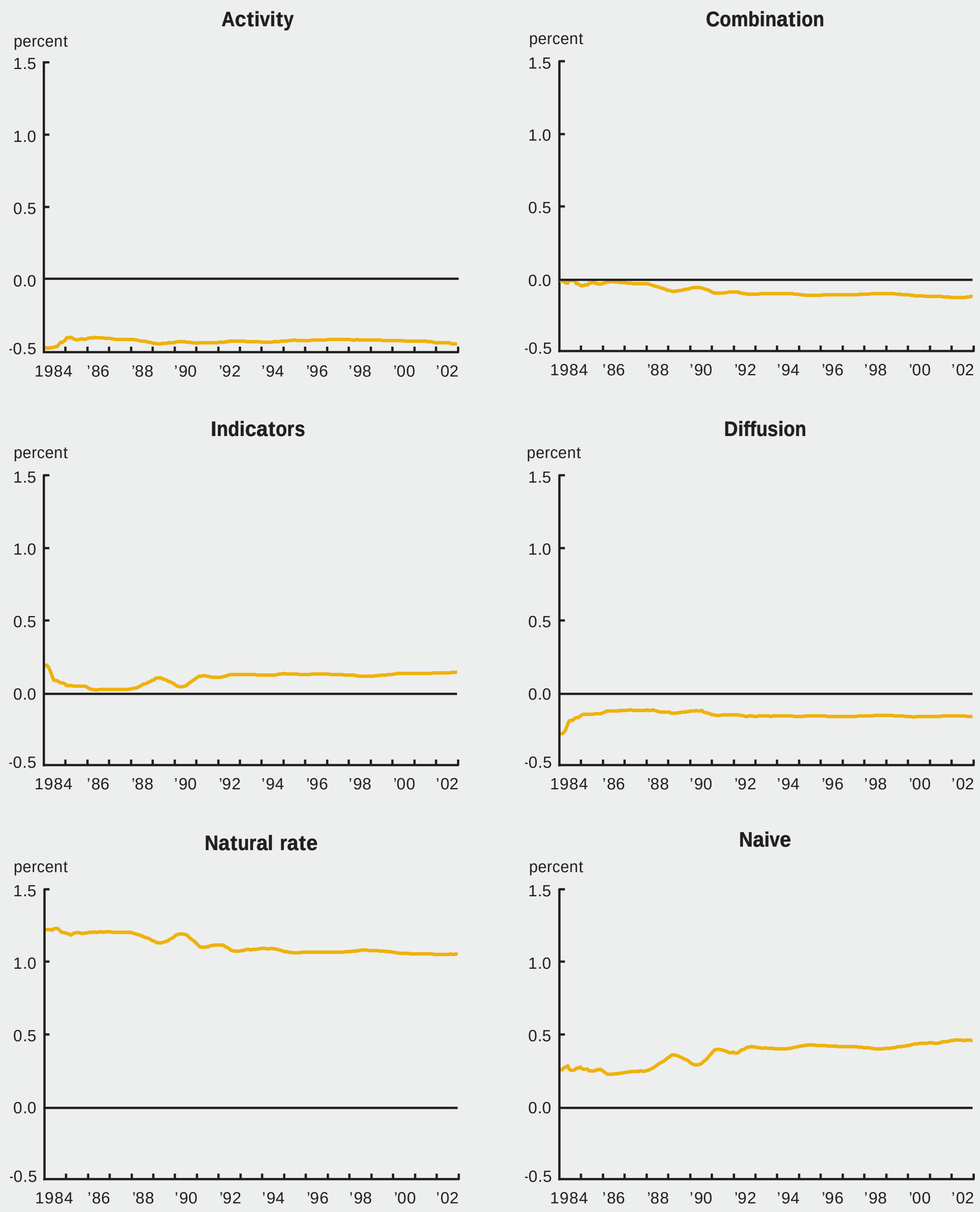
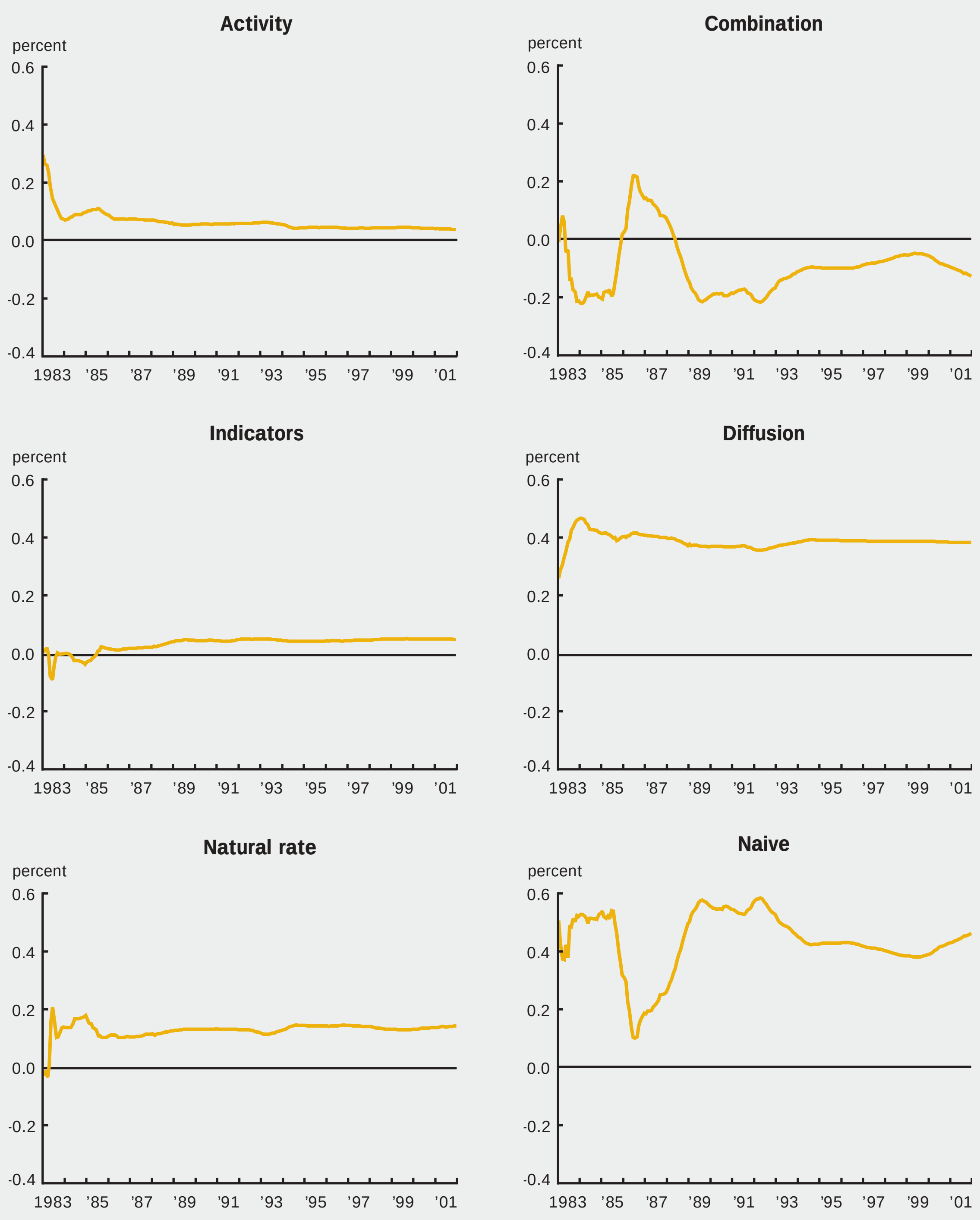


FIGURE 6

Regression weights for rolling forecasts, 2-year horizon



that the additional data improve the forecasts. In particular, there is not a consistent pattern of improvement with the quarterly models and when there is improvement it is typically much less than one-tenth of a percentage point. Sometimes the quarterly models are worse. One model does show consistent improvement at the quarterly frequency—the rolling optimally weighted model. This model does well at the two-year horizon, improving over its monthly counterpart by about one-tenth of a percentage point in all sub-samples after 1985.[17]

In a departure from the monthly analysis, a non-meta model shows up in the list of robust models when we incorporate the additional data. This model is the rolling output gap model, which we could not examine at the monthly frequency because gross domestic product data are only available quarterly. When the output gap model is estimated using the rolling procedure, it is the best performing model over 1977–84 and 1985–2003 and performs better than the naive model in all the sub-samples we consider when forecasting two years ahead. This model does not do as well forecasting at the one-year horizon. In particular, it is outperformed by the rolling optimally weighted model over all the sub-samples. Still, the fact that such a simple model does so well at forecasting two years ahead is interesting and deserves further study.[19]

Taking all the evidence into account, it seems reasonable to conclude that the quarterly data do not add much to forecast performance. Two exceptions are when the additional data are incorporated into the rolling output gap model and the rolling optimally weighted model, both of which perform well at the two-year horizon.

Conclusion

We have found that a robust forecast of the magnitude of inflation can be obtained by combining the forecasts of several models that incorporate the information in the available data in different ways. This suggests that a useful approach to building a reliable statistical forecasting framework is to be eclectic with respect to both the data used to formulate a forecast and the models used to incorporate the data into a forecast. Relying on a small number of inflation indicators and one forecasting model is not a good idea.

Having drawn this conclusion, we must note two caveats.[18] The most obvious caveat is that the conclusion we have just stated sows the seeds of future failure. We have concluded that one must not rely on a particular model, yet we have essentially described a particular model. While we realize the circularity of our conclusion, we would rather interpret our findings as suggesting that combining the forecasts from models that include the data in different ways is the main lesson to be learned. That is, we do not put a lot of weight on the particular models we worked with. We also want to emphasize the limitations of the kinds of forecasting models studied in this article. Clearly, these models are not structural and, therefore, are inadequate for assessing the impact of systematic changes in policy. This is what fully articulated general equilibrium economic models, which account for behavioral responses to policy changes, are for. However, such models, while beginning to be used at central banks, are still inadequate for the everyday needs of policymakers. The forecasting models discussed here have their uses and probably will continue to be popular for some time to come. Principally, these models are useful for understanding what current inflation expectations are. Since the past actions of the Fed are embedded in the coefficients, the models take into account "typical" Fed responses to current conditions. For these reasons, inflation forecasts serve as a useful benchmark for policymakers assessing the current stance of monetary policy. This article has shown that such forecasts can be improved reliably by taking into account information in variables other than inflation.

NOTES

[1]See, for example, Sims (2002) and Stock and Watson (2002). Fisher, Liu, and Zhou (2002) document that the failure of Phillips curve models after 1985 is essentially due to an especially poor performance in the 1985–92 period.

[2]One might view equation 1 as an odd choice to base inflation forecasts on since it involves *changes* of inflation rather than *levels* of inflation. The reason we use this equation is because it performs better than models based on the level of inflation. This reflects the fact that 12-month inflation is an extremely persistent variable, so that its level does not change much over short periods.

[3]Specifically, we use the Bayes information criterion (BIC) to select the number of lags. Intuitively, BIC selects the number of lags to improve the fit of the model without increasing by too much the sampling error in the lag coefficients.

[4]Another way to forecast inflation would be to formulate a vector autoregression in the level or change in one-month inflation and the indicator variables and project this system forward J periods from date T. Such a forecast would yield superior results if the vector autoregression were correctly specified. The conventional wisdom is that the direct approach taken here is in practice better. Marcellino, Stock, and Watson (2004) show that for many variables, but not for inflation, this conventional wisdom is apparently false. We have explored the "multi-step iterated forecasts" described in Marcellino, Stock, and Watson (2004) and concur with their finding that this approach is a poor forecasting strategy for inflation.

[5]To estimate the natural rate, we use a filter applied to the time series of unemployment available at the time of the forecast. The particular filter we use is called a *band-pass filter*. This is designed to isolate particular frequencies of the data. We use it to isolate "long-run" or low frequency fluctuations in the unemployment rate. Specifically, we focus on fluctuations of period (inversely related to the frequency) 12 years or greater. The particular implementation of the band-pass filter we use is the one due to Christiano and Fitzgerald (1999).

[6]The index methodology was proposed by Stock and Watson (1999, 2002). For more details on the CFNAI, see www.chicagofed.org/economic_research_and_data/cfnai.cfm.

[7]Technically, we compute the first six principal components of the 145 variables.

[8]The median of six forecasts is the average of the third and fourth ranked forecasts. We explored other ways of choosing among the six models, including using the mean and using the best out-of-sample forecasting performance (this is described later) up to the date of the forecast. These other ways of summarizing the forecasts performed similarly to the approach taken here.

[9]The word "meta" is often used to describe an analysis that synthesizes research results obtained using different approaches to a question. By this definition, the diffusion, combination, and indicator models might also be considered meta models. We prefer not to use this descriptor to classify these models since they combine the information from forecasts that, except for the indicators used, are based on the same forecasting strategy.

[10]We use this measure of inflation since it plays a prominent role in FOMC discussions.

[11]Compiling the data that were available at a particular point in time is a daunting task. A real-time dataset is available from the Philadelphia Fed. Unfortunately this dataset has a limited number of variables and excludes many that might be useful for forecasting inflation.

[12]Data revisions are a problem for the naive and autoregression models since the price index we use, the PCE deflator, is subject to revisions.

[13]Comparisons of models based on RMSE are subject to sampling variability and consequently subject to error. In principle, we could use Monte Carlo methods to assess the magnitude of this error. However, this would require specifying an underlying data-generating process for all the variables in our analysis (more than 150 of them). This sampling error should be kept in mind when interpreting the results. See Clark and McCracken (2001) and the references they cite for a useful discussion of some of the issues involved in assessing the statistical difference in the accuracy of forecasts.

[14]For another discussion of this point, see Cecchetti, Chu, and Steindel (2000).

[15]In principle there is nothing wrong with a negative weight. Conditional on all the other forecasts, a forecast of an increase in inflation from a model with a negative weight is a signal that the other models combined are forecasting an increase in inflation that is too big or a decrease in inflation that is not big enough, relative to past experience. If the model did not provide information about inflation, then it would get a zero weight.

[16]When computing the weighted forecasts at the quarterly horizon, we add the forecasts of the output gap model to the list of forecasts that are averaged.

[17]We also examine the impact of just averaging the monthly data to convert it to the quarterly frequency. When we do this, we find little evidence that monthly noise is a significant source of forecast error since there is not a consistent pattern of improvement in the quarterly models and when there is improvement it is typically much less than one-tenth of a percentage point.

[18]Another important caveat involves the use of rolling regressions. Sargent (1999) argues that the rise of inflation during the 1960s and 1970s and the subsequent decline can be explained by a process of the Fed learning and forgetting about its ability to exploit a perceived trade-off between inflation and unemployment. This analysis suggests a potential problem with using the rolling regression framework, because it may lead to a recurrence of the rise of inflation in the 1960s and 1970s. However, as Sargent (1999, p. 134) points out, a credible commitment by the Fed to low inflation should prevent such a recurrence. Under this view, there is no problem with using the rolling regression approach to forecasting.

[19]See Clark and McCracken (2004) for a recent analysis of the predictive content of the output gap for inflation.

DATA APPENDIX

Monthly data: 1967:01–2003:12[a]

Model	Transformation	Mnemonic	Constructed series mnemonic	Haver description	Haver database	Secondary source
Activity	log 1st diff	le		Civilian employment: Sixteen years & over: 16 yr + (SA, 000s)	usecon	
Activity	1st diff	lrm25		Civilian unemployment rate: Men, 25–54 years (SA, %)	usecon	
Activity	1st diff		LCUN = a0m005	Average weekly initial claims unemployment insurance (SA, 000s)	bci	
Activity, diffusion, combination	log 1st diff	cbhm		Personal consumption expenditures (SAAR, chained 2000$bil.)	usecon	
Activity, diffusion, combination	log 1st diff	cdbhm		Personal consumption expenditures: Durable goods (SAAR, chained 2000$bil.)	usecon	
Activity, diffusion, combination	log 1st diff	cnbhm		Personal consumption expenditures: Nondurable goods (SAAR, chained 2000$bil.)	usecon	
Activity, diffusion, combination	log 1st diff	csbhm		Personal consumption expenditures: Services (SAAR, chained 2000$bil.)	usecon	
Activity, diffusion, combination	log 1st diff	ypdhm		Real disposable personal income (SAAR, chained 2000$bil.)	usecon	
Activity, diffusion, combination	log 1st diff		CONSTPV = cpv –cpvr	Value of public construction put in place (SAAR, chained $mil.)	usecon	
Activity, diffusion, combination	log 1st diff		CONSTPU = cpg	Value of private construction put in place (SAAR, chained $mil.)	usecon	
Activity, diffusion, combination	log	hsm		Manufacturers' shipments of mobile homes (SAAR, units in 000s)	usecon	
Activity, diffusion, combination	log	hst		Housing starts (SAAR, units in 000s)	usecon	
Activity, diffusion, combination	log	hstmw		Housing starts: Midwest (SAAR, units in 000s)	usecon	
Activity, diffusion, combination	log	hstne		Housing starts: Northeast (SAAR, units in 000s)	usecon	
Activity, diffusion, combination	log	hsts		Housing starts: South (SAAR, units in 000s)	usecon	
Activity, diffusion, combination	log	hstw		Housing starts: West (SAAR, units in 000s)	usecon	
Activity, diffusion, combination	log 1st diff	ip		Industrial Production Index (SA, 1997=100)	usecon	
Activity, diffusion, combination	log 1st diff	ip51		Industrial Production: Consumer goods (SA, 1997=100)	usecon	
Activity, diffusion, combination	log 1st diff	ip511		Industrial Production: Durable consumer goods (SA, 1997=100)	usecon	
Activity, diffusion, combination	log 1st diff	ip512		Industrial Production: Nondurable consumer goods (SA, 1997=100)	usecon	
Activity, diffusion, combination	log 1st diff	ip521		Industrial Production: Business equipment (SA, 1997=100)	usecon	
Activity, diffusion, combination	log 1st diff	ip53		Industrial Production: Materials (SA, 1997=100)	usecon	
Activity, diffusion, combination	log 1st diff	ip531		Industrial Production: Durable goods materials (SA, 1997=100)	usecon	
Activity, diffusion, combination	log 1st diff	ip532		Industrial Production: Nondurable goods materials (SA, 1997=100)	usecon	
Activity, diffusion, combination	log 1st diff	ip54		Industrial Production: Nonindustrial supplies (SA, 1997=100)	usecon	
Activity, diffusion, combination	log 1st diff	ipb0		Industrial Production: Mining (SA, 1997=100)	usecon	
Activity, diffusion, combination	log 1st diff	ipfp		Industrial Production: Final products (SA, 1997=100)	usecon	
Activity, diffusion, combination	log 1st diff	ipmdg		Industrial Production: Durable goods [NAICS] (SA, 1997=100)	usecon	
Activity, diffusion, combination	log 1st diff	ipmfg		Industrial Production: Manufacturing [SIC] (SA, 1997=100)	usecon	
Activity, diffusion, combination	log 1st diff	ipmnd		Industrial Production: Nondurable manufacturing (SA, 1997=100)	usecon	
Activity, diffusion, combination	log 1st diff	iptp		Industrial Production: Final products and nonindustrial supplies (SA, 1997=100)	usecon	
Activity, diffusion, combination	log 1st diff	iputl		Industrial Production: Electric and gas utilities (SA, 1997=100)	usecon	
Activity, diffusion, combination	log 1st diff	laconsa		All employees: Construction (SA, 000s)	usecon	
Activity, diffusion, combination	log 1st diff	ladurga		All employees: Durable goods manufacturing (SA, 000s)	usecon	
Activity, diffusion, combination	log 1st diff	lafirea		All employees: Financial activities (SA, 000s)	usecon	
Activity, diffusion, combination	log 1st diff	lagooda		All employees: Goods-producing industries (SA, 000s)	usecon	
Activity, diffusion, combination	log 1st diff	lagovta		All employees: Government (SA, 000s)	usecon	
Activity, diffusion, combination	log 1st diff	lamanua		All employees: Manufacturing (SA, 000s)	usecon	
Activity, diffusion, combination	log 1st diff	laminga		All employees: Mining (SA, 000s)	usecon	
Activity, diffusion, combination	log 1st diff	lanagra		All employees: Total nonfarm (SA, 000s)	usecon	
Activity, diffusion, combination	log 1st diff	landura		All employees: Nondurable goods manufacturing (SA, 000s)	usecon	
Activity, diffusion, combination	log 1st diff	lapriva		All employees: Total private industries (SA, 000s)	usecon	
Activity, diffusion, combination	log 1st diff	lartrda		All employees: Retail trade (SA, 000s)	usecon	
Activity, diffusion, combination	log 1st diff	laserpa		All employees: Service-providing industries (SA, 000s)	usecon	
Activity, diffusion, combination	log 1st diff	LASRVSA = lainfoa + lapbsva + laeduha + laleiha + lasrvoa		All employees: Aggregate of categories	usecon	
Activity, diffusion, combination	log 1st diff	LATPUTA = lattula – lawtrda – lartrda		All employees: Aggregate of categories	usecon	
Activity, diffusion, combination	log 1st diff	lena		Civilian employment: Nonagricultural Industries: 16yr + (SA, 000s)	usecon	
Activity, diffusion, combination	log 1st diff	lhelpr		Ratio: Help-wanted advertising in newspapers/Number unemployed (SA)	usecon	
Activity, diffusion, combination	1st diff	lomanua		Average weekly hours: Overtime: Manufacturing (SA, Hrs)	usecon	
Activity, diffusion, combination	1st diff	lrmanua		Average weekly hours: Manufacturing (SA, Hrs)	usecon	
Activity, diffusion, combination	level	napmc		ISM Mfg: PMI Composite Index (SA, 50+ = Econ Expand)	usecon	
Activity, diffusion, combination	level	napmei		ISM Mfg: Employment Index (SA, 50+ = Econ Expand)	usecon	
Activity, diffusion, combination	level	napmii		ISM Mfg: Inventories Index (SA, 50+ = Econ Expand)	usecon	
Activity, diffusion, combination	level	napmni		ISM Mfg: New Orders Index (SA, 50+ = Econ Expand)	usecon	
Activity, diffusion, combination	level	napmoi		ISM Mfg: Production Index (SA, 50+ = Econ Expand)	usecon	
Activity, diffusion, combination	log 1st diff	rsdh		Real retail sales: Durable goods (SA, chained 2000$mil.)	usecon	
Activity, diffusion, combination	log 1st diff		RSH = rsh + rsh2	Retail sales: Retail trade (SA, Spliced, chained 2000$mil.)	usecon	
Activity, diffusion, combination	log 1st diff	rsnh		Real retail sales: Nondurable goods (SA, chained 2000$mil.)	usecon	
Activity, diffusion, combination	log 1st diff		TIMDH = timdh + timdh2	Real inventories: Mfg: Durable goods industries (SA, EOP, spliced, chained 2000$mil.)	usecon	
Activity, diffusion, combination	log 1st diff		TIMH = timh + timh2	Real manufacturing & trade inventories: Mfg industries (SA, EOP, spliced, chained 2000$mil.)	usecon	
Activity, diffusion, combination	log 1st diff		TIMNH = timnh + timnh2	Real mfg inventories: Nondurable goods industries (SA, EOP, spliced, chained 2000$mil.)	usecon	
Activity, diffusion, combination	log 1st diff		TIRH = tirh + tirh2	Real inventories: Retail trade industries (SA, EOP, spliced, chained 2000$mil.)	usecon	
Activity, diffusion, combination	log 1st diff		TITH = tith + tith2	Real manufacturing & trade inventories: Industries (SA, EOP, spliced, chained 2000$mil.)	usecon	
Activity, diffusion, combination	log 1st diff		TIWH = tiwh + tiwh2	Real inventories: Merchant wholesale trade industries (SA, EOP, spliced, chained 2000$mil.)	usecon	

DATA APPENDIX *(continued)*

Model	Transformation	Mnemonic	Constructed series mnemonic	Haver description	Haver database	Secondary source
Activity, diffusion, combination	1st diff		TRMH = trmh + trmh2	Real inventories/sales ratio: Manufacturing industries (SA, spliced, chained 2000$)	usecon	
Activity, diffusion, combination	1st diff		TRRH= trrh + trrh2	Inventories/sales ratio: Retail trade industries (SA, spliced, chained 2000$)	usna	
Activity, diffusion, combination	1st diff		TRTH= trth + trth2	Real manufacturing & trade: Inventories/sales ratio (SA, spliced, chained 2000$)	usna	
Activity, diffusion, combination	1st diff		TRWMH=trwmh + trwmh2	Inventories/sales ratio: Merchant wholesale trade industries(SA, spliced, chained 2000$)	usna	
Activity, diffusion, combination	log 1st diff		TSMDH= tsmdh + tsmdh2	Real sales: Mfg: Durable goods industries(SA, spliced, chained 2000$mil.)	usna	
Activity, diffusion, combination	log 1st diff		TSMH= tsmh + tsmh2	Real sales: Manufacturing industries (SA, spliced, chained 2000$mil.)	usna	
Activity, diffusion, combination	log 1st diff		TSMNH= tsmnh + tsmnh2	Real sales: Mfg: Nondurable goods industries (SA, spliced, chained 2000$mil.)	usna	
Activity, diffusion, combination	log 1st diff		TSTH= tsth + tsth2	Real manufacturing & trade sales: All industries (SA, spliced, chained 2000$mil.)	usna	
Activity, diffusion, combination	log 1st diff		TSWMDH= tswmdh	Real sales: Merchant wholesalers: Durable goods inds. (SA, spliced, chained 2000$mil.)	usna	
Activity, diffusion, combination	log 1st diff		TSWMH= tswmh + twsmh2	Real sales: Merchant wholesale trade industries (SA, spliced, chained 2000$mil.)	usna	
Activity, diffusion, combination	log 1st diff		TSWMNH= tswmnh + tswmnh2	Real sales: merchant wholesale: Nondurable goods inds. (SA, spliced, chained 2000$mil.)	usna	
Activity, diffusion, combination	log 1st diff	ypltpmh		Real personal income less transfer payments (SAAR, chained 2000$bil.)	usecon	
Activity, diffusion, combination	log 1st diff		CDVHM = cdvhm + cdvh	PCE: Durable goods: Motor vehicles and parts (SAAR, spliced and interpolated, chained 2000$mil.)	usna	
Activity, diffusion, combination	log 1st diff		MDOQ = a0m007	Manufacturers' new orders: Durable goods (SA, chained 2000$mil.)	bci	
Activity, diffusion, combination	log 1st diff		MOCGMC = a0m008	Manufacturers' new orders: Consumer goods & materials (SA, 1982$mil.)	bci	
Activity, diffusion, combination	log 1st diff		MOCNC = a0m027	Manufacturers' new orders: Nondefense capital goods (SA, 1982$mil.)	bci	
Activity, diffusion, combination, indicators (3)	log	hpt		New private housing units authorized by building permit (SAAR, units in 000s)	usecon	
Activity, diffusion, combination, indicators (2)	1st diff	cumfg		Capacity utilization: Manufacturing [SIC] (SA, % of capacity)	usecon	
Activity, diffusion, combination, indicators (2)	log 1st diff	lhelp		Index of help-wanted advertising in newspapers (SA, 1987=100)	usecon	
Activity, diffusion, combination, indicators	1st diff	lr		Civilian unemployment rate: 16yr + (SA, %)	usecon	
Activity, indicators (2)	level	napmvdi		ISM: Mfg: Vendor Deliveries Index (SA, 50+ = Econ Expand)	usecon	
Diffusion, combination	level	cexp		University of Michigan: Consumer expectations (NSA, 66Q1=100)	usecon	
Diffusion, combination	level	lu0		Civilians unemployed for less than 5 weeks (SA, 000s)	usecon	
Diffusion, combination	level	lu15		Civilians unemployed for 15–26 weeks (SA, 000s)	usecon	
Diffusion, combination	level	lu5		Civilians unemployed for 5–14 weeks (SA, 000s)	usecon	
Diffusion, combination	level	luad		Average {Mean} duration of unemployment (SA, weeks)	usecon	
Diffusion, combination	level	lut15		Civilians unemployed for 15 weeks and over (SA, 000s)	usecon	
Diffusion, combination	level	lut27		Civilians unemployed for 27 weeks and over (SA, 000s)	usecon	
Diffusion, combination	log 2nd diff	faram		Adjusted monetary base (SA, $mil.)	usecon	
Diffusion, combination	log 2nd diff	faran		Adjusted nonborrowed reserves of depository institutions (SA, $mil.)	usecon	
Diffusion, combination	log 2nd diff	faranp		Adjusted nonborrowed reserves plus extended credit (SA, $mil.)	usecon	
Diffusion, combination	log 2nd diff	farat		Adjusted reserves of depository institutions (SA, $mil.)	usecon	
Diffusion, combination	log 2nd diff	farmsr		Adj. monetary base including deposits to satisfy clearing balance contracts (SA, $bil.)	usecon	
Diffusion, combination	log 2nd diff	fm1		Money stock: M1 (SA, $bil.)	usecon	
Diffusion, combination	log 1st diff	fm2c		Real money stock: M2 (SA, chained 2000$bil.)	usecon	
Diffusion, combination	log 2nd diff	fm3		Money stock: M3 (SA, $bil.)	usecon	
Diffusion, combination[a]	log 1st diff	fxtwb[a]		Nominal broad trade-weighted exchange value of US$ (JAN 97=100)	usecon	
Diffusion, combination	log 1st diff	fxuk		Foreign exchange rate: United Kingdom (US$/Pound)	usecon	
Diffusion, combination	1st diff	faaa		Moody's seasoned Aaa corporate bond yield (% p.a.)	usecon	
Diffusion, combination	1st diff	fbaa		Moody's seasoned Baa corporate bond yield (% p.a.)	usecon	
Diffusion, combination	level		DAAA = faaa – ffed	Moody's seasoned Aaa corporate bond yield – fed funds rate(% p.a.)	usecon	
Diffusion, combination	level		DBAA = fbaa –ffed	Moody's seasoned Baa corporate bond yield – fed funds rate (% p.a.)	usecon	
Diffusion, combination	level	sdy5comm		S&P: Composite 500, dividend yield (%)	usecon	
Diffusion, combination	log 1st diff	sp500		Stock Price Index: Standard & Poor's 500 Composite (1941–43=10)	usecon	
Diffusion, combination	level	spe5comm		S&P: 500 Composite, P/E ratio, 4-qtr trailing earnings	usecon	
Diffusion, combination	log 1st diff	spny		Stock Price Index: NYSE Composite (Avg, Dec. 31, 2002=5000)	usecon	
Diffusion, combination	log 1st diff	spspi		Stock Price Index: Standard & Poor's 400 Industrials (1941–43=10)	usecon	
Diffusion, combination	1st diff	ftbs3		3-month Treasury bills, secondary market (% p.a.)	usecon	
Diffusion, combination	1st diff	ftbs6		6-month Treasury bills, secondary market (% p.a.)	usecon	
Diffusion, combination	level		DTBS03 = ftbs3 – ffed	3-month Treasury bills – fed funds rate, (% p.a.)	usecon	
Diffusion, combination	level		DTBS06 = ftbs6 – ffed	6-month Treasury bills – fed funds rate (% p.a.)	usecon	
Diffusion, combination	1st diff	fcm1		1-year Treasury bill yield at constant maturity (% p.a.)	usecon	
Diffusion, combination	1st diff	fcm5		5-year Treasury note yield at constant maturity (% p.a.)	usecon	
Diffusion, combination	level		DCM1 = fmc1 – ffed	1-year Treasury bill yield at constant maturity – fed funds rate (% p.a.)	usecon	
Diffusion, combination	level		DCM5 = fmc5 –ffed	5-year Treasury note yield at constant maturity – fed funds rate (% p.a.)	usecon	
Diffusion, combination	level		DCM10 = fcm10 –ffed	10-year Treasury note yield at constant maturity – fed funds rate (% p.a.)	usecon	
Diffusion, combination	log 2nd diff	sp1000		PPI: Crude materials for further processing (SA, 1982=100)	usecon	
Diffusion, combination	log 2nd diff	sp3100		PPI: Finished consumer goods (SA, 1982=100)	usecon	
Diffusion, combination	log 2nd diff	pcua		CPI-U: Apparel (SA, 1982–84=100)	usecon	
Diffusion, combination	log 2nd diff	pcucc		CPI-U: Commodities (SA, 1982–84=100)	usecon	
Diffusion, combination	log 2nd diff	pcuccd		CPI-U: Durables (SA, 1982–84=100)	usecon	
Diffusion, combination	log 2nd diff	pcucs		CPI-U: Services (SA, 1982–84=100)	usecon	
Diffusion, combination	log 2nd diff	pcum		CPI-U: Medical care (SA, 1982–84=100)	usecon	

DATA APPENDIX *(continued)*

Model	Transformation	Mnemonic	Constructed series mnemonic	Haver description	Haver database	Secondary source
Diffusion, combination	log 2nd diff	pcuslf		CPI-U: All items less food (SA, 1982–84=100)	usecon	
Diffusion, combination	log 2nd diff	pcuslm		CPI-U: All items less medical care (SA, 1982–84=100)	usecon	
Diffusion, combination	log 2nd diff	pcusls		CPI-U: All items less shelter (SA, 1982–84=100)	usecon	
Diffusion, combination	log 2nd diff	pcut		CPI-U: Transportation (SA, 1982–84=100)	usecon	
Diffusion, combination	log 2nd diff	jcdm		PCE: Durable goods: Chain Price Index (SA, 2000=100)	usna	
Diffusion, combination	log 2nd diff	jcm		PCE: Personal consumption expenditures: Chain Price Index (SA, 2000=100)	usna	
Diffusion, combination	log 2nd diff	jcnm		PCE: Nondurable goods: Chain Price Index (SA, 2000=100)	usna	
Diffusion, combination	log 2nd diff	jcsm		PCE: Services: Chain Price Index (SA, 2000=100)	usna	
Diffusion, combination	log 2nd diff	leconsa		Avg hourly earnings: Construction (SA, $/Hr)	usecon	
Diffusion, combination	log 2nd diff	lemanua		Avg hourly earnings: Manufacturing (SA, $/Hr)	usecon	
Diffusion, combination	1st diff		FCLQ = a0m101	Commercial & industrial loans outstanding (EOP, SA, chained 2000$mil.)	bci	
Diffusion, combination, indicators (5)	log 2nd diff	fm2		Money stock: M2 (SA, $bil.)	usecon	
Diffusion, combination, indicators (5)	1st diff	fcm10		10-year Treasury note yield at constant maturity (% p.a.)	usecon	
Diffusion, combination, indicators	1st diff	ffed		Federal funds [effective] rate (% p.a.)	usecon	
Diffusion, combination, indicators (4)	log 2nd diff	sp2000		PPI: Intermediate materials, supplies, and components (SA, 1982=100)	usecon	
Diffusion, combination, indicators (4)	log 2nd diff	sp3000		PPI: Finished goods (SA, 1982=100)	usecon	
Diffusion, combination, indicators (4)	level	napmpi		ISM: Mfg: Prices Index (NSA, 50+ = Econ Expand)	usecon	
Indicators (1) log	1st diff	zlead		Composite Index of 10 Leading Indicators (1996=100)	bci	
Indicators (3) log	1st diff		CPC = CONSTPV + CONSTPU	New construction put in place (SAAR, 2000$mil.)	usecon	
Indicators (3) log	1st diff	hn1us		New single-family houses sold: United States (SAAR, 000s)	usecon	
Indicators (1) log	1st diff	chm		Personal consumption expenditures (SAAR, chained 2000$mil.) (spliced from usna96 before 1990)	usna	
Indicators (1)	level	swxli2		Stock and Watson nonfinancial leading index %	usecon	
Indicators (5) level			CM03CM01 = fcm3 – fmc1	3-year/1-year T-bill spread	usecon	
Indicators[b] (5)	level	fxtwm[b]		Nominal trade-weighted exch value of US$/major currencies (MAR 73=100)	usecon	
Indicators (5)	log 1st diff		PZGLD = pzgld + mgold + fgold	Cash prices: gold, Handy & Harman Base Price (avg, spliced, $/Troy oz)	weekly	COMEX, FSC
Indicators (5)	log 1st diff		PZSIL	Cash price: silver, troy oz, Handy & Harman Base Price (avg, $/troy oz)	weekly	FSC
Indicators (4) log	1st diff	pzall		KR-CRB Spot Commodity Price Index: All commodities	usecon	
Indicators[c] (3) log	1st diff	spwpc[c]		SPOT COMMODITY PRICE - PLYWOOD, CROWS (PUIWMWPC_N.WT)		FAME
Indicators (4) log	1st diff		PFALL	KR-CRB Futures: All commodities (avg, 1967=100) weekly		BCRB
Indicators[d] (4) log	1st diff	pzdalu[d]		Aluminum ingot producer price: Delivered Midwest (avg, cents/lb)	weekly	
Indicators (4) log	1st diff	p101		PPI: Iron and steel (NSA, 1982=100)	usecon	
Indicators (4) log	1st diff	ueg		CPI-U: Energy (SA, 1982–84=100)	cpidata	
Natural rate	band-pass filtered	UGAP		Unemployment gap constructed from Perry-weighted unemployment rate	empl	
Prices	log 2nd diff	jcxfem		PCE less food and energy: Price Index (SA) (2000=100)	usna	

COMEX http://www.wrenresearch.com.au/downloads/index.htm
FSC http://www.webspace4me.net/~blhill2/data/commodities
BCRB http://economic-charts.com/em-cgi/data.exe/crb/crb01
FAME Federal Reserve Bank of San Francisco website

Indicator model groups:
1: Economic activity
2: Slackness measures
3: Housing and building activity
4: Industrial prices
5: Financial markets

[a] fxtwb begins in 1973:01
[b] fxtwm begins in 1973:01
[c] cspwpc begins in 1979:01
[d] pzdalu begins in 1988:07

Notes: SAAR is seasonally adjusted annual rate, SA is seasonally adjusted, NSA is not seasonally adjusted, NAICS is North American industry classification system, SIC is standard industrial classification, and EOP is end of period.

DATA APPENDIX *(continued)*

Quarterly data: 1967:1–2003:4

Model	Transformation	Mnemonic	Constructed series mnemonic	Haver description	Haver database
Activity	log 1st diff	ch		Real Personal Consumption Expenditures (SAAR, Bil. Chn. 2000 $)	usna
Activity	log 1st diff	cdh		Real Personal Consumption Expenditures: Durable Goods (SAAR, Bil. Chn. 2000 $)	usna
Activity	log 1st diff	cnh		Real Personal Consumption Expenditures: Non-Durable Goods (SAAR, Bil. Chn. 2000 $)	usna
Activity	log 1st diff	csh		Real Personal Consumption Expenditures: Services (SAAR, Bil. Chn. 2000 $)	usna
Activity	log 1st diff	ih		Real Gross Private Domestic Investment (SAAR, Bil. Chn. 2000 $)	usna
Activity	log 1st diff	fh		Real Private Fixed	usna
Activity	log 1st diff	fnsh		Real Private Nonresidential Structures	usna
Activity	log 1st diff	fneh		Real Private Nonresidential Equipment & Software	usna
Activity	1st diff	vh		Real Change in Private Inventories (SAAR, Bil. Chn. 2000 $)	usna
Activity	1st diff	xneth		Real Net Exports of Goods & Services (SAAR, Bil. Chn. 2000 $)	usna
Activity	log 1st diff	gh		Real Govt. Consumption Expenditures & Gross Investment (SAAR, Bil. Chn. 2000 $)	usna
Activity	log 1st diff	gfnh		Real Govt. Non-defense Consumption Expenditures & Gross Investment (SAAR, Bil. Chn. 2000 $)	usna
Activity	log 1st diff	ypdh		Real Disposable Personal Income (SAAR, Bil. Chn. 2000 $)	usna
Activity	log 1st diff	gdpbq		Index of Business Gross Value added	usna
Activity	log 1st diff	fsq		Index of Real Final Sales	usna
Activity, Indicators (1)	log 1st diff	gdph		Real Gross Domestic Product (SAAR, Bil. Chn. 2000 $)	usna
Activity, Indicators (3)	log 1st diff	fnh		Real Private Nonresidential	usna
Activity, Indicators (3)	log 1st diff	frh		Real Private Residential	usna
Diffusion, Combination	log 1st diff	lxba		Business Sector: Output per Hour of all Persons (SA,1992=100)	usecon
Diffusion, Combination	log 1st diff	lxbc		Business Sector: Compensation per Hour of all Persons (SA,1992=100)	usecon
Diffusion, Combination	log 1st diff	lxbr		Business Sector: Real Compensation per Hour of all Persons (SA,1992=100)	usecon
Diffusion, Combination	log 1st diff	lxbu		Business Sector: Unit Labor Costs (SA,1992=100)	usecon
Diffusion, Combination	log 1st diff	lxbn		Business Sector: Unit Non-Labor Payments (SA,1992=100)	usecon
Diffusion, Combination	log 1st diff	lxnfn		Non-farm Business Sector: Unit Non-Labor Payments (SA,1992=100)	usecon
Diffusion, Combination	log 1st diff	lxma		Manufacturing Sector: Output per Hour of all Persons (SA,1992=100)	usecon
Diffusion, Combination	log 1st diff	lxmda		Manufacturing Sector Durables: Output per Hour of all Persons (SA,1992=100)	usecon
Diffusion, Combination	log 1st diff	lxmna		Manufacturing Sector Non-durables: Output per Hour of all Persons (SA,1992=100)	usecon
Diffusion, Combination	log 1st diff	lxnca		Non-financial Corporations: Output per Hour, All employees (SA, 1992=100)	usecon
Diffusion, Combination	log 1st diff	lxncc		Non-financial Corporations: Compensation per Hour, All employees (SA, 1992=100)	usecon
Diffusion, Combination	log 1st diff	lxncr		Non-financial Corporations: Real Compensation per Hour, All employees (SA, 1992=100)	usecon
Diffusion, Combination	log 1st diff	lxncu		Non-financial Corporations: Unit Labor Costs, All employees (SA, 1992=100)	usecon
Diffusion, Combination	log 1st diff	lxncn		Non-financial Corporations: Unit Non-Labor Costs, All employees (SA, 1992=100)	usecon
Diffusion, Combination	log 1st diff	lxnct		Non-financial Corporations: Total Unit Costs, All employees (SA, 1992=100)	usecon
Diffusion, Combination	log 1st diff		BRULC= lxbu/lxbi	Business Sector: Real Unit Labor Costs (SA,1992=100)	usecon
Diffusion, Combination	log 1st diff		FRULC= lnxncu/lxnci	Non-financial Corporations: Real Unit Labor Costs, All employees (SA, 1992=100)	usecon
Diffusion, Combination	log 1st diff		BNLRULC= lxbn/lxbi	Business Sector: Real Unit Non-Labor Payments (SA,1992=100)	usecon
Diffusion, Combination	log 1st diff		NFNLRULC = lxnfn/lxnfi	Non-farm Business Sector: Real Unit Non-Labor Payments (SA,1992=100)	usecon
Diffusion, Combination	log 1st diff		FNLRULC= lxncn/lxnci	Non-financial Corporations: Real Unit Non-Labor Costs, All employees (SA, 1992=100)	usecon
Diffusion, Combination	log 1st diff		FTOTRUC= lnxnct/lxnci	Non-financial Corporations: Real Total Unit Costs, All employees (SA, 1992=100)	usecon
Diffusion, Combination	log 1st diff	grt		Government Total Receipts (SAAR, Bil. $)	usna
Diffusion, Combination	log 1st diff	get		Government Total Expenditures (SAAR, Bil. $)	usna
Diffusion, Combination	1st diff	gnl		Government Net Lending or Net Borrowing (SAAR, Bil. $)	usna
Diffusion, Combination	log 1st diff	dgdp		GDP Deflator	usna
Diffusion, Combination	log 1st diff	di		Gross Private Domestic Investment: Implicit Price Deflator (SA, 2000=100)	usna
Diffusion, Combination	log 1st diff	df		Private Fixed Investment: Implicit Price Deflator (SA, 2000=100)	usna
Diffusion, Combination	log 1st diff	dfn		Private Non-residential Fixed Investment: Implicit Price Deflator (SA, 2000=100)	usna
Diffusion, Combination	log 1st diff	dfns		Private Non-residential Structures: Implicit Price Deflator (SA, 2000=100)	usna
Diffusion, Combination	log 1st diff	dfne		Private Non-residential Equipment/Software: Implicit Price Deflator (SA, 2000=100)	usna
Diffusion, Combination	log 1st diff	dfr		Private Residential Investment: Implicit Price Deflator (SA, 2000=100)	usna
Diffusion, Combination	log 1st diff	dg		Government Consumption/Gross Investment: Implicit Price Deflator (SA, 2000=100)	usna
Diffusion, Combination	log 1st diff	dgfn		Federal Non-Defense Consumption/Investment: Implicit Price Deflator (SA, 2000=100)	usna
Diffusion, Combination	log 1st diff	dm		Imports of Goods & Services: Implicit Price Deflator (SA, 2000=100)	usna
Diffusion, Combination	log 1st diff	dx		Exports of Goods & Services: Implicit Price Deflator (SA, 2000=100)	usna
Diffusion, Combination, Indicators (6)	log 1st diff	lxnfa		Non-farm Business Sector: Output per Hour of all Persons (SA,1992=100)	usecon
Diffusion, Combination, Indicators (6)	log 1st diff	lxnfc		Non-farm Business Sector: Compensation per Hour of all Persons (SA,1992=100)	usecon
Diffusion, Combination, Indicators (6)	log 1st diff	lxnfr		Non-farm Business Sector: Real Compensation per Hour of all Persons (SA,1992=100)	usecon
Diffusion, Combination, Indicators (6)	log 1st diff	lxnfu		Non-farm Business Sector: Unit Labor Costs (SA,1992=100)	usecon
Diffusion, Combination, Indicators (6)	log 1st diff		NFRULC= lxnfu/lxnfi	Non-farm Business Sector: Real Unit Labor Costs (SA,1992=100)	usecon
Output Gap	band-pass filtered		OGAP	Output gap constructed from band-pass filtered Real GDP	usna
Real Unit Labor Cost Gap	band-pass filtered		RGAP	Band-pass filtered version of Non-farm Business Sector Real Unit Labor Costs	usecon
Prices	log 2nd diff	jcxfe		PCE less food and Energy: Price Index (SA) (2000=100)	usna

Indicator Model Groups:
1: Economic Activity
2: Slackness Measures
3: Housing and Building Activity
4: Industrial Prices
5: Financial Markets
6: Productivity and Marginal Cost

REFERENCES

Atkeson, Andrew, and Lee E. Ohanian, 2001, "Are Phillips curves useful for forecasting inflation?," *Quarterly Review*, Federal Reserve Bank of Minneapolis, Vol. 25, No. 1, Winter, pp. 2–11.

Cecchetti, Stephen G., Rita S. Chu, and Charles Steindel, 2000, "The unreliability of inflation indicators," *Current Issues in Economics and Finance*, April, Vol. 6, No. 4.

Christiano, Lawrence J., and Terry Fitzgerald, 1999, "The band-pass filter," National Bureau of Economic Research, working paper, No. 7257, July.

Clark, Todd E., and Michael W. McCracken, 2004, "The predictive content of the output gap for inflation: Resolving in-sample and out-of-sample evidence," manuscript.

__________, 2001, "Tests of equal forecast accuracy and encompassing for nested models," *Journal of Econometrics*, Vol. 105, November, pp. 85–110.

Fisher, Jonas D. M., Chin Liu, and Ruilin Zhou, 2002, "When can we forecast inflation?," *Economic Perspectives*, Federal Reserve Bank of Chicago, First Quarter, pp. 30–42.

Marcellino, Massimiliano, James H. Stock, and Mark Watson, 2004, "A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series," manuscript.

Sargent, Thomas, 1999, *The Conquest of American Inflation*, Princeton, NJ: Princeton University Press.

Sims, Christopher, 2002, "The role of models and probabilities in the monetary policy process," *Brookings Papers on Economic Activity*, Vol. 2, pp.1–62.

Stock, James H., and Mark Watson, 2003, "Forecasting output and inflation: The role of asset prices," *Journal of Economic Literature*.

__________, 2002, "Macroeconomic forecasting using diffusion indexes," *Journal of Business and Economic Statistics*, Vol. 20, No. 2, April, pp. 147–162.

__________, 1999, "Forecasting inflation," *Journal of Monetary Economics*, Vol. 44, No. 2.