ELSEVIER Journal of Monetary Economics 44 (1999) 293–335

Journal of MONETARY ECONOMICS

www.elsevier.nl/locate/econbase

Forecasting inflation☆

James H. Stock[a], Mark W. Watson[b,*]

[a]*Kennedy School of Government, Harvard University, Cambridge, MA 02138, USA and National Bureau of Economic Research, Cambridge, MA 02138, USA*

[b]*Woodrow Wilson School, Princeton University, Princeton, NJ 08544, USA*

Received 4 November 1998; received in revised form 27 July 1999; accepted 27 July 1999

Abstract

This paper investigates forecasts of US inflation at the 12-month horizon. The starting point is the conventional unemployment rate Phillips curve, which is examined in a simulated out-of-sample forecasting framework. Inflation forecasts produced by the Phillips curve generally have been more accurate than forecasts based on other macroeconomic variables, including interest rates, money and commodity prices. These forecasts can however be improved upon using a generalized Phillips curve based on measures of real aggregate activity other than unemployment, especially a new index of aggregate activity based on 168 economic indicators.

JEL classification: E31; C32

Keywords: Phillips curve; Forecast combination

1. Introduction

The Phillips curve has played a prominent role in empirical macroeconomics in the US over the past four decades. As a tool for forecasting inflation, it is

☆We have benefited from comments by Guido Kuersteiner, Brian O'Reilly, a referee, and participants at the Studienzendrum Gerzensee Conference on the 'Return of the Phillips Curve'. This research was supported by National Science Foundation grant SBR-9730489.

* Corresponding author. Tel.: + 1-609-258-4811; fax: + 1-609-258-5533.
E-mail address: mwatson@princeton.edu (M.W. Watson)

0304-3932/99/$ - see front matter
PII: S0304-3932(99)00027-6

widely regarded as stable, reliable and accurate, at least compared to the alternatives. Alan Blinder, former Vice Chairman of the Board of Governors of the US Federal Reserve System, called it the 'clean little secret' of empirical macroeconomics.

This paper reassesses the use of the Phillips curve for forecasting price inflation. We focus on three questions. First, has the US Phillips curve been stable? If not, what are the implications of the instability for forecasting future inflation? Second, the Phillips curve is conventionally specified in terms of unemployment, but at a conceptual level other measures of economic activity could be used instead. Do these alternative Phillips curves provide better forecasts of inflation than the unemployment rate Phillips curve? Third, these variables are, of course, a small subset of the many macroeconomic variables that are potentially useful for forecasting inflation. For example, monetary theories of inflation and the theory of the term structure of interest rates suggest alternative frameworks for forecasting inflation. How do inflation forecasts from the Phillips curve stack up against time-series forecasts made using interest rates, money, and other series? Put baldly, is it time for inflation forecasters to move beyond the Phillips curve?

The focus of this paper is on forecasting price inflation using monthly data for the US from 1959:1 to 1997:9. Attention is restricted to forecasts of inflation over a 12-month horizon. All forecasting comparisons are performed using a simulated out-of-sample methodology, that is, all models are estimated with data that is dated prior to the forecast period. This empirical analysis suggests some answers to these questions.

First, we find that there is statistical evidence that the parameters of the Phillips curve, as conventionally specified, have changed over this period. The major source of instability seems to be changes in the contribution of lags of inflation in the Phillips curve. While this instability is statistically significant, it appears to be quantitatively small.[1]

Second, Phillips curves specified with alternative measures of real economic activity can provide forecasts with smaller mean squared errors than those from unemployment-based Phillips curves. For example, Phillips curves that use housing starts, capacity utilization or the rate of growth of manufacturing and trades sales produce forecasts that are generally more accurate than forecasts constructed from Phillips curves using the unemployment rate.

[1] For the past few years, inflation has consistently been below the forecasts made by conventional Phillips curve specifications. This has raised the possibility of a large decline in the NAIRU in the mid-1990s or possibly a broader breakdown of the Phillips curve altogether; see Gordon (1998) for a discussion. Although they are important, these developments are, we believe, too recent to make a clear assessment about stability given the available data. We therefore focus on stability in larger subsamples and defer the issue of instability in the mid-1990s to future work.

Third, it is possible to improve upon traditional Phillips curve forecasts by using alternative economic indicators to forecast inflation. The investigation here casts a wide net: we consider forecasts of inflation based on 167 additional economic indicators. Several conclusions emerge. Although there are theoretical reasons to expect interest rates and interest rate spreads to be useful for predicting inflation, forecasts based on these variables fail to improve on Phillips curve forecasts, at least at the one year horizon. The evidence on nominal money is less clear cut: models that add indexes of the money supply to the Phillips curve provide marginal improvements for some sample periods and some measures of inflation, but they lead to a serious deterioration in accuracy for forecasts of inflation based on the consumer price index during the 1970s and early 1980s. Commodity prices do not improve inflation forecasts at the 12-month horizon. The only variables that consistently improve upon Phillips curve forecasts are measures of aggregate activity, and the best of these is a new index of 168 indicators of economic activity. These alternative forecasts, when combined with Phillips curve forecasts, produce forecasting gains that are both statistically and economically significant.

These results lead us to conclude that the unemployment rate Phillips curve can play a useful role in forecasting inflation, but that relying on it to the exclusion of other forecasts is a mistake. Forecasting relations based on other measures of aggregate activity can perform as well or better than those based on unemployment, and combining these forecasts produces still further improvements.

The remainder of the paper is organized as follows. In Section 2, we examine the stability of standard specifications of the Phillips curve. In Section 3, Phillips curves based on alternative measures of aggregate activity are considered. In Section 4, forecasts of inflation from the Phillips curve are compared with forecasts based on our full set of 168 economic indicators. Section 5 considers multivariate forecasts of inflation that use all 168 indicators. The results in Sections 2–5 maintain the conventional assumption that inflation is integrated of order 1 (i.e. $I(1)$), and the robustness of our results to this assumption is investigated in Section 6. Section 7 concludes.

2. Stability of the US Phillips curve, 1959–1997

Conventional specifications of the Phillips curve relate the change of inflation to past values of the unemployment gap (the difference between the unemployment rate and the NAIRU), past changes of inflation, and current and/or past values of variables that control for various supply shocks.[2] Because we are

[2] This framework is due to Gordon (1982) and forms the basis for estimates of the NAIRU (see for example, Congressional Budget Office (1994), Fuhrer (1995), and Council of Economic Advisors (1998)).

interested forecasting, we adopt this framework with two modifications: the dependent variable is the change in the inflation rate over periods longer than the sampling frequency, and supply shocks measures are not included in the equation. The first modification allows us to use the estimated equation directly for multiperiod (12-month-ahead) forecasting. Supply shock measures are omitted because preliminary results (not reported here) indicated that the forecasting performance of models that included these variables (the relative price of food and energy and the Nixon price control variable as in Gordon (1982,1997)) is worse, on a simulated out of sample basis, than the corresponding models in which these variables are excluded. This is not surprising: although the supply shock variables are statistically significant in full-sample specifications with unemployment, in a simulated out of sample setting their coefficients are poorly estimated for much of the sample and this produces poor out of sample forecasts. This is consistent with these supply shock measures being identified as useful in unemployment-based Phillips curves based on ex post analysis.

The Phillips curve specification used in this paper is

$$\pi^h_{t+h}\text{-}\pi_t = \phi + \beta(L)u_t + \gamma(L)\Delta\pi_t + e_{t+h}, \tag{1}$$

where $\pi^h_t = (1200/h)\ln(P_t/P_{t-h})$ is the h-period inflation in the price level P_t, reported at an annual rate; $\pi_t \equiv \pi^1_t = 1200 * \ln(P_t/P_{t-1})$ is monthly inflation at an annual rate; u_t is the unemployment rate; and $\beta(L)$ and $\gamma(L)$ are polynomials in the lag operator L.

This specfication imposes two important restrictions. The first is that inflation is integrated of order one (i.e. $I(1)$). The specification (1) is equivalent to a specification with π^h_{t+h} as the left hand variable and replacing $\gamma(L)\Delta\pi_t$ with, say, $\mu(L)\pi_t$, subject to the restriction that $\mu(1) = 1$. Thus, for $h = 12$, this specification can be thought of as predicting inflation over the next twelve months using a distributed lag of current and past inflation, subject to the restiction that the distributed lag coefficients sum to one. Modeling US price inflation as $I(1)$ is standard in this literature, and as we discuss below, is consistent with recursive unit-root tests of various inflation series over most of the sample period. The robustness of the main substantive results to relaxing the unit root assumption is examined in the penultimate section of this paper.

The second restriction imposed in Eq. (1) is that the NAIRU is constant. To see this, note that the Phillips curve is conventionally written as

$$\pi^h_{t+h} - \pi_t = \beta(L)(u_t - \bar{u}_t) + \gamma(L)\Delta\pi_t + e_{t+h}, \tag{2}$$

where $\bar{u}_t$ is the NAIRU. When $\bar{u}_t$ is time invariant so that $\bar{u}_t = \bar{u}$, then Eq. (2) can be written as Eq. (1) with the constant term $\phi = -\beta(1)\bar{u}$. There is a large recent literature on the constancy of the NAIRU, and the constancy of the Phillips curve more generally (see Gordon (1997,1998)), King and Watson (1994), Shimer (1998), Staiger et al. (1997a,b), Stock (1998)). This research documents instability in the coefficients of specifications like (1) using post-war data for the US.

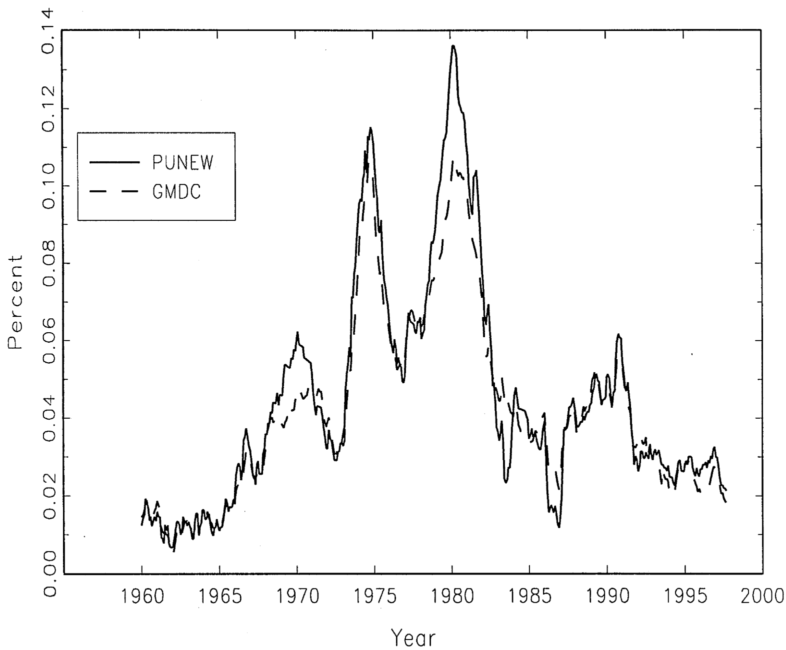

Fig. 1. Annual inflation.

Instability in Eq. (1) has obviously important implications for forecasting, and thus we will examine stability of the coefficients in Eq. (1) before discussing the forecasting performance of the Phillips curve.

Our estimates use monthly data for the US, 1959:1–1997:9. Fig. 1 plots annual inflation rates, π_t^{12}, for two closely watched US monthly price indexes: the consumer price index (CPI-U; the mnemonic in the figure is PUNEW[3]) and the personal consumption expenditure (PCE) deflator (GMDC in the figure). Although the two measures of inflation are generally similar, there are marked differences in 1970, 1975 and 1980 (when CPI inflation was much higher than PCE inflation) and in 1983 and 1986 (when CPI inflation was much lower than PCE inflation). The causes of the differences in the series are well known: the CPI is essentially a Laspeyres index which uses a fixed basket to weight its constituent prices, while the PCE deflator uses chain weighting; the CPI data are not historically revised when methods or data change, while the PCE deflator is

[3] For series taken from the database formerly known as CITIBASE, the CITIBASE mnemonics are used consistently in the tables and in the appendix.

subject to revision. Because a major change in the CPI occurred in 1983, when the owner-occupied housing component was changed, results will also be presented for CPI inflation with housing services eliminated (*PUXHS*). Two unemployment rates are considered: the total civilian unemployment rate (*LHUR*), and the unemployment rate for males in the age group 25–54 (*LHMU25*). The latter series is included to control for potential demographic shifts that could affect the stability of the coefficients, in particular the large increase in female labor force participation rates over this period.

Several tests for the stability of the parameters in Eq. (1) were performed. All are variants of the Quandt (1960) likelihood ratio (QLR) procedure, which tests for a single breakpoint in the regression. The tests were implemented as the maximum of HAC-robust Wald statistics for shifts in the coefficients over all possible break dates in the middle 70% of the sample; p-values for the statistics are computed using the approximation given in Hansen (1997). Results are shown in Table 1 for regressions estimated over horizons $h = 1$ and $h = 12$. The first statistic (QLR_{all}) tests for the constancy of all the parameters in Eq. (1). The next statistic ($QLR_{\phi,\beta}$) tests for stability of the constant term (and hence the NAIRU) together with the coefficients on the lags of the unemployment rate ($\beta(L)$) assuming that the coefficients in $\gamma(L)$ are constant. Similarly, QLR_{γ} tests for the stability of the coefficients on lagged changes in inflation ($\gamma(L)$) assuming ϕ and $\beta(L)$ are constant. For each combination of price and unemployment rate data, the number of lags in $\beta(L)$ and $\gamma(L)$ were chosen separately by the Bayes information criterion (BIC) over the full sample, where in both cases the number of lags was permitted to be between 0 and 11.

The QLR statistics in Table 1 indicate statistically significant evidence of instability in these empirical Phillips curves. This instability appears to be concentrated in the coefficients on lagged inflation: while the QLR_{all} and QLR_{γ} statistics are statistically significant, the $QLR_{\phi,\beta}$ statistics provide far less evidence of instability in the NAIRU and in the effect of unemployment on future values of inflation. Importantly, while the instability in $\gamma(L)$ is statistically significant, it does not seem to be quantitatively large, particularly in its effect on 12-month ahead forecasts. Fig. 2 plots estimates of the accumulated values of $[1 - L\gamma(L)]^{-1}$ (the impulse responses from e_t to future values of π_t holding the unemployment rate constant) estimated over the first and second half of the samples for the CPI and PCE deflator using LHUR. These impulse responses are broadly similar across the two sample periods, and most of the differences occur for horizons less than 12 months. This evidence is consistent with results presented in King and Watson (1994), who found statistically significant shifts in the coefficients of a bivariate VAR fit to post-war US inflation and unemployment data, but found that these shifts had little effect on the forecasts produced by the VAR.

In the forecasting experiments that we carry out in later sections we will ignore this instability, except to the extent that it is captured in recursive

Table 1
Stability tests for the Phillips curve regression model

$$\pi^h_{t+h} - \pi_t = \phi + \beta(L)u_t + \gamma(L)\Delta\pi_t + e_{t+h}$$

Panel A: One-month ahead regressions ($h = 1$)

		P-values for QLR test statistics		
Price index	Unemp. rate	QLR_{all}	$QLR_{\phi,\beta}$	QLR_{γ}
Punew	Lhur	0.00	0.58	0.01
	Lhmu25	0.00	0.62	0.02
GMDC	Lhur	0.13	0.99	0.05
	Lhmu25	0.12	0.94	0.05
Puxhs	Lhur	0.00	0.68	0.00
	Lhmu25	0.00	0.85	0.00

Panel B: One-year ahead regressions ($h = 12$)

Punew	Lhur	0.00	0.00	0.00
	Lhmu25	0.00	0.01	0.00
GMDC	Lhur	0.01	0.09	0.07
	Lhmu25	0.03	0.37	0.03
Puxhs	Lhur	0.00	0.03	0.00
	Lhmu25	0.00	0.19	0.00

Notes: QLR_{all} tests all of the regression coefficients over all possible break points in the middle 70% of the sample. The other statistics test subsets of the coefficients under the maintained assumption that the other coefficients are constant. $QLR_{\phi,\beta}$ tests ϕ and the coefficients of $\gamma(L)$, and QLR_{γ} test the coefficients of the lag polynomial $\gamma(L)$. The Wald form of the QLR statistics using a HAC covariance matrix for the estimated parameters (constructed using a Bartlett kernel using $h-1$ lags); p-values are computed using the approximation given in Hansen (1997). The sample period is $t = $ 1960:2–1996:9.

estimates of the regression coefficients. We do this for two reasons. First, Fig. 2 shows that the instability is small, so that gains from incorporating this instability are likely to be modest at best. In fact, when instability is small, existing statistical forecasting methods that incorporate parameter instability (rolling regression, TVP models, etc.) perform no better than recursive least squares, and in many cases perform significantly worse (for some empirical evidence, see Stock and Watson (1996)). Second, this instability has been identified in a full-sample analysis, and incorporating it into the models is inconsistent with the simulated real-time methodology of the forecasting exercise.

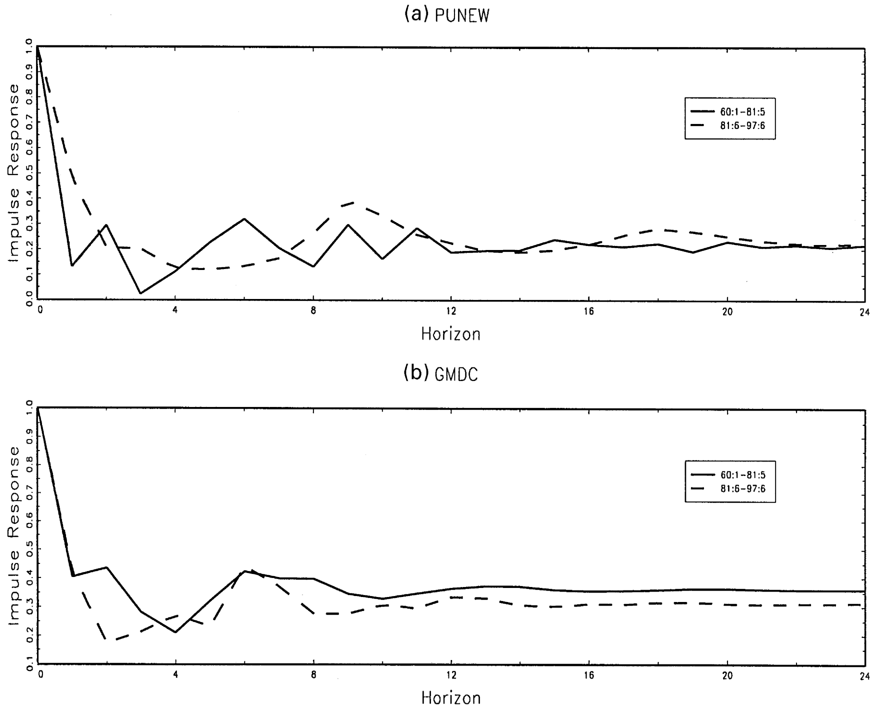

Fig. 2. Estimated impulse responses for different sample periods.

3. Inflation forecasts based on measures of aggregate real activity

Although the Phillips curve is typically specified in terms of the deviation of unemployment from its natural rate, more generally it is a relation between inflation and aggregate real activity. This section compares the forecasting performance of the conventional unemployment rate Phillips curve to generalized Phillips curves that use other measures of aggregate activity.

The forecasting models used here are analogous to Eq. (1) except that the alternative indicator, x_t, replaces unemployment:

$$\pi^h_{t+h} - \pi_t = \phi + \beta(L)x_t + \gamma(L)\Delta\pi_t + e_{t+h}. \quad (3)$$

In (3), it is assumed that x_t has already been transformed so that it is $I(0)$. This assumes that inflation and the alternative demand measure are not cointegrated, an assumption that is theoretically and empirically plausible for real activity measures (robustness to this assumption is examined in Section 6). Specification (3) mirrors specification (1). The constant intercept implies that, under (3), the 'natural rate of x_t' is constant.

Seven alternative measures of aggregate activity are considered: industrial production (*IP*), real personal income (*GMPYQ*), total real manufacturing and trade sales (*MSMTQ*), the number of employees on nonagricultural payrolls (*LPNAG*), the capacity utilization rate in manufacturing (*IPXMCA*), and housing starts (*HSBP*). We also consider the unemployment rate for males age 25 to 54 (*LHMU25*). The data source and full definitions of each series are summarized in Appendix A.

The last three activity variables (*IPXMCA*, *HSBP*, *LHMU25*) are approximately $I(0)$ variables and can be used directly in Eq. (3). The first four variables (*IP*, *GMPYQ*, *MSMTQ*, *LPNAG*) contain significant trend components so that Eq. (3) applies when x_t is interpreted as deviations from trend. There is a large literature on methods for detrending these variables so as to construct estimates of an 'output gap'. Familiar approaches include methods that use segmented trends with break points determined by historically dated business cycles, methods based on estimates of aggregate production functions, time-series filtering methods, and combinations of these methods; (see Kuttner (1994) for a brief survey). An important limitation of many of these methods is that they estimate x_t using both future and past values of the series, making them unsuitable for forecasting. We experimented with several methods that are suitable for forecasting and report results for estimates of x_t based on a one-sided version of the Hodrick–Prescott (1981) (HP) filter. This procedure produces plausible trend and gap estimates for each of the variables analyzed here. The one-sided HP filter is convenient and preserves the temporal ordering of the data. Of course, improved forecasting performance might obtain if alternative, possibly multivariate, one-sided estimates of the trend components of these series were used.

The one-sided HP trend estimate is constructed as the Kalman filter estimate of τ_t in the model:

$$y_t = \tau_t + \varepsilon_t, \tag{4}$$

$$(1 - L)^2 \tau_t = \eta_t, \tag{5}$$

where y_t is the logarithm of the data series, τ_t is the unobserved trend component and $\{\varepsilon_t\}$ and $\{\eta_t\}$ are mutually uncorrelated white noise sequences with relative variance $q = \mathrm{var}(\eta_t)/\mathrm{var}(\varepsilon_t)$. As discussed in Harvey and Jaeger (1993) and King and Rebelo (1993), the HP-filter is the optimal (linear minimum mean square error) two-sided trend extraction filter for Eqs. (4) and (5). Because our focus is on forecasting, we use the optimal one-sided analogue of this filter, so that future values of y_t (which would not be available for real time forecasting) are not used in the detrending operation. We use a value of q for our monthly data ($q_{\mathrm{monthly}} = 0.75 * 10^{-6}$) that approximately matches the spectral gain for the HP-filter typically applied to quarterly data (which uses $q_{\mathrm{quarterly}} = 0.675 * 10^{-3}$). We also report forecasting results using $x_t = \Delta y_t$ to gauge the robustness of our results to this choice of detrending.

The empirical analysis examines the forecasting performance of the candidate series x_t in a simulated out of sample forecasting exercise. This entails making forecasts using only data dated before the forecast period. For example, consider the forecast of the (twelve month) inflation rate from 1980:1 to 1981:1, made in 1980:1. To compute this forecast, all the models are estimated, information criteria are computed, and lag lengths are selected using data through 1980:1, at which point the forecast of inflation over 1980:1 to 1981:1 is made. Moving forward one month, all the models are reestimated (and information criteria computed and models selected) using data through 1980:2, and the forecast of inflation over 1980:2–1981:2 is computed. For each series x_t, this produces a single series of forecast errors based on simulated out-of-sample (also termed recursive) estimation and model selection. The data set begins in 1959:1, and the first observation used in the regressions is 1960:2 (earlier observations are used for initial conditions in the regressions). The period over which simulated out of sample forecasts are computed and compared is 1970:1 through 1996:9.

The dependent variables in this and subsequent sections are based on the CPI and, alternatively, the PCE deflator. The results using the CPI without housing are similar to those for the CPI and are not reported.

Several statistics are computed to summarize the performance of the simulated out of sample forecasts. One is the mean-squared-error (MSE) of forecasts based on x_t, relative to the MSE of forecasts based on the unemployment rate (*LHUR*). A HAC standard error of this relative mean-squared-error is also reported. (See West (1996) for an asymptotic justification of this procedure using recursively estimated models.)

The remaining statistics assess whether the candidate variable makes a useful forecasting contribution, relative to unemployment. A forecast combining regression provides a simple device for comparing the simulated out of sample performance of the two non-nested models (the model incorporating x_t and the model using the unemployment rate). This is done in the forecast combination regression,

$$\pi^h_{t+h}\text{-}\pi_t = \lambda f^x_t + (1-\lambda) f^u_t + \varepsilon_{t+h}, \tag{6}$$

where f^x_t is the forecast of $\pi^h_{t+h} - \pi_t$ based on the candidate series x, made at date t, f^u_t is the corresponding forecast based on the unemployment rate, and ε_{t+h} is the forecast error associated with the combined forecast. If $\lambda = 0$, then forecasts based on x_t add nothing to forecasts based on unemployment; if $\lambda = 1$, then forecasts based on the unemployment rate add nothing to forecasts based on x_t.

The results are summarized in Table 2. Results are shown for two forecast sub-samples: 1970–1983 and 1984–1996. The last row of the table, labeled LHUR RMSE, shows the root-mean-square-error for the benchmark Phillips curve specification. The other entries in the table show the relative mean square error of the alternative models and the OLS estimates of λ.

Table 2
Forecasting performance of alternative real activity measures

		PUNEW				GMDC			
		1970–1983		1984–1996		1970–1983		1984–1996	
Variable	Trans	Rel. MSE	λ	Rel. MSE	λ	Rel. MSE	λ	Rel. MSE	λ
No change		1.90	0.11	2.44	0.06	1.30	0.30	2.78	− 0.05
		(0.59)	(0.07)	(1.59)	(0.08)	(0.18)	(0.15)	(1.31)	(0.05)
Univariate	–	1.26	− 0.13	0.98	0.53	1.00	0.50	1.06	0.27
		(0.19)	(0.25)	(0.15)	(0.33)	(0.15)	(0.38)	(0.09)	(0.29)
'Gaps' specifications									
ip	DT	1.11	0.04	0.91	0.84	0.97	0.62	0.99	0.58
		(0.11)	(0.34)	(0.08)	(0.29)	(0.08)	(0.37)	(0.04)	(0.26)
gmpyq	DT	1.23	− 0.11	1.11	0.33	1.14	0.04	1.11	0.26
		(0.16)	(0.26)	(0.12)	(0.13)	(0.12)	(0.29)	(0.10)	(0.18)
msmtq	DT	0.96	0.67	0.87	0.83	0.90	1.03	0.92	0.83
		(0.08)	(0.35)	(0.11)	(0.24)	(0.09)	(0.43)	(0.09)	(0.37)
lpnag	DT	1.08	0.14	0.93	0.73	1.09	0.02	0.93	0.83
		(0.12)	(0.46)	(0.08)	(0.28)	(0.11)	(0.45)	(0.08)	(0.35)
ipxmca	LV	0.99	0.56	0.85	0.95	0.95	0.91	0.96	0.72
		(0.06)	(0.32)	(0.09)	(0.27)	(0.06)	(0.49)	(0.06)	(0.30)
hsbp	LN	0.85	0.94	1.03	0.47	0.89	0.81	0.90	0.65
		(0.10)	(0.26)	(0.24)	(0.23)	(0.15)	(0.37)	(0.17)	(0.26)
lhmu25	LV	1.04	0.21	1.04	0.32	1.00	0.52	1.01	0.44
		(0.06)	(0.41)	(0.10)	(0.36)	(0.06)	(0.50)	(0.06)	(0.36)
First differences specifications									
ip	DLN	1.00	0.51	1.09	0.26	0.88	1.11	1.13	0.13
		(0.05)	(0.30)	(0.12)	(0.25)	(0.15)	(0.60)	(0.09)	(0.19)
gmpyq	DLN	0.88	0.79	1.25	0.30	0.65	1.38	1.20	0.33
		(0.08)	(0.20)	(0.24)	(0.14)	(0.22)	(0.29)	(0.18)	(0.13)
msmtq	DLN	0.83	1.38	0.97	0.55	0.84	1.23	1.02	0.45
		(0.07)	(0.27)	(0.13)	(0.24)	(0.16)	(0.51)	(0.11)	(0.23)
lpnag	DLN	0.94	0.82	0.92	0.74	0.87	1.21	0.92	0.84
		(0.06)	(0.27)	(0.09)	(0.28)	(0.13)	(0.53)	(0.08)	(0.35)
dipxmca	DLV	0.97	0.64	1.13	0.21	0.90	0.96	1.15	0.14
		(0.07)	(0.36)	(0.16)	(0.29)	(0.15)	(0.57)	(0.10)	(0.16)
dhsbp	DLN	1.28	− 0.05	1.05	0.42	1.03	0.43	1.05	0.31
		(0.19)	(0.26)	(0.16)	(0.23)	(0.16)	(0.35)	(0.09)	(0.28)
dlhmu25	DLV	0.97	0.67	1.16	− 0.09	0.94	0.80	1.10	0.07
		(0.08)	(0.44)	(0.12)	(0.28)	(0.15)	(0.67)	(0.08)	(0.23)
dlhur	DLV	0.95	1.03	1.12	− 0.47	0.90	1.05	1.07	0.20
		(0.06)	(0.55)	(0.11)	(0.68)	(0.17)	(0.79)	(0.08)	(0.25)
Phillips curve RMSEs (% per annum)									
LHUR RMSE			2.4		1.4		1.9		1.0

Table 2 (continued)

Notes: All results are for simulated out-of-sample forecasts as discussed in the text. The first row of the table shows results for the 'No change' (martingale) forecast of inflation and the next row, 'Univariate', shows results for a univariate autoregression. *LHUR RMSE* denotes the root-mean-square forecast error constructed using recursively estimated coefficients in Eq. (3) using $x_t = LHUR$ and a forecasting horizon of one year ($h = 12$). For a series y_t, the transformations $x_t = f(y_t)$ are; $x_t = y_t$ (LV), $x_t = \Delta y_t$ (DLV), $x_t = \Delta^2 y_t$ (DDLV), $x_t = \ln(y_t)$ (LN), $x_t = \Delta[\ln(y_t)]$ (DLN), $x_t = \Delta^2[\ln(y_t)]$ (DDLN), $x_t = \ln(y_t) - \tau_t$ (DT) where τ_t is the one-sided HP-trend component of $\ln(y_t)$ described in the text. The entries Rel. MSE show the mean-square-forecast-error (MSE) using the variable given in the first column and computed from recursively estimated coefficients in Eq. (3) relative to the MSE using LHUR. The column labeled λ shows OLS estimate of λ from Eq. (6). HAC robust standard errors (estimated using a Bartlett kernel with 12 lags) are shown in parentheses. The forecasts were computed over the sample period 1970:1–1996:9. The first sample used for the using recursively estimated regressions was 1960:2.

Several findings emerge from the table. There are important differences in the forecastability of inflation across price series and over time. PCE inflation forecasts are more accurate than CPI forecasts: over the entire sample period the RMSE for the PCE is approximately 25% smaller than for the CPI. Forecast errors are much smaller in the second half of the forecast period (1984–1996) than in the first half (1970–1983): the RMSE drops by over 40% for both inflation measures. There is considerable forecastable variation in inflation changes: the relative MSEs of the 'No Change' forecast (i.e., the model that forecasts no change in the inflation rate) are much larger than the relative MSEs of any an other forecasting models. Forecasts using the unemployment rate generally outperform univariate autoregressions (the relative MSEs for the univariate autoregressions are greater than 1.0), but the forecasting gain is quantitatively large only for CPI inflation in the 1970–1983 subsample.

Two variables (capacity utilization (*IPXMCA*) and manufacturing and trade sales (*MSMTQ*)) outperform the unemployment rate uniformly across series and sample period. Many of the estimated values of λ are significantly greater than 0, suggesting that these alternative activity measures contain useful information not included in lags of the unemployment rate or past inflation. Finally, specifications using the first difference of the activity variables produce more accurate forecasts than specifications using 'gaps' for the early sample period, but this reverses in the later sample period, when gaps perform better than first differences.

4. Bivariate inflation forecasts using other economic indicators

We now turn to the broader question of how these activity-based forecasts of inflation compare with forecasts based on other economic indicators. Some of these series are suggested by theory. For example, the expectations hypothesis of the term structure of interest rates suggests that spreads between interest rates of different maturities incorporate the forecasts of inflation made by market participants. Similarly, the quantity theory of money predicts that, in the long run, the rate of inflation is determined by the long-run growth rate of monetary aggregates. In addition, we also consider series that are not necessarily identified by a macroeconomic theory but which represent various aspects of the macroeconomy and/or have previously been used as leading indicators.

In all, 168 candidate series are used to generate simulated forecasts of inflation that can be compared with forecasts based on unemployment and on the alternative activity measures. The series are listed and described in Appendix A. The methodology for assessing the performance of the candidate indicator is identical to that of Section 3.

The results are contained in Table 3. The first panel of the table gives results for interest rates (first difference form) and yield curve spreads (all relative to the three month Treasury bill rate). Bivariate models with these interest rates perform worse than the benchmark Phillips curve model, and indeed their performance is typically inferior to the univariate autoregression. With only one exception, the relative mean-square-errors exceed unity for all of interest rate variables, for all sample periods and for both price series. (The exception is a value of 0.97 for the one-year yield curve spread for the CPI in the 1984–1996 sample period.) Some of the estimating combining weights are positive and statistically significant, which suggests that including interest rates may improve the forecasting performance of the benchmark model. However, it is important to note that the estimated combining weights for the univariate model are also greater than zero in the second subsample (although not statistically significant), suggesting that the benchmark model relies too heavily on the unemployment rate.

The next panel shows results for measures of the nominal money supply. Included are results using money growth rates and their first differences. These models do not perform well. The best performing money supply models are comparable to the univariate autoregressions, presumably because the estimated coefficients on money are very close to zero in these specifications.

Results for 140 additional indicators are also included in the table. Models incorporating exchange rates do not perform as well as the benchmark Phillips curve model or the univariate autoregression. Models incorporating different price indexes, including commodity prices, produce forecasts that are very similar to forecasts produced by the univariate models. Lags of nominal wages do not seem to add information beyond that contained in lags of prices. The conclusion that emerges from looking across all of the variables in Table 3 is that many of the models that use real activity variables dominate univariate autoregressions and the benchmark Phillips curve (for example, see the rows labeled *PMI*, *HSFR*, *LP*, *LHELX*), but models that use other variables (asset

Table 3
Forecasting performance of various economic indicators

		PUNEW				GMDC			
		1970–1983		1984–1996		1970–1983		1984–1996	
Variables	Trans	Rel. MSE	λ	Rel. MSE	λ	Rel. MSE	λ	Rel. MSE	λ
Univariate	–	1.26	– 0.13	0.98	0.53	1.00	0.50	1.06	0.27
		(0.19)	(0.25)	(0.15)	(0.33)	(0.15)	(0.38)	(0.09)	(0.29)
Interest rates									
fyff	DLV	1.34	0.05	1.02	0.44	1.07	0.37	1.06	0.25
		(0.33)	(0.16)	(0.15)	(0.33)	(0.20)	(0.35)	(0.08)	(0.29)
fycp	DLV	1.25	0.06	1.04	0.42	1.03	0.42	1.07	0.23
		(0.18)	(0.17)	(0.16)	(0.33)	(0.16)	(0.38)	(0.08)	(0.30)
fygm3	DLV	1.27	0.06	1.01	0.47	1.09	0.31	1.06	0.25
		(0.24)	(0.20)	(0.15)	(0.31)	(0.19)	(0.38)	(0.08)	(0.29)
fygm6	DLV	1.25	0.03	1.04	0.42	1.02	0.46	1.06	0.24
		(0.21)	(0.22)	(0.15)	(0.31)	(0.16)	(0.43)	(0.08)	(0.29)
fygt1	DLV	1.21	0.08	1.03	0.42	1.02	0.45	1.06	0.25
		(0.17)	(0.22)	(0.15)	(0.32)	(0.15)	(0.40)	(0.08)	(0.30)
fygt5	DLV	1.24	– 0.03	1.13	0.37	1.01	0.48	1.06	0.27
		(0.18)	(0.24)	(0.24)	(0.21)	(0.16)	(0.38)	(0.09)	(0.29)
fygt10	DLV	1.23	0.19	1.11	0.41	1.02	0.45	1.06	0.26
		(0.21)	(0.25)	(0.25)	(0.19)	(0.15)	(0.36)	(0.09)	(0.29)
fyaaac	DLV	1.26	0.26	1.26	0.34	1.14	0.32	1.07	0.25
		(0.22)	(0.17)	(0.39)	(0.20)	(0.19)	(0.19)	(0.10)	(0.29)
fybaac	DLV	1.12	0.40	1.23	0.36	1.15	0.33	1.08	0.34
		(0.18)	(0.14)	(0.38)	(0.18)	(0.18)	(0.17)	(0.12)	(0.20)
fyfha	DLV	1.31	0.19	1.26	0.30	1.02	0.45	1.07	0.26
		(0.24)	(0.20)	(0.29)	(0.16)	(0.16)	(0.37)	(0.09)	(0.29)
sp_fyff	LV	1.21	0.00	1.11	0.31	1.04	0.41	1.17	0.02
		(0.18)	(0.29)	(0.18)	(0.27)	(0.19)	(0.46)	(0.11)	(0.21)
sp_fycp	LV	1.17	0.12	1.09	0.38	0.99	0.52	1.11	0.25
		(0.15)	(0.26)	(0.21)	(0.24)	(0.14)	(0.39)	(0.13)	(0.26)
sp_fygm6	LV	1.14	0.37	1.16	0.34	1.06	0.43	1.19	0.15
		(0.21)	(0.17)	(0.26)	(0.20)	(0.16)	(0.18)	(0.17)	(0.23)
sp_fygt1	LV	1.40	– 0.13	0.97	0.55	1.06	0.38	1.07	0.28
		(0.29)	(0.18)	(0.18)	(0.29)	(0.15)	(0.28)	(0.10)	(0.30)
sp_fygt5	LV	1.08	0.42	1.62	0.18	1.25	0.25	1.44	0.12
		(0.12)	(0.11)	(0.73)	(0.19)	(0.21)	(0.16)	(0.41)	(0.20)
sp_fygt10	LV	1.10	0.39	1.68	0.14	1.23	0.24	1.51	0.05
		(0.15)	(0.15)	(0.73)	(0.19)	(0.20)	(0.17)	(0.40)	(0.20)
sp_fyaaac	LV	1.10	0.37	1.54	0.10	1.21	0.24	1.39	0.05
		(0.15)	(0.18)	(0.45)	(0.20)	(0.21)	(0.19)	(0.28)	(0.23)
sp_fybaac	LV	1.18	0.30	1.32	0.05	1.29	0.15	1.12	0.07
		(0.21)	(0.20)	(0.26)	(0.18)	(0.26)	(0.19)	(0.07)	(0.19)
sp_fyfha	LV	1.22	0.27	1.30	0.22	1.29	0.16	1.11	0.18
		(0.22)	(0.19)	(0.28)	(0.18)	(0.26)	(0.18)	(0.10)	(0.25)

Table 3. (continued)

		PUNEW				GMDC			
		1970–1983		1984–1996		1970–1983		1984–1996	
Variables	Trans	Rel. MSE	λ	Rel. MSE	λ	Rel. MSE	λ	Rel. MSE	λ
Nominal money									
fm1	DLN	1.25	0.11	1.08	0.42	1.06	0.38	1.05	0.37
		(0.19)	(0.20)	(0.26)	(0.23)	(0.17)	(0.32)	(0.10)	(0.24)
fm2	DLN	1.29	− 0.01	0.97	0.53	1.05	0.39	0.98	0.54
		(0.19)	(0.23)	(0.13)	(0.17)	(0.16)	(0.34)	(0.08)	(0.21)
fm3	DLN	1.27	− 0.07	1.00	0.50	1.03	0.43	1.01	0.49
		(0.20)	(0.25)	(0.12)	(0.17)	(0.15)	(0.35)	(0.08)	(0.19)
fml	DLN	1.28	0.05	1.12	0.35	1.06	0.38	1.06	0.37
		(0.26)	(0.26)	(0.14)	(0.14)	(0.18)	(0.35)	(0.09)	(0.19)
fmfba	DLN	1.27	− 0.03	1.11	0.33	1.04	0.43	1.13	0.12
		(0.21)	(0.26)	(0.27)	(0.35)	(0.18)	(0.35)	(0.16)	(0.36)
fmbase	DLN	1.36	− 0.18	1.05	0.42	1.11	0.29	1.08	0.23
		(0.23)	(0.23)	(0.19)	(0.31)	(0.18)	(0.33)	(0.11)	(0.30)
fmrra	DLN	1.28	− 0.14	0.99	0.51	1.00	0.51	1.06	0.31
		(0.18)	(0.26)	(0.17)	(0.27)	(0.16)	(0.39)	(0.10)	(0.27)
fmrnba	DLN	1.26	− 0.11	1.07	0.37	1.01	0.47	1.07	0.24
		(0.18)	(0.26)	(0.16)	(0.27)	(0.15)	(0.38)	(0.09)	(0.28)
fmrnbc	DLN	1.25	− 0.12	1.04	0.43	1.00	0.49	1.07	0.24
		(0.18)	(0.25)	(0.16)	(0.29)	(0.15)	(0.39)	(0.09)	(0.28)
fm1	DDLN	1.26	− 0.12	0.98	0.53	1.00	0.50	1.06	0.28
		(0.18)	(0.25)	(0.16)	(0.33)	(0.15)	(0.39)	(0.09)	(0.29)
fm2	DDLN	1.26	− 0.15	0.99	0.53	1.00	0.50	1.07	0.26
		(0.19)	(0.25)	(0.16)	(0.32)	(0.15)	(0.39)	(0.09)	(0.29)
fm3	DDLN	1.26	− 0.14	0.98	0.53	1.00	0.49	1.06	0.27
		(0.19)	(0.25)	(0.15)	(0.33)	(0.16)	(0.39)	(0.09)	(0.29)
fml	DDLN	1.26	− 0.13	0.99	0.53	1.00	0.50	1.06	0.27
		(0.19)	(0.25)	(0.16)	(0.33)	(0.16)	(0.39)	(0.09)	(0.30)
fmfba	DDLN	1.25	− 0.10	0.99	0.53	0.99	0.51	1.06	0.29
		(0.18)	(0.25)	(0.16)	(0.32)	(0.16)	(0.39)	(0.09)	(0.29)
fmbase	DDLN	1.26	− 0.13	0.98	0.53	1.00	0.50	1.06	0.28
		(0.19)	(0.25)	(0.16)	(0.32)	(0.16)	(0.39)	(0.09)	(0.29)
fmrra	DDLN	1.26	− 0.12	0.98	0.54	1.00	0.51	1.06	0.30
		(0.18)	(0.25)	(0.16)	(0.32)	(0.16)	(0.39)	(0.09)	(0.29)
fmrnba	DDLN	1.26	− 0.14	0.99	0.53	0.99	0.51	1.06	0.27
		(0.19)	(0.25)	(0.16)	(0.33)	(0.16)	(0.39)	(0.09)	(0.29)
fmrnbc	DDLN	1.26	− 0.14	0.98	0.54	0.99	0.52	1.06	0.27
		(0.19)	(0.25)	(0.16)	(0.33)	(0.16)	(0.39)	(0.09)	(0.30)
Exchange rates									
exrus	DLN	1.33	0.24	1.94	0.19	1.32	0.24	1.66	0.12
		(0.36)	(0.13)	(0.18)	(0.18)	(0.37)	(0.16)	(0.69)	(0.21)
exrger	DLN	1.32	− 0.12	1.38	0.24	0.99	0.52	1.62	0.05
		(0.22)	(0.24)	(0.54)	(0.24)	(0.12)	(0.20)	(0.60)	(0.23)

Table 3. (continued)

		PUNEW				GMDC			
		1970–1983		1984–1996		1970–1983		1984–1996	
Variables	Trans	Rel. MSE	λ	Rel. MSE	λ	Rel. MSE	λ	Rel. MSE	λ
exrsw	DLN	1.32	−0.07	1.31	0.26	1.62	−0.12	1.39	0.03
		(0.22)	(0.22)	(0.50)	(0.27)	(0.71)	(0.21)	(0.39)	(0.28)
exrjan	DLN	1.42	0.30	1.49	0.30	1.49	0.26	1.14	0.19
		(0.33)	(0.08)	(0.50)	(0.15)	(0.34)	(0.09)	(0.16)	(0.26)
exruk	DLN	1.27	−0.15	1.01	0.47	1.04	0.39	1.08	0.22
		(0.19)	(0.25)	(0.17)	(0.32)	(0.13)	(0.36)	(0.10)	(0.30)
exrcan	DLN	1.28	−0.20	0.98	0.54	1.01	0.48	1.06	0.31
		(0.18)	(0.25)	(0.16)	(0.33)	(0.15)	(0.38)	(0.09)	(0.28)
Prices and wages									
pmcp	LV	1.25	−0.16	1.08	0.39	1.06	0.33	1.09	0.20
		(0.18)	(0.31)	(0.20)	(0.26)	(0.14)	(0.39)	(0.09)	(0.28)
pwfsa	DDLN	1.26	−0.11	0.97	0.56	1.00	0.51	1.05	0.31
		(0.18)	(0.25)	(0.15)	(0.32)	(0.15)	(0.38)	(0.09)	(0.28)
pwfcsa	DDLN	1.25	−0.11	0.98	0.55	0.99	0.53	1.05	0.32
		(0.18)	(0.25)	(0.15)	(0.32)	(0.16)	(0.38)	(0.09)	(0.28)
pwimsa	DDLN	1.26	−0.12	0.98	0.54	1.00	0.50	1.06	0.28
		(0.19)	(0.25)	(0.15)	(0.32)	(0.16)	(0.39)	(0.09)	(0.29)
pwcmsa	DDLN	1.26	−0.12	0.98	0.54	1.04	0.41	1.06	0.29
		(0.18)	(0.25)	(0.15)	(0.32)	(0.18)	(0.36)	(0.09)	(0.29)
psm99q	DDLN	1.37	−0.24	1.27	0.24	1.02	0.46	1.06	0.28
		(0.23)	(0.22)	(0.28)	(0.21)	(0.15)	(0.37)	(0.09)	(0.29)
punew	DDLN	1.26	−0.13	0.98	0.53	1.01	0.48	1.06	0.29
		(0.19)	(0.25)	(0.15)	(0.33)	(0.15)	(0.38)	(0.09)	(0.29)
pu83	DDLN	1.26	−0.13	0.99	0.51	1.00	0.49	1.07	0.26
		(0.19)	(0.25)	(0.16)	(0.32)	(0.16)	(0.38)	(0.09)	(0.29)
pu84	DDLN	1.26	−0.13	0.98	0.54	1.00	0.50	1.06	0.30
		(0.19)	(0.25)	(0.15)	(0.32)	(0.15)	(0.39)	(0.09)	(0.27)
puc	DDLN	1.26	−0.13	0.98	0.54	1.00	0.49	1.05	0.31
		(0.19)	(0.25)	(0.15)	(0.32)	(0.15)	(0.39)	(0.09)	(0.29)
pucd	DDLN	1.24	−0.08	0.99	0.52	1.00	0.49	1.06	0.29
		(0.18)	(0.26)	(0.16)	(0.32)	(0.15)	(0.38)	(0.09)	(0.29)
pus	DDLN	1.26	−0.13	0.99	0.53	1.00	0.51	1.06	0.27
		(0.19)	(0.25)	(0.16)	(0.33)	(0.16)	(0.39)	(0.09)	(0.29)
puxf	DDLN	1.26	−0.12	0.98	0.54	1.00	0.50	1.06	0.28
		(0.18)	(0.25)	(0.15)	(0.32)	(0.15)	(0.39)	(0.09)	(0.29)
puxhs	DDLN	1.26	−0.13	0.98	0.54	1.00	0.49	1.06	0.27
		(0.19)	(0.25)	(0.15)	(0.33)	(0.15)	(0.38)	(0.09)	(0.29)
puxm	DDLN	1.25	−0.12	0.98	0.54	1.00	0.50	1.06	0.30
		(0.18)	(0.26)	(0.15)	(0.32)	(0.15)	(0.39)	(0.09)	(0.29)
gmdc	DDLN	1.26	−0.12	0.99	0.53	1.00	0.50	1.06	0.27
		(0.19)	(0.24)	(0.15)	(0.33)	(0.15)	(0.38)	(0.09)	(0.29)

Table 3. (continued)

		PUNEW				GMDC			
		1970–1983		1984–1996		1970–1983		1984–1996	
Variables	Trans	Rel. MSE	λ	Rel. MSE	λ	Rel. MSE	λ	Rel. MSE	λ
gmdcd	DDLN	1.26	− 0.12	0.99	0.53	1.00	0.49	1.06	0.28
		(0.19)	(0.25)	(0.16)	(0.33)	(0.15)	(0.38)	(0.09)	(0.29)
gmdcn	DDLN	1.27	− 0.12	0.98	0.54	1.00	0.49	1.04	0.35
		(0.19)	(0.24)	(0.15)	(0.32)	(0.15)	(0.38)	(0.08)	(0.27)
gmdcs	DDLN	1.26	− 0.14	0.98	0.53	0.99	0.51	1.05	0.33
		(0.19)	(0.25)	(0.16)	(0.32)	(0.16)	(0.39)	(0.09)	(0.29)
lehcc	DDLN	1.26	− 0.13	0.99	0.53	1.00	0.49	1.06	0.28
		(0.19)	(0.25)	(0.15)	(0.32)	(0.15)	(0.38)	(0.09)	(0.29)
lehm	DDLN	1.26	− 0.13	0.98	0.53	1.00	0.50	1.06	0.27
		(0.19)	(0.25)	(0.15)	(0.33)	(0.15)	(0.38)	(0.09)	(0.29)
Output									
ipp	DLN	0.97	0.67	1.14	0.19	0.85	1.13	1.15	0.20
		(0.06)	(0.30)	(0.15)	(0.24)	(0.16)	(0.53)	(0.11)	(0.17)
ipf	DLN	1.03	0.34	1.07	0.33	0.90	0.96	1.17	0.12
		(0.08)	(0.36)	(0.12)	(0.23)	(0.15)	(0.57)	(0.12)	(0.19)
ipc	DLN	1.10	0.20	1.00	0.50	0.92	0.76	1.13	0.22
		(0.12)	(0.28)	(0.10)	(0.19)	(0.16)	(0.49)	(0.09)	(0.17)
ipcd	DLN	1.21	− 0.10	1.10	0.31	1.02	0.45	1.12	0.14
		(0.16)	(0.31)	(0.12)	(0.22)	(0.15)	(0.41)	(0.09)	(0.23)
ipcn	DLN	1.16	0.06	0.97	0.53	0.99	0.52	1.09	0.37
		(0.15)	(0.26)	(0.15)	(0.17)	(0.15)	(0.45)	(0.13)	(0.18)
ipe	DLN	1.00	0.48	1.11	0.27	0.87	1.05	1.17	− 0.02
		(0.07)	(0.42)	(0.12)	(0.18)	(0.18)	(0.60)	(0.11)	(0.20)
ipi	DLN	0.84	1.05	1.12	0.33	0.84	0.96	1.11	0.31
		(0.07)	(0.20)	(0.12)	(0.14)	(0.17)	(0.42)	(0.10)	(0.16)
ipm	DLN	1.10	0.09	1.04	0.40	0.98	0.61	1.14	0.00
		(0.10)	(0.39)	(0.12)	(0.26)	(0.14)	(0.62)	(0.09)	(0.20)
ipmd	DLN	1.12	0.07	1.03	0.42	1.00	0.48	1.13	0.10
		(0.12)	(0.39)	(0.09)	(0.23)	(0.15)	(0.62)	(0.09)	(0.18)
ipmnd	DLN	1.15	− 0.03	1.03	0.45	1.01	0.48	1.08	0.31
		(0.09)	(0.27)	(0.12)	(0.17)	(0.15)	(0.52)	(0.07)	(0.17)
ipmfg	DLN	0.96	0.76	1.11	0.21	0.87	1.17	1.11	0.19
		(0.05)	(0.36)	(0.13)	(0.25)	(0.15)	(0.61)	(0.08)	(0.18)
ipd	DLN	0.99	0.56	1.14	0.18	0.91	0.97	1.18	0.05
		(0.07)	(0.39)	(0.15)	(0.26)	(0.15)	(0.66)	(0.10)	(0.16)
ipn	DLN	1.08	0.21	1.07	0.41	0.91	0.80	1.08	0.35
		(0.10)	(0.29)	(0.11)	(0.13)	(0.16)	(0.50)	(0.08)	(0.14)
ipmin	DLN	1.25	− 0.08	1.00	0.51	1.01	0.48	1.06	0.28
		(0.18)	(0.25)	(0.16)	(0.33)	(0.15)	(0.37)	(0.09)	(0.29)
iput	DLN	1.28	− 0.23	1.00	0.50	1.03	0.41	1.05	0.33
		(0.17)	(0.24)	(0.15)	(0.25)	(0.15)	(0.39)	(0.09)	(0.25)

Table 3. (continued)

Variables	Trans	PUNEW 1970–1983 Rel. MSE	PUNEW 1970–1983 λ	PUNEW 1984–1996 Rel. MSE	PUNEW 1984–1996 λ	GMDC 1970–1983 Rel. MSE	GMDC 1970–1983 λ	GMDC 1984–1996 Rel. MSE	GMDC 1984–1996 λ
pmi	LV	0.74	1.73	0.96	0.56	0.73	1.51	0.99	0.53
		(0.08)	(0.29)	(0.15)	(0.24)	(0.21)	(0.46)	(0.13)	(0.28)
pmp	LV	0.89	0.89	1.01	0.49	0.80	1.18	1.15	0.29
		(0.07)	(0.26)	(0.13)	(0.21)	(0.20)	(0.50)	(0.15)	(0.19)
gmyxpq	DLN	0.84	0.86	1.42	0.08	0.73	1.20	1.27	0.25
		(0.07)	(0.23)	(0.38)	(0.17)	(0.18)	(0.27)	(0.22)	(0.13)
Employment									
lhel	DLN	0.79	1.37	1.18	0.28	0.79	1.47	1.19	0.27
		(0.08)	(0.24)	(0.23)	(0.21)	(0.18)	(0.54)	(0.16)	(0.16)
lhelx	LN	0.84	2.49	0.85	1.21	0.90	2.17	0.88	1.10
		(0.05)	(0.39)	(0.10)	(0.29)	(0.07)	(0.85)	(0.08)	(0.32)
lhem	DLN	0.98	0.57	1.11	0.28	0.94	0.69	1.02	0.45
		(0.08)	(0.32)	(0.19)	(0.31)	(0.16)	(0.53)	(0.09)	(0.20)
lhnag	DLN	0.97	0.63	1.11	0.29	0.92	0.80	1.00	0.50
		(0.08)	(0.37)	(0.18)	(0.30)	(0.16)	(0.57)	(0.12)	(0.21)
lp	DLN	0.89	1.03	0.96	0.63	0.88	1.06	0.96	0.69
		(0.07)	(0.27)	(0.08)	(0.29)	(0.14)	(0.50)	(0.06)	(0.29)
lpgd	DLN	0.92	0.89	0.94	0.72	0.85	1.14	0.98	0.57
		(0.06)	(0.31)	(0.08)	(0.34)	(0.17)	(0.58)	(0.08)	(0.28)
lpmi	DLN	1.25	−0.10	0.99	0.53	1.01	0.48	1.06	0.28
		(0.19)	(0.25)	(0.16)	(0.32)	(0.16)	(0.38)	(0.09)	(0.29)
lpcc	DLN	1.05	0.34	1.24	0.06	0.98	0.58	1.00	0.50
		(0.13)	(0.35)	(0.31)	(0.30)	(0.16)	(0.51)	(0.10)	(0.26)
lpem	DLN	0.94	0.95	0.89	0.86	0.90	1.15	0.97	0.66
		(0.06)	(0.46)	(0.07)	(0.27)	(0.13)	(0.63)	(0.05)	(0.25)
lped	DLN	0.93	1.00	0.92	0.75	0.88	1.17	0.98	0.60
		(0.07)	(0.50)	(0.08)	(0.31)	(0.14)	(0.57)	(0.06)	(0.30)
lpen	DLN	1.06	0.17	0.92	0.67	0.93	0.83	0.95	0.64
		(0.08)	(0.35)	(0.06)	(0.14)	(0.12)	(0.56)	(0.08)	(0.21)
lpsp	DLN	1.00	0.51	0.98	0.53	0.99	0.57	0.97	0.58
		(0.11)	(0.41)	(0.13)	(0.25)	(0.07)	(0.36)	(0.11)	(0.30)
lptu	DLN	1.32	0.00	1.03	0.42	1.13	0.30	1.07	0.19
		(0.24)	(0.14)	(0.15)	(0.33)	(0.20)	(0.22)	(0.07)	(0.28)
lpt	DLN	1.01	0.42	1.12	0.32	0.94	0.84	1.22	0.18
		(0.09)	(0.47)	(0.19)	(0.22)	(0.11)	(0.58)	(0.21)	(0.19)
lpfr	DLN	1.03	0.41	1.51	0.25	0.92	0.74	1.10	0.40
		(0.13)	(0.35)	(0.42)	(0.14)	(0.15)	(0.43)	(0.21)	(0.20)
lps	DLN	1.17	0.02	1.01	0.47	1.10	0.15	1.04	0.34
		(0.16)	(0.31)	(0.10)	(0.30)	(0.10)	(0.24)	(0.07)	(0.28)
lpgov	DLN	1.27	−0.16	0.97	0.56	1.01	0.47	1.05	0.34
		(0.19)	(0.25)	(0.14)	(0.27)	(0.15)	(0.39)	(0.08)	(0.23)

Table 3. (continued)

		PUNEW				GMDC			
		1970–1983		1984–1996		1970–1983		1984–1996	
Variables	Trans	Rel. MSE	λ	Rel. MSE	λ	Rel. MSE	λ	Rel. MSE	λ
lphrm	LV	1.08	0.17	1.44	0.06	0.99	0.61	1.43	− 0.11
		(0.09)	(0.32)	(0.29)	(0.23)	(0.04)	(0.28)	(0.23)	(0.21)
lpmosa	LV	1.02	0.44	1.70	0.02	0.91	0.95	1.63	− 0.06
		(0.10)	(0.33)	(0.41)	(0.18)	(0.11)	(0.57)	(0.36)	(0.17)
pmemp	LV	0.80	1.39	1.04	0.45	0.79	1.47	1.02	0.46
		(0.07)	(0.35)	(0.23)	(0.28)	(0.17)	(0.56)	(0.15)	(0.30)
luinc	LV	1.08	0.26	0.94	0.62	1.01	0.44	0.93	0.87
		(0.09)	(0.28)	(0.11)	(0.24)	(0.06)	(0.31)	(0.07)	(0.29)
lhu680	LV	1.28	− 0.19	1.03	0.44	1.00	0.50	1.07	0.25
		(0.18)	(0.23)	(0.13)	(0.28)	(0.16)	(0.41)	(0.08)	(0.27)
lhu5	LV	1.49	− 0.23	1.14	0.30	1.38	0.24	1.14	0.18
		(0.36)	(0.28)	(0.20)	(0.22)	(0.31)	(0.24)	(0.13)	(0.22)
lhu14	LV	1.18	− 0.36	1.06	0.38	1.25	−0.46	1.06	0.34
		(0.13)	(0.48)	(0.14)	(0.26)	(0.21)	(0.37)	(0.10)	(0.26)
lhu15	LV	1.22	− 0.34	1.10	0.28	1.08	0.16	1.07	0.28
		(0.11)	(0.22)	(0.11)	(0.23)	(0.07)	(0.22)	(0.07)	(0.18)
lhu26	LV	1.37	− 0.29	1.00	0.50	1.20	− 0.12	1.07	0.34
		(0.22)	(0.18)	(0.11)	(0.19)	(0.13)	(0.21)	(0.09)	(0.22)
lhu27	LV	1.28	− 0.30	1.15	0.15	1.08	0.26	1.11	0.15
		(0.20)	(0.31)	(0.12)	(0.22)	(0.10)	(0.26)	(0.08)	(0.23)
Real retail, manufacturing and trade sales									
msmq	DLN	0.94	0.83	1.09	0.35	0.95	0.71	1.14	0.14
		(0.06)	(0.31)	(0.11)	(0.16)	(0.15)	(0.55)	(0.09)	(0.15)
msdq	DLN	1.01	0.48	1.07	0.37	0.98	0.58	1.04	0.38
		(0.08)	(0.34)	(0.11)	(0.17)	(0.15)	(0.51)	(0.06)	(0.17)
msnq	DLN	1.05	0.34	0.92	0.65	0.97	0.60	1.10	0.22
		(0.10)	(0.24)	(0.13)	(0.24)	(0.15)	(0.43)	(0.07)	(0.17)
wtq	DLN	0.98	0.56	0.82	0.83	0.88	0.82	0.86	0.87
		(0.09)	(0.31)	(0.16)	(0.25)	(0.17)	(0.34)	(0.13)	(0.27)
wtdq	DLN	0.91	0.75	0.98	0.54	0.89	0.83	1.00	0.51
		(0.07)	(0.16)	(0.10)	(0.20)	(0.17)	(0.43)	(0.08)	(0.24)
wtnq	DLN	1.22	− 0.09	0.78	0.88	0.99	0.52	0.92	0.70
		(0.17)	(0.26)	(0.13)	(0.16)	(0.15)	(0.37)	(0.10)	(0.22)
rtq	DLN	1.02	0.44	1.00	0.49	0.92	0.69	1.12	0.28
		(0.11)	(0.28)	(0.13)	(0.19)	(0.16)	(0.34)	(0.11)	(0.18)
rtnq	DLN	1.18	0.02	0.86	0.76	1.00	0.49	0.94	0.63
		(0.15)	(0.29)	(0.11)	(0.17)	(0.13)	(0.35)	(0.10)	(0.24)
Consumption									
gmcq	DLN	1.05	0.38	0.96	0.55	0.89	0.76	1.15	0.28
		(0.09)	(0.23)	(0.09)	(0.13)	(0.15)	(0.35)	(0.15)	(0.20)

Table 3. (continued)

		PUNEW				GMDC			
		1970–1983		1984–1996		1970–1983		1984–1996	
Variables	Trans	Rel. MSE	λ	Rel. MSE	λ	Rel. MSE	λ	Rel. MSE	λ
gmcdq	DLN	1.21	0.00	0.97	0.54	1.01	0.48	1.13	0.14
		(0.15)	(0.26)	(0.11)	(0.17)	(0.15)	(0.40)	(0.11)	(0.22)
gmcnq	DLN	1.12	0.17	0.87	0.71	0.94	0.69	0.99	0.53
		(0.12)	(0.24)	(0.08)	(0.11)	(0.14)	(0.40)	(0.12)	(0.26)
gmcsq	DLN	1.08	0.32	1.36	0.21	1.05	0.35	1.21	0.19
		(0.14)	(0.27)	(0.41)	(0.19)	(0.13)	(0.40)	(0.20)	(0.23)
gmcanq	DLN	1.24	− 0.07	0.99	0.51	0.99	0.52	1.06	0.31
		(0.17)	(0.25)	(0.15)	(0.31)	(0.16)	(0.38)	(0.09)	(0.27)
Housing									
hsfr	LN	0.84	0.93	1.07	0.43	0.87	0.87	0.95	0.58
		(0.11)	(0.26)	(0.22)	(0.22)	(0.15)	(0.38)	(0.16)	(0.25)
hsne	LN	1.01	0.47	1.19	0.40	0.83	0.92	1.15	0.39
		(0.13)	(0.25)	(0.29)	(0.15)	(0.19)	(0.43)	(0.24)	(0.16)
hsmw	LN	0.96	0.55	0.73	0.85	0.88	0.76	0.92	0.69
		(0.14)	(0.19)	(0.16)	(0.17)	(0.18)	(0.36)	(0.10)	(0.25)
hssou	LN	1.12	0.27	1.15	0.26	1.05	0.40	1.05	0.37
		(0.13)	(0.24)	(0.20)	(0.27)	(0.16)	(0.32)	(0.12)	(0.28)
hswst	LN	1.05	0.33	0.97	0.56	1.03	0.39	0.97	0.59
		(0.12)	(0.35)	(0.19)	(0.37)	(0.10)	(0.32)	(0.12)	(0.37)
hsbp	LN	0.85	0.94	1.03	0.47	0.89	0.81	0.90	0.65
		(0.10)	(0.26)	(0.24)	(0.23)	(0.15)	(0.37)	(0.17)	(0.26)
dhsbp	DLN	1.28	− 0.05	1.05	0.42	1.03	0.43	1.05	0.31
		(0.19)	(0.26)	(0.16)	(0.23)	(0.16)	(0.35)	(0.09)	(0.28)
hsbr	LN	0.85	0.94	1.03	0.47	0.89	0.81	0.90	0.65
		(0.10)	(0.26)	(0.24)	(0.23)	(0.15)	(0.37)	(0.17)	(0.26)
hmob	LN	1.33	− 0.24	1.10	0.33	1.07	0.34	1.04	0.34
		(0.18)	(0.29)	(0.19)	(0.28)	(0.16)	(0.36)	(0.06)	(0.25)
condo9	LN	1.23	0.03	1.25	0.16	1.05	0.40	1.02	0.42
		(0.21)	(0.29)	(0.20)	(0.16)	(0.21)	(0.39)	(0.09)	(0.30)
Inventories and orders									
ivmtq	DLN	1.22	− 0.18	1.06	0.37	0.97	0.58	1.09	0.22
		(0.16)	(0.26)	(0.12)	(0.22)	(0.13)	(0.43)	(0.07)	(0.21)
ivmfgq	DLN	1.27	− 0.17	0.99	0.51	1.06	0.31	1.02	0.42
		(0.19)	(0.25)	(0.13)	(0.29)	(0.11)	(0.31)	(0.07)	(0.25)
ivmfdq	DLN	1.23	− 0.10	0.98	0.54	1.03	0.41	1.03	0.40
		(0.17)	(0.26)	(0.12)	(0.28)	(0.11)	(0.34)	(0.07)	(0.22)
ivmfnq	DLN	1.26	− 0.05	1.01	0.48	1.01	0.48	1.09	0.21
		(0.20)	(0.24)	(0.17)	(0.32)	(0.15)	(0.35)	(0.11)	(0.28)
ivwrq	DLN	1.24	− 0.11	0.98	0.54	1.00	0.49	1.05	0.34
		(0.18)	(0.24)	(0.15)	(0.30)	(0.15)	(0.39)	(0.08)	(0.27)

Table 3. (continued)

Variables	Trans	PUNEW 1970–1983 Rel. MSE	PUNEW 1970–1983 λ	PUNEW 1984–1996 Rel. MSE	PUNEW 1984–1996 λ	GMDC 1970–1983 Rel. MSE	GMDC 1970–1983 λ	GMDC 1984–1996 Rel. MSE	GMDC 1984–1996 λ
ivrrq	DLN	1.22	– 0.11	1.14	0.20	0.99	0.52	1.11	0.14
		(0.17)	(0.28)	(0.17)	(0.28)	(0.16)	(0.42)	(0.09)	(0.24)
ivsrq	DLV	1.03	0.40	1.07	0.41	0.99	0.52	1.07	0.33
		(0.09)	(0.26)	(0.11)	(0.15)	(0.16)	(0.39)	(0.08)	(0.17)
ivsrmq	DLV	1.11	0.22	1.17	0.26	0.98	0.55	1.17	0.04
		(0.14)	(0.30)	(0.16)	(0.18)	(0.16)	(0.39)	(0.12)	(0.23)
ivsrwq	DLV	1.10	0.29	0.91	0.65	0.98	0.55	1.02	0.46
		(0.10)	(0.25)	(0.09)	(0.14)	(0.18)	(0.37)	(0.08)	(0.21)
ivsrrq	DLV	1.26	– 0.12	0.98	0.54	1.01	0.48	1.05	0.33
		(0.18)	(0.25)	(0.15)	(0.31)	(0.13)	(0.33)	(0.08)	(0.27)
pmnv	LV	1.02	0.43	0.90	0.66	0.87	1.05	0.94	0.64
		(0.10)	(0.36)	(0.11)	(0.19)	(0.17)	(0.57)	(0.09)	(0.21)
pmno	LV	0.85	1.30	1.01	0.49	0.80	1.40	1.08	0.39
		(0.06)	(0.32)	(0.14)	(0.19)	(0.19)	(0.57)	(0.16)	(0.21)
mocmq	DLN	1.05	0.34	1.07	0.38	0.94	0.68	1.13	0.17
		(0.10)	(0.30)	(0.09)	(0.14)	(0.15)	(0.46)	(0.08)	(0.15)
mdoq	DLN	0.91	0.92	1.15	0.27	0.95	0.72	1.08	0.30
		(0.06)	(0.21)	(0.17)	(0.16)	(0.15)	(0.54)	(0.07)	(0.13)
msondq	DLN	1.08	0.27	1.28	0.03	0.98	0.56	1.14	0.11
		(0.13)	(0.31)	(0.27)	(0.26)	(0.15)	(0.44)	(0.11)	(0.20)
mpconq	DLN	1.08	0.26	1.32	0.02	1.00	0.51	1.13	0.04
		(0.13)	(0.32)	(0.31)	(0.25)	(0.15)	(0.40)	(0.10)	(0.27)
Stock prices									
fsncom	DLN	1.24	– 0.03	1.14	0.31	1.02	0.45	1.07	0.27
		(0.18)	(0.23)	(0.22)	(0.25)	(0.16)	(0.35)	(0.09)	(0.27)
fspcom	DLN	1.24	-0.04	1.17	0.27	1.02	0.46	1.07	0.27
		(0.18)	(0.24)	(0.23)	(0.24)	(0.16)	(0.36)	(0.09)	(0.26)
fspin	DLN	1.24	– 0.03	1.14	0.31	1.02	0.45	1.07	0.27
		(0.18)	(0.24)	(0.22)	(0.25)	(0.16)	(0.35)	(0.09)	(0.26)
fspcap	DLN	1.23	0.00	1.23	0.22	1.04	0.42	1.08	0.24
		(0.17)	(0.23)	(0.25)	(0.22)	(0.16)	(0.35)	(0.09)	(0.25)
fsput	DLN	1.26	– 0.12	0.99	0.52	1.02	0.46	1.06	0.27
		(0.19)	(0.25)	(0.16)	(0.33)	(0.15)	(0.37)	(0.09)	(0.29)
fsdxp	LV	1.55	0.09	1.04	0.44	1.20	0.32	1.16	0.16
		(0.68)	(0.25)	(0.19)	(0.26)	(0.41)	(0.31)	(0.13)	(0.23)
fspxe	LV	1.33	– 0.08	1.18	0.33	1.03	0.42	1.28	0.12
		(0.22)	(0.25)	(0.24)	(0.19)	(0.17)	(0.38)	(0.26)	(0.19)
Other variables									
fm2dq	DLN	1.22	0.13	0.91	0.58	0.99	0.52	0.89	0.68
		(0.17)	(0.24)	(0.14)	(0.13)	(0.15)	(0.32)	(0.11)	(0.17)

Table 3. (continued)

		PUNEW				GMDC			
		1970–1983		1984–1996		1970–1983		1984–1996	
Variables	Trans	Rel. MSE	λ	Rel. MSE	λ	Rel. MSE	λ	Rel. MSE	λ
fclnq	DLN	1.30	− 0.13	0.99	0.51	1.11	0.21	1.06	0.32
		(0.20)	(0.22)	(0.14)	(0.33)	(0.11)	(0.26)	(0.07)	(0.20)
fclbmc	LV	1.36	− 0.13	2.79	0.07	1.21	0.15	1.16	0.15
		(0.25)	(0.26)	(3.96)	(0.11)	(0.29)	(0.35)	(0.16)	(0.22)
hhsntn	LV	1.52	− 0.47	0.91	0.74	1.27	− 0.10	0.97	0.64
		(0.33)	(0.27)	(0.08)	(0.20)	(0.18)	(0.21)	(0.07)	(0.29)
pmdel	LV	1.02	0.45	0.89	0.76	0.91	0.78	0.88	0.77
		(0.09)	(0.25)	(0.11)	(0.27)	(0.14)	(0.41)	(0.10)	(0.23)

Notes: See notes to Table 2.

prices, money, commodity prices, etc.) do not perform this well. Moreover, many of these indicators appear to have unstable forecasting relations. For example, the National Association of Purchasing Managers' new orders index (*pmno*) significantly outperforms the unemployment rate during the first subsample, but has a relative MSE exceeding one during the second subsample for both CPI and PCE inflation; the reverse is true for the index of consumer expectations (*hhsntn*).

5. Multivariate inflation forecasts using leading indicators

In this section we move beyond the bivariate models of Sections 3 and 4 to compare the benchmark Phillips curve to forecasts constructed using multiple predictors. Moving from bivariate to multivariate models raises the important problem of parsimony. On a priori grounds, many of the variables listed in Table 3 would be expected to provide useful information about future inflation. However, including more than just a few of these variables in unrestricted regressions like (3) would result in overfitting and poor forecast performance. One approach is to estimate a large number of relatively simple models (say, all possible models that use no more than three leading indicators) and then use a model selection criterion to choose one of these for forecasting. However, the large number of possible models makes this statistically suspect: serious overfitting would likely spoil the resulting forecasts.

In this section we therefore consider two alternative approaches for constructing multivariate forecasts. The first approach is to treat as data the bivariate forecasts constructed in Section 4 and to combine them using various forecast combination procedures that are designed to handle large numbers of dependent forecasts. The second approach is to construct a small number of composite indexes from larger groups of variables, using methods in dynamic factor analysis, and then to use these indexes (estimated factors) to construct small multiviarate forecasting models.

The forecast combination methods begin with the forecasting models (3), now written as

$$\pi^h_{t+h} - \pi_t = \phi_i + \beta_i(L)x_{i,t} + \gamma_i(L)\Delta\pi_t + e_{i,t+h}, \tag{7}$$

where a subscript $i = 1, \dots, n$ has been added to denote the model constructed using the leading indicator $x_{i,t}$. Let $f_{i,t} = \hat{\phi}_i + \hat{\beta}_i(L)x_{i,t} + \hat{\gamma}_i(L)\Delta\pi_t$ denote the time 't' forecast implied by this model, where $\hat{\phi}_i$ (etc.) are coefficients estimated using data through date t. The combined forecasts are constructed as

$$f_{c,t} = \sum_{i=1}^{n} \omega_{it} f_{i,t}. \tag{8}$$

Three different procedures are used for choosing the weights $\{\omega_{it}\}$. The first sets $\omega_i = 1/n$, so that $f_{c,t}$ is the sample mean of the date t forecasts. The second uses the sample median instead of the mean. In the third, the weights are determined from the regression

$$\pi^h_{s+h} - \pi_s = \sum_{i=1}^{n} \omega_{it} f_{i,s} + \varepsilon_{s+h}, \quad s = 1, \dots, t, \tag{9}$$

estimated using data through period t. Because n is large, OLS estimation of Eq. (9) generally produces poor results. However, alternative estimators, better suited to the problem at hand, are available. In particular, if the forecasts have an approximate dynamic factor structure, then the problem of minimizing out of sample forecast error from the forecast combining regression (9) has similarities to the problem that leads to James and Stein (1961) estimation and to ridge regression, modified so that they shrink towards equal weighting (this argument is spelled out in Chan et al. (1999)). The third forecast combination procedure therefore is the ridge regression estimator of $\omega_t = (\omega_{1t}\ \omega_{2t}\ \cdots\ \omega_{nt})'$, which can be written as

$$\hat{\omega}_{t,RR} = \left(cI_n + \sum_{s=1}^{t} F_s F_s'\right)^{-1} \left(\sum_{s=1}^{t} F_s(\pi^h_{s+h} - \pi_s) + c/n\right). \tag{10}$$

where $F_s = (f_{1,s} \dots f_{n,s})'$, and $c = k \times TR(n^{-1}\sum_{s=1}^{t} F_s F_s')$. The parameter k governs the amount of shrinkage. When $k = 0$, $\hat{\omega}_{t,RR} = \hat{\omega}_{t,OLS}$ and as k grows large $\hat{\omega}_{t,RR} \to 1/n$. Results were computed for $k = 0.25, 0.5, 1$ and 10; the forecasts constructed using $k = 1$ generally were most accurate, so to save space only results for $k = 1$ are reported. Loosely speaking, $k = 1$ corresponds to shrinking the OLS estimator half way to the equal weighted value of $1/n$.

The second approach to multivariate forecasting in this high-dimensional setting utilizes estimated factors (or indexes) constructed from the set of predictor variables. Let X_t denote the set of predictors at date t. Then the factors are constructed as the principal components of X_t. A theoretical justification for this estimator, provided in Stock and Watson (1998), is that it produces consistent estimates of the factors under fairly general conditions in an approximate factor model when the number of elements in X_t grows large. This approach is applied here mainly when the number of predictors is large (all of the variables in Table 3), although in some cases the number of predictors is moderate to small (e.g. the nine money supply variables in Table 3). The rationale in the latter case is simply that the estimator provides a simple procedure for summarizing the data. Let D_s^t, $s = 1, \ldots, t$, denote the m-vector time series of factors extracted at date t. Then forecasts are constructed from the regression model:

$$\pi_{s+h}^h - \pi_s = \phi + \beta(L)' D_s^t + \gamma(L)\Delta\pi_s + e_{s+h}, \quad s = 1, \ldots, t. \tag{11}$$

The recursive design used in this section parallels the design used in the last two sections. Specifically, at date t, the coefficients in (7) are estimated for each x by OLS using only data through date t. The orders of the lag polynomials $\beta(L)$ and $\gamma(L)$ are determined separately by BIC for each date over orders 0–11. The recursive model selection also allows $\gamma(L) = 0$. With the coefficients of (7) estimated, the forecasts $f_{i,t}$ are formed. For the ridge regression combined forecast, ridge regression estimates of ω are computed using the set of forecasts and inflation data for dates t and earlier.

Similarly, at date t, factors are constructed as principal components using data on the various indicators from dates t and earlier. These estimated factors are then used in regression (11), which is estimated by OLS using data on inflation and the factors dated t and earlier. BIC model selection is recursively carried out over the number of factors and the orders of the lag polynomials. Two factor models are estimated. The first model allows several underlying factors to help forecast inflation, and recursively chooses models with 1–6 factors, each entering with 0–5 lags. The second model uses a single factor (representing, say 'activity' or 'money') and allows from 0–11 lags of the factor to enter (11). Both models allow up to 11 lags in $\gamma(L)$.

Results for four categories of variables are summarized in Table 4. Panel A of the table shows results constructed from all of the variables shown in Table 3 together with the variables in Table 2 and the unemployment rate (*LHUR*). The row labeled 'Mul. factors' shows results from the multiple factor model; the next row shows results from the single factor version of the model. The following three rows show forecasts constructed from the forecast combining Eq. (8), the mean forecast, the median, and the ridge regression combination forecast. Panel B include all of the variables shown in Table 2 (using the first differenced values of the trending variables) and the variables under the categories labeled output, employment (except the unemployment duration variables), sales, consumption, and inventories and orders. The results in Panel C are constructed using all of the interest variables in Table 3 (including the interest rate spreads). The results in Panel D use the variables in the 'Nominal money' section of Table 3 transformed as the second difference of logarithms. Results using the first differences of the money variables are similar to the second difference results and are not reported.[4]

Three conclusions emerge from Table 4. First, single factor models either using all of the indicators or using only the real activity indicators produce the best overall forecasts of inflation. The all-indicates single factor model performs marginally better than real-activity single factor model in the the first sample period; their performance is identical in the second period. These single factor forecasts are significantly better (economically and statistically) than the benchmark unemployment rate Phillips curve model. They also dominate forecasts constructed from any of the bivariate models. For example, no bivariate model has a smaller mean square for both price series in the first sample period than the all-indicators single factor model. In the second sample period only two bivariate models (*LHELX* and *WTQ*) have a smaller mean-squared-errors than this single factor model. The second conclusion is that there is little if any improvement in the interest rate and nominal money combination forecasts over their bivariate analogues in Table 3. The variables continue to perform relatively poorly. Finally, the ridge regression forecasts outperform unemployment in both subsamples for both inflation variables (both economically and statistically) using either the real activity indicators or the full set of indicators. The ridge regression forecasts typically improve upon the mean and median forecasts. However, none of the combination forecasts perform as well as the single factor models.

Fig. 3 plots the all-indicators and real-activity-indicators single factors together with the unemployment rate. The series are expressed in standard deviation units and the unemployment rate has been multiplied by -1 so that peaks in both series corresponds to expansions in real activity. To facilitate

---

[4] Before the factors were estimated, each data series was automatically screened for 'outliers'. Specically, the inter-quartile range (IQR) was computed for each series and observations with median deviations exceeding $\lambda \times \mathrm{IQR}$ were labeled as outliers. These observations were handled using two different methods. The first method was simply to discard series that contained any outliers. In the second method, the specific outlying observations were treated as 'missing data' and a factor estimation method that allows for missing observations was used. (This method is described in Stock and Watson (1998).) For the first method a large value of λ ($\lambda = 20$) was used, and this resulted in three series being dropped from the analysis (*LPMI*, *LPTU* and *FCLBMC*). For the second method a smaller value of λ ($\lambda = 6$) was used, and this resulted in outliers being identified in 40 of the series. The resulting factor estimates and forecasts using these two different methods were very similar. (For example, the sample correlation coefficient for the first factor was 0.995), and to save space we report results for the first method only.

Table 4
Forecasting performance of multivariate models

	PUNEW				GDMC			
	1970–1983		1984–1996		1970–1983		1984–1996	
Variables	Rel. MSE	λ	Rel. MSE	λ	Rel. MSE	λ	Rel. MSE	λ
Univariate	1.26	− 0.13	0.98	0.53	1.00	0.50	1.06	0.27
	(0.19)	(0.25)	(0.15)	(0.33)	(0.15)	(0.38)	(0.09)	(0.29)
Panel A. All indicators								
Mul. factors	0.80	0.86	0.91	0.60	0.97	0.59	0.84	0.69
	(0.11)	(0.19)	(0.16)	(0.17)	(0.08)	(0.23)	(0.16)	(0.16)
1 factor	0.72	1.63	0.86	0.82	0.71	1.53	0.90	0.87
	(0.08)	(0.27)	(0.10)	(0.21)	(0.21)	(0.43)	(0.11)	(0.33)
Comb. mean	1.04	0.29	0.88	0.87	0.91	0.89	0.97	0.63
	(0.09)	(0.39)	(0.09)	(0.27)	(0.14)	(0.55)	(0.05)	(0.26)
Comb. median	1.14	0.03	0.92	0.70	0.95	0.64	1.02	0.40
	(0.13)	(0.31)	(0.11)	(0.29)	(0.15)	(0.46)	(0.07)	(0.28)
Comb. ridge reg.	0.86	1.20	0.87	0.73	0.90	1.15	0.94	0.63
	(0.06)	(0.22)	(0.17)	(0.29)	(0.09)	(0.39)	(0.10)	(0.23)
Panel B. Real activity indicators								
Mul. factors	0.72	1.26	1.03	0.47	0.87	0.99	0.78	0.81
	(0.10)	(0.19)	(0.16)	(0.17)	(0.10)	(0.30)	(0.16)	(0.18)
1 factor	0.75	1.40	0.86	0.86	0.80	1.12	0.90	0.71
	(0.08)	(0.26)	(0.10)	(0.21)	(0.20)	(0.52)	(0.13)	(0.26)
Comb. mean	0.97	0.72	0.88	0.96	0.87	1.29	0.94	0.78
	(0.07)	(0.47)	(0.06)	(0.19)	(0.14)	(0.66)	(0.06)	(0.26)
Comb. median	1.00	0.53	0.90	0.87	0.89	1.04	0.97	0.64
	(0.08)	(0.46)	(0.06)	(0.19)	(0.15)	(0.65)	(0.05)	(0.25)
Comb. ridge reg.	0.84	1.47	0.90	0.74	0.84	1.55	0.95	0.61
	(0.06)	(0.34)	(0.11)	(0.25)	(0.13)	(0.50)	(0.10)	(0.21)
Panel C. Interest rates								
Mul. factors	1.17	0.34	1.19	0.29	1.04	0.43	1.25	0.14
	(0.18)	(0.15)	(0.24)	(0.21)	(0.13)	(0.24)	(0.18)	(0.15)
1 factor	1.12	0.34	1.19	0.27	1.10	0.23	1.05	0.29
	(0.17)	(0.21)	(0.25)	(0.24)	(0.12)	(0.30)	(0.08)	(0.30)
Comb. mean	1.03	0.42	0.96	0.59	0.96	0.62	1.06	0.28
	(0.11)	(0.28)	(0.15)	(0.31)	(0.13)	(0.40)	(0.08)	(0.30)
Comb. median	1.11	0.21	0.95	0.60	0.97	0.58	1.05	0.29
	(0.13)	(0.30)	(0.15)	(0.32)	(0.15)	(0.43)	(0.08)	(0.29)
Comb. ridge reg.	1.04	0.42	1.00	0.51	1.00	0.49	1.14	0.09
	(0.12)	(0.24)	(0.17)	(0.31)	(0.13)	(0.34)	(0.11)	(0.23)

Table 4. (continued)

Variables	PUNEW				GDMC			
	1970–1983		1984–1996		1970–1983		1984–1996	
	Rel. MSE	λ	Rel. MSE	λ	Rel. MSE	λ	Rel. MSE	λ
Panel D. Money								
Mul. factors	1.26	− 0.13	0.98	0.53	0.99	0.52	1.06	0.28
	(0.19)	(0.25)	(0.16)	(0.33)	(0.16)	(0.39)	(0.09)	(0.30)
1 factor	1.26	− 0.13	0.98	0.53	0.99	0.52	1.06	0.28
	(0.19)	(0.25)	(0.16)	(0.33)	(0.16)	(0.39)	(0.09)	(0.30)
Comb. mean	1.25	− 0.05	0.97	0.56	1.03	0.44	1.03	0.39
	(0.19)	(0.26)	(0.15)	(0.27)	(0.16)	(0.37)	(0.08)	(0.26)
Comb. median	1.25	− 0.08	0.97	0.56	1.02	0.46	1.04	0.37
	(0.18)	(0.27)	(0.14)	(0.27)	(0.16)	(0.37)	(0.09)	(0.27)
Comb. ridge reg.	1.22	0.02	1.02	0.47	1.00	0.51	1.14	0.17
	(0.17)	(0.26)	(0.16)	(0.25)	(0.17)	(0.36)	(0.12)	(0.22)

Notes: Results are shown for multivariate models using different groups of variables. Panel A include all the variables listed in Tables 2 and 3. Panel B include the variables in Table 2 and the output, employment, consumption, sales, housing, inventory and orders variables in Table 3. Panel C includes the interest rate variables in Table 3, and Panel D includes the money variables in Table 3 using the DDLN transformation. The results for rows labeled 'Mul. factors' and '1 factor' are for forecasts constructed using multiple factors and single factor as described in the text. The other rows entries are for forecasts constructed from the mean, median and ridge regression combining formulas. See the notes for Table 2 for additional details.

interpretation, the estimated factors has also been smoothed using the filter $(1/3)\times(L^{-1}+1+L)$ to eliminate some of its high frequency variation. Three features are particularly noteworthy. First, the estimated factor computed using all of the indicators is essentially identical to the factor computed using only the real activity indicators. Slight differences can only be seen at the series' peaks and troughs, and the correlation between the series is 0.98. Thus, the all-indicator single factor should be interpreted as an index of real economic activity. Second, the estimated activity factors have more cyclical variability than the unemployment rate. (For example, compare the series during the 1967 growth recession and the two recessions of the early 1980s). Third, the factors tend to lead the unemployment rate by several months, as can be seen by comparing the business cycle peak and trough dates of the series.

The final issue addressed in this section is whether forecasts based on a Phillips curve, reinterpreted as a relationship between inflation and the single actitivity factor, can be improved upon by including additional variables

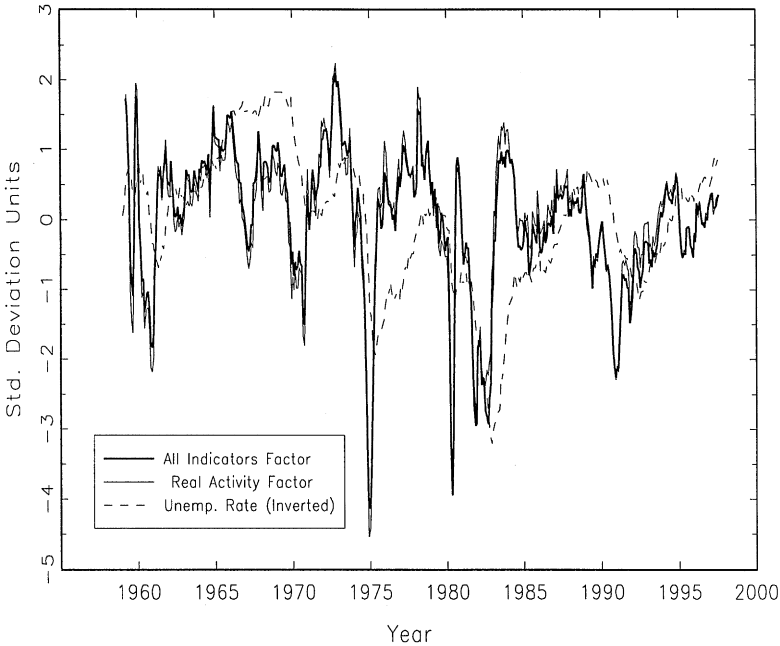

Fig. 3. Activity indicators.

(money, interest rates, commodity prices, etc.) To answer this question the analysis of Table 4 is repeated, except that the benchmark model uses the single activity factor from the all-indicators model rather than the unemployment rate. We then ask whether more accurate forecasts can be constructed by combining this new benchmark Phillips curve forecast with the forecasts constructed from the interest rate, money or real-indicators factor models. These combined forecasts are computed from Eq. (8) using OLS to recursively estimate the coefficients ω_{it}.

The results are sumarized in Table 5. There is no evidence suggesting that any of the other models dominate the new benchmark model for predicting CPI inflation over this period. As expected from the results in Table 4, the interest rate and money models are dominated by the benchmark model. More interesting are the results shown in the bottom of the table, where these forecasts are combined with the new benchmark Phillips curve forecast: there is no indication that any of these models is preferred to the new benchmark model. In summary, Table 5 indicates that it is difficult to improve upon forecasts made using the single activity factor.

Fig. 4 plots realizations of CPI inflation with corresponding forecasts constructed 12 months earlier. (The series are aligned so that the vertical distance

Table 5
Forecasting performance relative to real activity single factor model

Variables	PUNEW				GDMC			
	1970–1983		1984–1996		1970–1983		1984–1996	
	Rel. MSE	λ	Rel. MSE	λ	Rel. MSE	λ	Rel. MSE	λ
Univariate	1.74	− 0.16	1.15	0.32	1.41	0.10	1.18	0.08
	(0.44)	(0.16)	(0.29)	(0.28)	(0.27)	(0.15)	(0.21)	(0.33)
1hur	1.38	− 0.63	1.17	0.18	1.41	− 0.53	1.11	0.13
	(0.15)	(0.27)	(0.14)	(0.21)	(0.41)	(0.43)	(0.13)	(0.33)
Panel A. All indicators								
Mul. factors	1.10	0.24	1.06	0.38	1.36	− 0.31	0.94	0.61
	(0.15)	(0.36)	(0.17)	(0.34)	(0.35)	(0.43)	(0.13)	(0.22)
Comb. mean	1.44	− 0.16	1.03	0.43	1.28	0.06	1.08	0.15
	(0.26)	(0.23)	(0.18)	(0.37)	(0.22)	(0.21)	(0.14)	(0.47)
Comb. median	1.58	− 0.16	1.07	0.38	1.34	0.10	1.14	0.10
	(0.34)	(0.19)	(0.22)	(0.31)	(0.25)	(0.17)	(0.18)	(0.37)
Comb. ridge reg.	1.18	− 0.32	1.02	0.47	1.27	− 0.37	1.05	0.31
	(0.13)	(0.44)	(0.23)	(0.46)	(0.26)	(0.44)	(0.09)	(0.36)
Panel B. Real activity indicators								
Mul. factors	0.99	0.54	1.20	0.09	1.22	− 0.20	0.87	0.80
	(0.11)	(0.51)	(0.19)	(0.35)	(0.25)	(0.50)	(0.12)	(0.24)
1 factor	1.03	0.26	1.00	0.51	1.13	− 0.59	1.00	0.48
	(0.05)	(0.45)	(0.07)	(0.84)	(0.07)	(0.25)	(0.06)	(0.37)
Comb. mean	1.34	− 0.20	1.02	0.42	1.22	0.04	1.05	0.16
	(0.20)	(0.26)	(0.13)	(0.41)	(0.18)	(0.25)	(0.09)	(0.57)
Comb. median	1.38	− 0.22	1.05	0.35	1.25	0.04	1.08	− 0.03
	(0.22)	(0.24)	(0.13)	(0.38)	(0.20)	(0.21)	(0.10)	(0.51)
Comb. ridge reg.	1.17	− 0.17	1.05	0.29	1.18	− 0.06	1.05	0.25
	(0.12)	(0.31)	(0.13)	(0.58)	(0.17)	(0.34)	(0.07)	(0.34)
Panel C. Interest rates								
Mul. factors	1.62	0.11	1.39	0.17	1.46	0.07	1.39	− 0.03
	(0.40)	(0.15)	(0.40)	(0.20)	(0.42)	(0.22)	(0.24)	(0.18)
1 factor	1.55	0.13	1.39	0.20	1.55	− 0.03	1.17	0.07
	(0.40)	(0.21)	(0.45)	(0.22)	(0.41)	(0.21)	(0.20)	(0.34)
Comb. mean	1.43	0.03	1.12	0.38	1.35	0.05	1.17	0.09
	(0.29)	(0.26)	(0.28)	(0.26)	(0.27)	(0.22)	(0.20)	(0.33)
Comb. median	1.53	− 0.08	1.11	0.37	1.37	0.07	1.17	0.07
	(0.33)	(0.24)	(0.28)	(0.28)	(0.27)	(0.19)	(0.20)	(0.34)
Comb. ridge reg.	1.44	0.05	1.16	0.34	1.41	0.02	1.27	− 0.02
	(0.30)	(0.24)	(0.30)	(0.25)	(0.28)	(0.19)	(0.22)	(0.28)

Table 5
Forecasting performance relative to real activity single factor model

Variables	PUNEW				GDMC			
	1970–1983		1984–1996		1970–1983		1984–1996	
	Rel. MSE	λ	Rel. MSE	λ	Rel. MSE	λ	Rel. MSE	λ
Panel D. Money								
Mul. factors	1.74	– 0.15	1.15	0.32	1.40	0.12	1.18	0.09
	(0.44)	(0.16)	(0.29)	(0.27)	(0.26)	(0.15)	(0.21)	(0.33)
1 factor	1.74	– 0.15	1.15	0.32	1.40	0.12	1.18	0.09
	(0.44)	(0.16)	(0.29)	(0.27)	(0.26)	(0.15)	(0.21)	(0.33)
Comb. mean	1.72	– 0.11	1.13	0.33	1.44	0.09	1.15	0.13
	(0.43)	(0.17)	(0.27)	(0.29)	(0.29)	(0.15)	(0.19)	(0.36)
Comb. median	1.73	– 0.13	1.13	0.34	1.43	0.10	1.15	0.14
	(0.43)	(0.17)	(0.27)	(0.28)	(0.28)	(0.15)	(0.20)	(0.34)
Comb. ridge reg.	1.69	– 0.09	1.19	0.28	1.40	0.13	1.27	– 0.00
	(0.40)	(0.17)	(0.30)	(0.27)	(0.22)	(0.12)	(0.24)	(0.28)
Panel E. Activity factor combined with								
Real activity ind.	1.12	– 0.20	1.00	0.86	0.99	0.58	1.10	– 0.75
	(0.10)	(0.38)	(0.02)	(1.88)	(0.03)	(0.28)	(0.05)	(0.44)
Interest rates	1.65	– 0.07	0.98	1.31	1.47	0.11	1.08	– 0.55
	(0.51)	(0.16)	(0.03)	(1.75)	(0.58)	(0.15)	(0.04)	(0.38)
Money	1.60	0.15	1.04	– 1.32	1.26	0.13	1.05	– 0.56
	(0.67)	(0.06)	(0.02)	(0.93)	(0.29)	(0.15)	(0.04)	(0.70)
Int. rates, money	1.73	0.12	1.13	– 0.21	1.66	0.12	1.04	– 0.27
	(0.72)	(0.06)	(0.11)	(0.47)	(0.95)	(0.12)	(0.04)	(0.63)
Activity factor RMSE (% per annum)	2.1		1.3		1.7		1.0	

Notes: See notes to Tables 2 and 4 for a description of the table entries. The benchmark forecast used in this table is constructed from the single factor all-indicators model. The results shown under Panel E are for forecasts constructed using a recursive OLS combination of the benchmark forecast with single factor forecasts from the models listed in the first column.

between the plot of inflation and the forecast represents the forecast error.) Forecasts constructed using LHUR and the single activity factor are shown. While the two forecasts are usually similar, they differ in some periods, and the single factor forecast is on average more accurate than the unemployment rate forecast over the entire sample period.

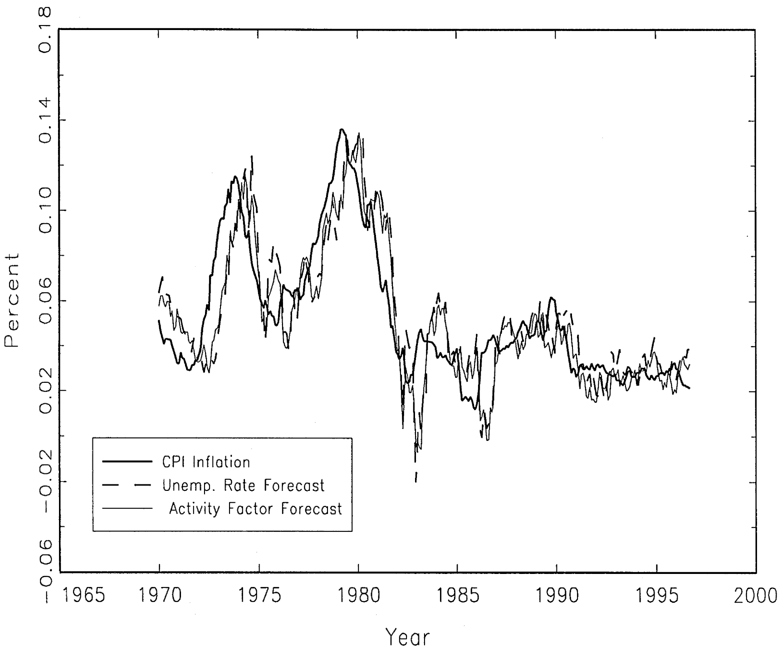

Fig. 4. Annual CPI inflation and forecasts made 12 months earlier.

6. Robustness to the assumption that inflation is $I(1)$

The results reported thus far rely on a specification that imposes the restriction that inflation is $I(1)$. In this section we study the robustness of the forecasting results by respecifying Eq. (3) as

$$\pi^h_{t+h} - \pi_t = \phi + \beta(L)x_t + \mu(L)\pi_t + e_{t+h}. \tag{12}$$

Eq. (12) reduces to (3) after imposing the restriction $\mu(1) = 1$.

Results are reported in Table 6. The benchmark model in Table 6 is the $I(0)$ specification (12), with x_t equal to the single activity factor computed using all of the indicators. The first row of Table 6 compares this benchmark $I(0)$ model to the benchmark $I(1)$ model from Table 5, in which x_t is the single activity factor. The remaining comparisons in Table 6 are between selected forecasts, computed using $I(0)$ specifications, and the benchmark $I(0)$ model.

The relative performance of the $I(0)$ and $I(1)$ specifications that use the single activity index vary across sample periods: imposing the unit root restriction leads to more accurate predictions in the first sample period but less accurate predictions in the second sample period. These results are consistent with unit

Table 6
Forecasting performance of $I(0)$ models of inflation relative to real activity single factor model

	PUNEW				GMDC			
	1970–1983		1984–1996		1970–1983		1984–1996	
Variable	Rel. MSE	λ	Rel. MSE	λ	Rel. MSE	λ	Rel. MSE	λ
Table 5 benchmark	0.89	2.44	1.09	−0.61	0.86	2.22	1.08	− 0.35
	(0.07)	(0.77)	(0.10)	(0.85)	(0.09)	(0.73)	(0.08)	(0.66)
lhur	1.44	− 0.29	1.22	0.05	1.58	− 0.57	1.12	0.02
	(0.34)	(0.38)	(0.16)	(0.25)	(0.62)	(0.41)	(0.11)	(0.36)
Univariate	1.55	− 0.16	1.18	0.30	1.36	0.01	1.17	0.07
	(0.30)	(0.21)	(0.29)	(0.28)	(0.25)	(0.19)	(0.18)	(0.33)
Panel A. All indicators								
Mul. factors	1.06	0.37	1.32	0.15	1.44	0.10	1.16	0.29
	(0.15)	(0.29)	(0.33)	(0.29)	(0.46)	(0.24)	(0.23)	(0.26)
Panel B. Real activity indicators								
Mul. factors	0.87	0.80	1.22	0.17	1.09	0.31	1.03	0.45
	(0.12)	(0.32)	(0.29)	(0.38)	(0.24)	(0.45)	(0.15)	(0.29)
1 factor	1.03	0.25	1.05	− 0.27	1.07	− 0.27	1.07	− 0.01
	(0.05)	(0.47)	(0.08)	(0.97)	(0.05)	(0.34)	(0.06)	(0.38)
ipxmca	1.26	− 0.44	0.97	0.59	1.24	− 0.28	1.07	0.11
	(0.13)	(0.31)	(0.12)	(0.43)	(0.25)	(0.50)	(0.12)	(0.51)
hsbp	1.46	− 0.57	1.14	0.08	1.52	− 0.31	1.03	0.40
	(0.33)	(0.36)	(0.12)	(0.25)	(0.57)	(0.35)	(0.11)	(0.38)
lhmu25	1.25	− 0.31	1.19	0.12	1.26	− 0.26	1.17	− 0.02
	(0.13)	(0.31)	(0.20)	(0.33)	(0.15)	(0.24)	(0.10)	(0.26)
msmtq	1.09	0.28	1.19	0.25	1.04	0.42	1.00	0.50
	(0.14)	(0.32)	(0.39)	(0.41)	(0.12)	(0.28)	(0.18)	(0.48)
Panel C. Interest rates								
Mul. factors	3.13	− 0.08	1.42	0.23	3.01	− 0.01	1.40	− 0.04
	(2.68)	(0.13)	(0.49)	(0.19)	(3.16)	(0.11)	(0.35)	(0.25)
1 factor	2.94	− 0.24	1.17	0.28	2.53	− 0.01	1.18	0.05
	(2.81)	(0.08)	(0.24)	(0.25)	(2.33)	(0.14)	(0.17)	(0.29)
fygm3	2.20	− 0.22	1.25	0.17	2.03	− 0.00	1.20	− 0.03
	(1.13)	(0.13)	(0.24)	(0.23)	(1.03)	(0.14)	(0.17)	(0.28)
fygt1	2.66	− 0.19	1.22	0.24	2.14	0.02	1.19	0.04
	(2.28)	(0.12)	(0.26)	(0.24)	(1.29)	(0.15)	(0.18)	(0.29)
fygm3-CI	2.00	− 0.13	1.33	− 0.02	1.80	0.01	1.26	− 0.34
	(0.79)	(0.15)	(0.39)	(0.27)	(0.77)	(0.18)	(0.24)	(0.36)
fygt1-CI	2.05	− 0.10	1.34	− 0.08	1.93	0.05	1.27	− 0.39
	(0.95)	(0.15)	(0.37)	(0.26)	(1.04)	(0.16)	(0.24)	(0.37)

Table 6. (continued)

	PUNEW				GMDC			
	1970–1983		1984–1996		1970–1983		1984–1996	
Variable	Rel. MSE	λ	Rel. MSE	λ	Rel. MSE	λ	Rel. MSE	λ
Panel D. Money								
Mul. factors	1.52	0.10	1.58	0.05	1.26	0.13	1.36	0.02
	(0.32)	(0.16)	(0.61)	(0.30)	(0.21)	(0.21)	(0.28)	(0.23)
1 factor	1.56	0.02	1.33	0.28	1.35	0.07	1.12	0.31
	(0.32)	(0.21)	(0.45)	(0.26)	(0.20)	(0.16)	(0.20)	(0.28)
fm2	1.82	− 0.14	1.02	0.48	1.46	− 0.03	1.01	0.46
	(0.43)	(0.17)	(0.24)	(0.30)	(0.29)	(0.17)	(0.13)	(0.36)
fmbase	1.47	0.07	1.74	0.14	1.35	0.09	1.35	− 0.04
	(0.27)	(0.22)	(0.76)	(0.23)	(0.20)	(0.17)	(0.29)	(0.29)
Panel E. Prices								
pmcp	1.22	− 0.15	1.28	0.20	1.17	0.03	1.28	− 0.01
	(0.15)	(0.34)	(0.33)	(0.27)	(0.23)	(0.45)	(0.29)	(0.36)
psm99q	0.92	0.64	1.32	0.23	0.97	0.57	1.43	− 0.04
	(0.11)	(0.19)	(0.28)	(0.21)	(0.09)	(0.20)	(0.27)	(0.21)
exrus	1.57	0.13	2.47	0.16	1.45	0.20	1.95	0.07
	(0.34)	(0.12)	(1.96)	(0.15)	(0.35)	(0.13)	(0.99)	(0.19)
Activity factor RMSE (% per annum)	2.2		1.3		1.8		1.0	

Notes: See notes to Tables 2 and 4 for a description of the table entries. The benchmark forecast used in this table is constructed from the single factor all-indicators model. All of the specifications except the entry labeled ‘Table 5 Benchmark’ are based on Eq. (12).

root tests applied to the inflation rate. Recursively computed unit root tests ($DFGLS^{\mu}$ from Elliott et al. (1996)) have p-values larger 10% for both inflation series through 1982 and p-values less than 10% after 1982 for CPI and after 1985 for the PCE. Of course, these univariate tests are merely suggestive: a formal unit root pretest strategy for the models considered in this paper would involve multivariate unit root and cointegration tests.

The results for the other variables are generally consistent with the results presented earlier. The $I(0)$ single activity index model produces more accurate forecasts than the $I(0)$ model that uses the unemployment rate (*LHUR*) as the activity indicator, particularly in the first half of the forecast period. This specification allows inflation and the unemployment rate to be cointegrated as in Ireland (1999). Looking at the individual real indicators, there is only one relative mean-square-error that is less than unity: capacity utilization provides a more accurate forecast for CPI inflation in the 1984–1996 sample period.

The other entries in the table rely on transformations of the indicators consistent with the levels specification for inflation. Thus, interest rates are allowed to enter in levels, and the interest rate factors are constructed using the levels of interest rates. Letting interest rates enter (12) in levels introduces an important variant: inflation and interest rates could be $I(1)$ and cointegrated, where the cointegrating vector is implicitly estimated by recursive nonlinear least squares. A further variant is to impose that these series are cointegrated and have a cointegrating vector of $(1, -1)$, consistent with the hypothesis that real interest rates are $I(0)$. This is done in the rows labeled *fygm3–CI* and *fygt1–CI*. Nominal money enters in growth rates. Finally, the price indexes, *pmcp* and *psm99q* are entered as first difference of logarithms.

Even though these models introduce richer low frequency dynamics than the $I(1)$ models of the earlier sections, they produce poor forecasts. Although there is some evidence that the index of sensitive material prices (*psm99q*) helped to forecast inflation during the 1970s, the forecasting performance of this model deteriorated significantly in the 1984–1996 sample period. No forecast outperforms the benchmark model for both inflation series in both sample period. The models that impose $I(0)$ real rates do particularly poorly, especially in the 1970–1983 sample. Comparison of the corresponding entries in Tables 5 and 6 indicates that the single activity model does relatively better than the alternative forecasts when comparisons are made among $I(0)$ specifications, than among $I(1)$ specifications.

In summary, these results suggest that the forecasts with $I(1)$ specifications of inflation are generally (but not always) preferred to those with $I(0)$ specifications; that in some cases the $I(0)$ forecasts perform extremely poorly; and that the results of the previous section are robust to specifying inflation as $I(0)$ rather than $I(1)$.

7. Discussion and conclusion

Some caveats are in order. First, the approach in this paper has been to evaluate forecasting performance using a simulated out-of-sample methodology. This methodology provides a degree of protection against overfitting and detects model instability. However, because a large number of forecasts were used, some overfitting bias nonetheless remains. This suggests that some of the best-performing forecasts produced using individual economic indicators might deteriorate as one moves beyond the end of our sample. Because the pool of forecasts is larger for the individual indicators considered in Section 4 than for the composite indexes considered in Section 5, overfitting is arguably more of an issue for the individual indicator forecasts than the composite forecasts. Second, we have considered only linear models. To the extent that the relation between inflation and some of the candidate variables is nonlinear, these results understate the forecasting improvements that might be obtained, relative to the conventional linear Phillips curve. Finally, our analysis has been limited to a one-year-ahead forecasting horizon.

The major conclusion of this study is that the Phillips curve, interpreted broadly as a relation between current real economic activity and future inflation, produced the most reliable and accurate short-run forecasts of US price inflation across all of the models that we considered over the 1970–1996 period. This conclusion will come as no surprise to applied macroeconomic forecasters in business and government, where the Phillips curve plays a central role in short-run inflation forecasting. The conclusion is also consistent with the recent academic literature on short-run inflation forecasting. For example, in a comparison of 71 potential leading indicators of inflation, Staiger et al. (1997a) report that the unemployment rate ranks 7th over the 1975–1984 forecasting period and 10th over 1985–1993. The only variable which dominates the unemployment rate over both periods is another indicator of real activity, the rate of capacity utilization.

The conventionally specified Phillips curve, based on the unemployment rate, was found to perform reasonably well. Its forecasts are better than univariate forecasting models (both autoregressions and random walk models), which in many situations have proven to be surprisingly strong benchmarks.[5] Moreover, with few exceptions, incoporating other variables does not significantly improve upon its short-run forecasts. Specifically, there are no gains from including money supply measures (consistent with results in Estrella and Mishkin (1997)), interest rates and spreads (consistent with the 'short-end of the term structure' results reported in Mishkin (1990)), or commodity prices (in contrast to the 'price puzzle' rationale for including commodity prices in VARs first suggested in Sims (1992)).

The few forecasts that do consistently improve upon unemployment rate Phillips curve forecasts are in fact from alternative Phillips curves, specified using other measures of aggregate activity instead of the unemployment rate. These measures include the capacity utilization rate and real manufacturing and trade sales. Interestingly, combining the forecasts produced by 85 separate generalized Phillips curve specifications, each with a different activity measure, also improved upon forecasts made solely using the unemployment rate.

Perhaps the most intriguing result is that the best-performing forecast is a Phillips curve forecast that uses a new composite index of aggregate activity comprised of the 168 individual activity measures. The forecasting gains from using this index are economically large and statistically significant over the 1970–1996 sample period, and we were unable to improve upon this forecast by

[5] Granger and Newbold (1976) provide a survey of early comparisons of forecasting performance of univariate and multivariate models, and Zarnowitz and Braun (1993) compare forecasts from univariate and VAR models with forecasts constructed by professional forecasters for the US over the 1968–1990 period.

combining it with other forecasts. This conclusion is consistent with the findings of the most recent studies of the apparent breakdown of the unemployment rate Phillips curve during 1997–1998. As is discussed in Gordon (1998) and Stock (1998), this poor performance seems to be associated with the specific use of unemployment rate as the activity indicator; they find that Phillips curve forecasts using alternative real activity measures perform much better than unemployment rate Phillips curves over this sample period.

Appendix A. Data description

This appendix lists the time series used to construct the diffusion index forecasts discussed in Section 5. The format is: series number; series mnemonic; data span used; and brief series description. The series were either taken directly from the DRI–McGraw Hill Basic Economics database, in which case the original mnemonics are used, or they were produced by authors' calculations based on data from that database, in which case the authors calculations and original DRI/McGraw series mnemonics are summarized in the data description field. The following abbreviations appear in the data definitions: SA = seasonally adjusted; NSA = not seasonally adjusted; SAAR = seasonally adjusted at an annual rate; FRB = Federal Reserve Board.

Real output and income

ip	industrial production: total index (1992 = 100, SA)
ipp	industrial production: products, total (1992 = 100, SA)
ipf	industrial production: final products (1992 = 100, SA)
ipc	industrial production: consumer goods (1992 = 100, SA)
ipcd	industrial production: durable consumer goods (1992 = 100, SA)
ipcn	industrial production: nondurable consumer goods (1992 = 100, SA)
ipe	industrial production: business equipment (1992 = 100, SA)
ipi	industrial production: intermediate products (1992 = 100, SA)
ipm	industrial production: materials (1992 = 100, SA)
ipmd	industrial production: durable goods materials (1992 = 100, SA)
ipmnd	industrial production: nondurable goods materials (1992 = 100, SA)
ipmfg	industrial production: manufacturing (1992 = 100, SA)
ipd	industrial production: durable manufacturing (1992 = 100, SA)
ipn	industrial production: nondurable manufacturing (1992 = 100, SA)
ipmin	industrial production: mining (1992 = 100, SA)
iput	industrial production: utilities (1992 = 100, SA)
ipxmca	capacity utilization rate: manufacturing, total (% of capacity, SA) (FRB)
pmi	purchasing managers' index (SA)
pmp	NAPM production index (%)

gmpyq	personal income (chained) ($92 b, SAAR)
gmyxpq	personal income less transfer payments (chained) ($92 b) (SAAR)

Employment and hours

lhel	index of help-wanted advertising in newspapers (1967 = 100; SA)
lhelx	employment: ratio; help-wanted ads: no. unemployed in civilian labour force
lhem	civilian labor force: employed, total (in thousands, SA)
lhnag	civilian labor force: employed, nonagricultural industries (in thousands, SA)
lhur	unemployment rate: all workers, 16 years and over (%, SA)
lhu680	unemployment by duration: average(mean)duration in weeks (SA)
lhu5	unemployment by duration: persons unemployed less than weeks (in thousands, SA)
lhu14	unemployment by duration: persons unemployed 5 to 14 weeks (in thousands, SA)
lhu15	unemployment by duration: persons unemployed 15 weeks + (in thousands, SA)
lhu26	unemployment by duration: persons unemployed 15 to 26 weeks (in thousands, SA)
lhu27	unemployment by duration: persons unemployed 27 weeks + (in thousands, SA)
lpnag	employees on nonagricultural payrolls: total (in thousands, SA)
lp	employees on nonagricultural payrolls: total, private (in thousands, SA)
lpgd	employees on nonagricultural payrolls: goods-producing (in thousands, SA)
lpmi	employees on nonagricultural payrolls: mining (in thousands, SA)
lpcc	employees on nonagricultural payrolls: contract construction (in thousands, SA)
lpem	employees on nonagricultural payrolls: manufacturing (in thousands, SA)
lped	employees on nonagricultural payrolls: durable goods (in thousands, SA)
lpen	employees on nonagricultural payrolls: nondurable goods (in thousands, SA)
lpsp	employees on nonagricultural payrolls: service-producing (in thousands, SA)
lptu	employees on nonagricultural payrolls: transport and public utilities (in thousands, SA)
lpt	employees on nonagricultural payrolls: wholesale and retail trade (in thousands, SA)
lpfr	employees on nonagricultural payrolls: finance, insured and real estate (in thousands, SA)

lps	employees on nonagricultural payrolls: services (in thousands, SA)
lpgov	employees on nonagricultural payrolls: government (in thousands, SA)
lphrm	average weekly hours of production workers: manufacturing (SA)
lpmosa	average weekly hours of production workers: manufacturing, overtime hours (SA)
luinc	average weekly initial claims, state unemployment insured, exc p. rico (in thousands, SA)
pmemp	NAPM employment index (%)

Real retail, manufacturing and trade sales

msmtq	manufacturing and trade: total (millions of chained 1992 dollars) (SA)
msmq	manufacturing and trade: manufacturing; total (millions of chained 1992 dollars) (SA)
msdq	manufacturing and trade: manufacturing; durable goods (millions of chained 1992 dollars) (SA)
msnq	manufacturing and trade: manufacturing; nondurable goods (millions of chained 1992 dollars) (SA)
wtq	merchant wholesalers: total (milions of chained 1992 dollars) (SA)
wtdq	merchant wholesalers: durable goods total (millions of chained 1992 dollars) (SA)
wtnq	merchant wholesalers: nondurable goods (millions of chained 1992 dollars) (SA)
rtq	retail trade: total (millions of chained 1992 dollars) (SA)
rtnq	retail trade: nondurable goods (millions of 1992 dollars) (SA)

Consumption

gmcq	personal consumption expenditure (chained) – total ($92 b) (SAAR)
gmcdq	personal consumption expenditure (chained) – total durables ($92 b) (SAAR)
gmcnq	personal consumption expenditure (chained) – nondurables ($92 b) (SAAR)
gmcsq	personal consumption expenditure (chained) – services ($92 b) (SAAR)
gmcanq	personal consumption expenditure (chained) – new cars ($92 b) (SAAR)

Housing starts and sales

hsfr	housing starts: nonfarm (1947–1958); total farm and nonfarm (1959–) (in thousands, SA)
hsne	housing starts: northeast (thousand units) SA
hsmw	housing starts: midwest (thousand units) SA
hssou	housing starts: south (thousand units) SA

hswst	housing starts: west (thousand units) SA
hsbp	building permits for new private housing units (thousands)
hsbr	housing authorized: total new private housing units (thousands, SAAR)
hmob	mobile homes: manufacturers' shipments (thousand units, SAAR)
condo9	construction contracts: commertial industrial buildings (million square feet floor sp., SA)

Inventories and orders

ivmtq	manufacturing and trade inventories: total (millions of chained dollars 1992) (SA)
ivmfgq	inventories, business, manufacturing (millions of chained 1992 dollars, SA)
ivmfdq	inventories, business durables (millions of chained 1992 dollars, SA)
ivmfnq	inventories, business nondurables (millions of chained 1992 dollars, SA)
ivwrq	manufacturing and trade inventories: merchant wholesalers (millions of chained 1992 dollars)
ivrrq	manufacturing and trade inventories: retail trade (millions of chained 1992 dollars) (SA)
ivsrq	ratio for manufacturing and trade: inventory/sales (chained 1992 dollars, SA)
ivsrmq	ratio for manufacturing and trade: manufacturing; inventory/sales ($87) (SA)
ivsrwq	ratio for manufacturing and trade: wholesaler; inventory/sales ($87) (SA)
ivsrrq	ratio for manufacturing and trade: retail trade; inventory/sales ($87) (SA)
pmnv	NAPM inventories index (%)
pmno	NAPM new orders index (%)
mocmq	new orders (net) – consumer goods and materials, in 1992 dollars (BCI)
mdoq	new orders, durable goods industries, in 1992 dollars (BCI)
msondq	new orders, nondefense capital goods, in 1992 dollars (BCI)
mpconq	contracts and orders for plant and equipment, in 1992 dollars (BCI)

Stock prices

fsncom	NYSE common stock price index: composite (12/31/65 = 50)
fspcom	S&P's common stock price index: composite (1941–1943 = 10)
fspin	S&P's common stock price index: industrials (1941–1943 = 10)
fspcap	S&P's common stock price index: capital goods (1941–1943 = 10)
fsput	S&P's common stock price index: utilities (1941–1943 = 10)
fsdxp	S&P's composite common stock: dividend yield (% per annum)

fspxe	S&P's composite common stock: price–earnings ratio (%, NSA)

Exchange rates

exrus	United States; effective exchange rate (merm) (index no.)
exrger	foreign exchange rate: Germany (DM per US$)
exrsw	foreign exchange rate: Switzerland (Swiss Franc per US$)
exrjan	foreign exchange rate: Japan (Yen per US$)
exruk	foreign exchange rate: United Kingdom (Cents per Pound)
exrcan	foreign exchange rate: Canada (Canadian $ per US$)

Interest rates

fyff	interest rate: Federal funds (effective) (% per annum, NSA)
fycp	interest rate: commercial paper, 6-month (% per annum, NSA)
fygm3	interest rate: US treasury bills, securing market, 3-monthly (% per annum, NSA)
fygm6	interest rate: US treasury bills, securing market, 6-monthly (% per annum, NSA)
fygt1	interest rate: US treasury const maturities, 1-year (% per annum, NSA)
fygt5	interest rate: US treasury const maturities, 5-year (% per annum, NSA)
fygt10	interest rate: US treasury const maturities, 10-year (% per annum, NSA)
fyaaac	bond yield: Moody's AAA corporate (% per annum)
fybaac	bond yield: Moody's BAA corporate (% per annum)
fyfha	secondary market yields on FHA mortgages (% per annum)
sp_fycp	Spread *fycp–fygm3*
sp_fyff	Spread *fyff–fygm3*
sp_fygm6	Spread *fygm6–fygm3*
sp_fygt1	Spread *fygt1–fygm3*
sp_fygt5	Spread *fygt5–fygm3*
sp_fygt10	Spread *fygt10–fygm3*
sp_fyaaac	Spread *fyaaac–fygm3*
sp_fybaac	Spread *fybaac–fygm3*
sp_fyfha	Spread *fyfha–fygm3*

Money and credit quantity aggregates

fm1	money stock: M1 (current travellers' checks, demand deposits other checkable deposits) (SA)
fm2	money stock: M2 (M1 + o'nite rps, euro$, G/P&B/D mmmfs sav. small time deposits (billion $)

fm3	money stock: M3 (M2 + long time deposits term RP's&INST only MMMFS) (billion $, SA)
fml	money stock: L (M3 + other liquid assets) (billion $, SA)
fmfba	monetary base, adjusted for reserve requirement change (million $, SA)
fmbase	monetary base, adjusted for reserve requirement changes (FRB of St. Louis) (billion $, SA)
fmrra	depository inst. reserves: total, adjusted for reserve requirement changes (million $, SA)
fmrnba	depository inst. reserves: nonborrowed, adjusted reserve requirement changes (million $, SA)
fmrnbc	depository inst. reserves: nonborrowed + external credit, adjusted reserve requirement changes (million $, SA)
fclbmc	weekly report LG commercial banks: net change commercial and industrial loans (billion $, SAAR)
fclnq	commercial and industrial loans oustanding, in 1992 dollars (BCI)
fm2dq	money supply – M2 in 1992 dollars (BCI)

Price indexes and wages

pmcp	NAPM commodity prices index (%)
pwfsa	producer price index: finished goods (1982 = 100, SA)
pwfcsa	producer price index: finished consumer goods (1982 = 100, SA)
pwimsa	producer price index: intermed materials supplies and components (1982 = 100, SA)
pwcmsa	producer price index: crude materials (1982 = 100, SA)
psm99q	index of sensitive materials prices (1990 = 100) (BCI-99A)
punew	CPI-U: all items (1982–1984 = 100, SA)
pu83	CPI-U: apparel and upkeep (1982–1984 = 100, SA)
pu84	CPI-U: transportation (82–84 = 100, SA)
puc	CPI-U: commodities (1982–1984 = 100, SA)
pucd	CPI-U: durables (1982–1984 = 100, SA)
pus	CPI-U: services (1982–1984 = 100, SA)
puxf	CPI-U: all items less food (1982–1984 = 100, SA)
puxhs	CPI-U: all items less shelter (1982–1984 = 100, SA)
puxm	CPI-U: all items less medical care (1982–1984 = 100, SA)
gmdc	PCE, impl pr defl: PCE (1987 = 100)
gmdcd	PCE, impl pr defl: PCE; durables (1987 = 100)
gmdcn	PCE, impl pr defl: PCE; nondurables (1987 = 100)
gmdcs	PCE, impl pr defl: PCE; services (1987 = 100)
lehcc	average hourly earnings of construction workers: construction ($, SA)
lehm	average hourly earnings of production workers: manufacturing ($, SA)

Misellaneous (Others)

hhsntn	University of Michigen index of consumer expectations (BCD-83)
pmdel	NAPM vendor deliveries index (%)

References

Chan, Y.L., Stock, J.H., Watson, M.W., 1999. A dynamic factor model for forecast combination. Spanish Economic Review 1 (2), 91–122.

Congressional Budget Office, 1994. Reestimating the NAIRU. In: The Economic and Budget Outlook, August.

Council of Economic Advisors, 1998. Economic Report of the President. US Government Printing Office, Washington, D.C.

Elliott, G., Rothenberg, T.J., Stock, J.H., 1996. Efficient tests for an autoregressive unit root. Econometrica 64, 813–836.

Estrella, A., Mishkin, F.S., 1997. Is there a role for monetary aggregates in the conduct of monetary policy?. Journal of Monetary Economics 40 (2), 279–304.

Fuhrer, J.C., 1995. The Phillips curve is alive and well. New England Economic Review of the Federal Reserve Bank of Boston, March/April, pp. 41–56.

Gordon, R.J., 1982. Price inertia and ineffectiveness in the United States. Journal of Political Economy 90, 1087–1117.

Gordon, R.J., 1997. The time-varying NAIRU and its implications for economic policy. Journal of Economic Perspectives 11, 11–32.

Gordon, R.J., 1998. Foundations of the Goldilocks economy: Supply shocks and the time-varying NAIRU. Brookings Papers on Economic Activity 2, 297–333.

Granger, C.W.J., Newbold, P., 1976. Forecasting economic time series. Academic Press, New York.

Hansen, B.E., 1997. Approximate asymptotic p-values for structural-change tests. Journal of Business and Economic Statistics 15 (1), 60–67.

Harvey, A.C., Jaeger, A., 1993. Detrending, stylized facts and the business cycle. Journal of Applied Econometrics 8 (3), 231–248.

Hodrick, R., Prescott, E., 1981. Post-war US business cycles: An empirical investigation. Working Paper, Carnegie-Mellon University; printed in: Journal of Money, Credit and Banking 29 (1997), 1–16.

Ireland, P.N., 1999. Does the time-consistency problem explain the behavior of inflation in the United States. Journal of Monetary Economics, this issue.

James, W., Stein, C., 1961. Estimation with quadratic loss. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley, pp. 361–379.

King, R.G., Rebelo, S.T., 1993. Low frequency filtering and real business cycles. Journal of Economic Dynamics and Control 17, 207–231.

King, R.G., Watson, M.W., 1994. The post-war US Phillips curve: A revisionist econometric history. Carnegie-Rochester Conference on Public Policy 41, 157–219.

Kuttner, K.N., 1994. Estimating potential output as a latent variable. Journal of Business and Economic Statistics 12 (3), 361–368.

Mishkin, F.S., 1990. What does the term structure tell us about future inflation?. Journal of Monetary Economics 25 (1), 77–96.

Quandt, R.E., 1960. Tests of the hypothesis that a linear regression system obeys two separate regimes. Journal of the American Statistical Association 55, 324–330.

Shimer, R., 1998. Why is the US unemployment rate so much lower? NBER Macroeconomics Annual, forthcoming.

Sims, C.A., 1992. Interpreting the macroeconomic time series facts: The effects of monetary policy. European Economic Review 36, 975–1011.

Staiger, D., Stock, J.H., Watson, M.W., 1997a. The NAIRU, unemployment, and monetary policy. Journal of Economic Perspectives 11, 33–51.

Staiger, D., Stock, J.H., Watson, M.W., 1997b. How precise are estimates of the natural rate of unemployment? in: Romer, C., Romer, D. (Eds.), Reducing Inflation: Motivation and Strategy, University of Chicago Press for the NBER, Chicago, pp. 195–242.

Stock, J.H., 1998. Comment on Gordon's 'Foundations of the Goldilocks economy: Supply shocks and the time-varying NAIRU'. Brookings Papers on Economic Activity 2, 334–341.

Stock, J.H., Watson, M.W., 1996. Evidence on structural instability in macroeconomic time series relations. Journal of Business and Economic Statistics 14, 11–30.

Stock, J.H., Watson, M.W., 1998. Diffusion indexes, NBER Working Paper.

West, K.D., 1996. Asymptotic inference about predictive ability. Econometrica 64, 1067–1084.

Zarnowitz, V., Braun, P., 1993. Twenty-two years of the NBER-ASA quarterly economic outlook surveys: Aspects and comparisons of forecasting performance. In: Stock, J.H., Watson, M.W. (Eds.), Business Cycles, Indicators, and Forecasting, NBER Studies in Business Cycles, Vol. 28. University of Chicago Press, Chicago.