

MONEY GROWTH AND INFLATION IN THE UNITED STATES

LANCE BACHMEIER

Kansas State University

SITTISAK LEELAHANON

Thammasat University

Qi Li

Texas A&M University

Specification tests reject a linear inflation forecasting model over the period 1959–2002. Based on this finding, we evaluate the out-of-sample inflation forecasts of a fully nonparametric model for 1994–2002. Our two main results are that: (i) nonlinear models produce much better forecasts than linear models, and (ii) including money growth in the nonparametric model yields marginal improvements, but including velocity reduces the mean squared forecast error by as much as 40%. A threshold model fits the data well over the full sample, offering an interpretation of our findings. We conclude that it is important to account for both nonlinearity and the behavior of monetary aggregates when forecasting inflation.

Keywords: Inflation Forecasting, Money Growth, Nonlinearity, Nonparametric

1. INTRODUCTION

It has become widely accepted that, for most purposes, changes in monetary aggregates are of little interest for the U.S. monetary policy process. This viewpoint is summarized nicely by the title of the paper by Leeper and Roush (2003): “Putting ‘M’ Back into Monetary Policy.” Numerous recent papers have presented evidence that money growth has no predictive power for inflation, and this finding is robust to changes in the sample period and econometric methodology.¹ Svensson and Woodford (2003) summarize the empirical literature, “Under normal circumstances, the information content of money growth for inflation forecasts in the short and medium term seems to be quite low. Only in the long run does a high correlation between money growth and inflation result.” The importance of these papers has grown, as inflation forecasting has come to occupy a central

Li’s research is partially supported by the Private Enterprise Research Center, Texas A&M University. We are grateful to two anonymous referees, an associate editor, and the editor, W.A. Barnett, for many insightful comments that have led to a much improved paper. We would also like to thank Ben Keen, Jeff Racine, and participants at the Southern Economic Association meetings for useful comments and discussions. Address correspondence to: Lance Bachmeier, Department of Economics, 327 Waters Hall, Kansas State University, Manhattan, KS 66502-4001, USA; e-mail: lanceb@ksu.edu.

place in monetary economics, both in practice and in theory. It is clear that the Federal Reserve relies on inflation forecasts when setting monetary policy [see, e.g., Clarida, et al. (2000)]. Additionally, a large academic literature has promoted the benefits of “inflation-forecast targeting,” whereby the central bank chooses a target inflation rate and adjusts its policy instrument to eliminate deviations of the expected inflation rate from the target [see, e.g., Bernanke and Woodford (2003), Svensson (2003), and Woodford (2005)].

This paper asks whether such a conclusion is warranted. Specifically, the papers cited earlier have focused on forecasts from linear vector autoregressive (VAR) models:

$$X_t = \alpha + \beta(L)X_{t-1} + \epsilon_t,$$

where $X_t = (\pi_t, \Delta m_t)'$, π_t is the inflation rate at time t , and Δm_t is the growth rate of a monetary aggregate. The popularity of VAR modeling arises from the fact that it is an atheoretical approach, and as such requires very few assumptions. The Wold Decomposition Theorem suggests that linear models are a good starting point [see Diebold (1998, pp. 179–180) for a related discussion]. If the system of interest is nonlinear, it may be better to directly estimate a nonlinear model rather than a linear approximation.

Aside from the obvious importance of correct specification in practical forecasting situations, the finding of a nonlinear relationship between inflation and fundamentals would have implications for VAR models of monetary policy. The price puzzle, a finding that tighter monetary policy is initially followed by a rise in the price level, is argued to be the result of information omitted from the Federal Reserve’s reaction function. “Solutions” to the price puzzle consist of adding variables to the model, such as commodity prices [see, e.g., Hanson (2004)]. Yet the omitted information also might take the form of a more complicated functional form. Predictable movements in the variables are interpreted as shocks, which can lead to a price puzzle in the same way that omitting relevant variables from the model leads to a price puzzle. A different literature has studied changes in the parameters of the Federal Reserve’s reaction function. It is common to include a measure of money growth in the reaction function, based on the argument that money growth has played an important role in monetary policy decisions. When the relationship between inflation and money growth is nonlinear, the reaction function will in general also be nonlinear. Boivin (2001) accounts for nonlinearities with a time-varying parameter VAR model, using an estimation strategy that accommodates the many estimated parameters in a VAR model.

Our baseline model is a fully nonparametric model that allows for any kind of nonlinearity in the relationship between money growth and inflation. We compare the linear univariate and bivariate VAR inflation forecasts that have been used in previous studies to their nonparametric counterparts. We then compare the forecast performance of the nonparametric models that include either money growth or velocity to autoregressive models, which provides a measure of the out-of-sample information content of these variables for inflation. Finally, we present evidence

that a threshold model captures the nonlinearity in inflation well, although a threshold model does not always forecast well out-of-sample. The idea that a parametric nonlinear model can fit the data well without providing large improvements in forecast performance is not new [see, e.g., Kilian and Taylor (2003) and Clements et al. (2004)].

A nonparametric approach should in principal be preferred to parametric approaches, because it is more general (the linear model is nested by the nonparametric model). As few other macroeconomic papers have attempted to exploit the gains from nonparametric modeling,² we now discuss some reasons why this may be so. First, the curse of dimensionality requires that only parsimonious models be considered. This limitation poses a special problem for macroeconomic models because of the importance, in many cases, of including multiple lags of the variables. Second, the computational burden of implementing a nonparametric approach with a data-driven method for selecting the smoothing parameters is nontrivial. A third disadvantage is that nonparametric models are less efficient when the data generating process can be approximated well by a linear model. The assumption of linearity is prevalent in out-of-sample macroeconomic forecasting exercises, presumably because of a lack of evidence to the contrary.³ Finally, it can be difficult to interpret the results obtained from a nonparametric regression. This is a problem when fitting prediction models, insofar as forecasters need to ask whether an estimated model has sensible properties and decide whether adjustments are necessary.

The paper proceeds as follows. Section 2 describes the data. Section 3 presents evidence on the importance of allowing for a flexible functional form, and discusses our findings on the information content of monetary aggregates for inflation. Section 4 summarizes our findings and suggests directions for future research.

2. DATA

All of the data series were downloaded from the St. Louis Federal Reserve bank Web site, and cover the period from January 1959 to May 2002. The monetary aggregate data we analyze are simple sum M1, M2, and M3, and the corresponding M1, M2, and M3 Divisia monetary services index data [see, e.g., Belongia and Chalfant (1989), Belongia (1996), and the collection of papers in Barnett and Serletis (2000) for discussion and empirical evidence on the advantages and disadvantages of using Divisia monetary aggregates]. The consumer price index series is used in place of other possible measures of the price level, because it is available at a monthly frequency, and has been studied widely in the literature. Use of data available at a quarterly frequency, such as the GDP deflator, would prohibit the use of nonparametric methods, because the sample size would be too small to allow for an informative out-of-sample forecast comparison. The velocity series is calculated as $V = PQ/M$, where PQ is the index of industrial production (nominal) and M is one of the six monetary aggregates.

3. EMPIRICAL RESULTS

As a benchmark, we first replicate for our updated data the well-known result that money growth is useless as an inflation indicator. Then we show how the results change when relaxing assumptions on functional form, and expand the models to include velocity. We finish by evaluating the in-sample fit and out-of-sample forecasts of threshold models. The advantage of these parametric nonlinear models is that they are easy to interpret.

3.1. Linear Models

Forecasts here and throughout this paper are made using a recursive estimation procedure with an increasing window of data, so that each forecast is based on a model estimated using only data available through the date that forecast would have been made. Because our first forecasted inflation rate is January 1994, the corresponding one-step ahead model was estimated using data available through December 1993. Observations for January 1994 were then added to the dataset, all models were reestimated, and one-step ahead forecasts were produced for the inflation rate in February 1994. The procedure was repeated to make a series of 100 forecasts for each model and forecast horizon, covering the period January 1994 through April 2002. Given the different forecast series and the observed historical inflation series, the mean squared prediction error (MSPE) was calculated for each model as $[\sum_{t=1994:1}^{2002:4} (\hat{\pi}_t - \pi_t)^2]/100$. The out-of-sample period was chosen for two reasons. First, the nonparametric models we look at below require a sufficiently large estimation sample. Second, this time period is more interesting than previous time periods. It has been argued that there may have been stable money demand relationships in earlier years but that they had broken down by the early 1990s. Good forecast performance in this time period would be an important finding.

Table 1 reports the MSPE of a linear model including money growth as a regressor:

$$\pi_t = \alpha_0 + \alpha_1 \pi_{t-s} + \alpha_2 \pi_{t-s-1} + \alpha_3 \Delta m_{t-s} + \alpha_4 \Delta m_{t-s-1} + \varepsilon_t,$$

relative to the autoregressive model:

$$\pi_t = \alpha_0 + \alpha_1 \pi_{t-s} + \alpha_2 \pi_{t-s-1} + \varepsilon_t, \quad (1)$$

where the forecast horizon is given by $s = 1, 6, 12,$ and 24 . To choose the lag length, the Schwarz information criterion (SIC) selected a VAR model with two lags of inflation and money growth.⁴ The statistics reported in Table 1 are the ratio of MSPE of one of the VAR models to the MSPE of the autoregressive model at that horizon, so that values less than one imply the VAR model forecasts are more accurate. p-values for the CCS test of out-of-sample Granger causality [Chao et al. (2001)] are reported in parenthesis below the relative MSPE statistics. The CCS test statistic is a measure of the correlation between the forecast errors of the AR model and the additional terms in the VAR model. As such the CCS test

TABLE 1. Relative MSPE (Linear VAR/Linear AR)

Horizon	M1	M2	M3	M1D	M2D	M3D
1 Month	1.04 (0.44)	0.99 (0.04)	1.04 (0.02)	1.00 (0.91)	1.00 (0.25)	1.00 (0.05)
6 Months	1.09 (0.55)	1.05 (0.00)	1.13 (0.00)	0.98 (0.95)	1.01 (0.07)	1.05 (0.01)
12 Months	0.90 (0.54)	0.93 (0.00)	1.04 (0.00)	0.98 (0.92)	0.97 (0.08)	1.01 (0.02)
24 Months	1.05 (0.13)	0.76 (0.00)	1.07 (0.00)	0.99 (0.34)	0.87 (0.02)	0.94 (0.01)

Notes: Entries in Table 1 are the MSPE of the linear VAR model divided by the MSPE of the linear AR model, for the period January 1994 to April 2002. Numbers in parentheses are p-values for the CCS test [Chao et al. (2001)] of no Granger causality from the money variable to inflation.

is not a direct comparison of the loss from the two forecasting models as is the DM test of Diebold and Mariano (1995) or West (1996), and the relative MSPE and CCS test results may sometimes contradict one another (for instance, the relative MSPE might be greater than one yet the CCS test rejects). The CCS test is closer to an encompassing test of whether the benchmark AR model contains all of the information in the VAR model. It is possible that one model may forecast better than another even if encompassing is rejected. Clark and McCracken (2001) propose a one-step ahead test of equal forecast accuracy of non-tested linear models, but no DM-type test is available for comparisons between nested models when the forecast horizon is greater than one.

At one-month and six-month forecast horizons, consistent with our prior expectations, any gains from including money are unimportant. In fact, the inefficiency associated with including money growth variables actually leads to as much as a 13% increase in MSPE! Interestingly, at the longer horizons (12 and 24 months) money growth does have value in some cases, contrary to the conclusions in recent studies, with an MSPE reduction in one case of 23%. Outside of three cases, however, any gains are small, with the VAR model offering little or no improvement over the AR model. Overall, the case for using money growth as an indicator variable for inflation is weak.

3.2. Relaxing Assumptions about Functional Form

We have confirmed the well-known result that inflation forecasts from linear VAR models are usually not more accurate than an autoregressive model. This section compares the linear inflation forecasting models earlier, used by previous authors (see the papers cited in the Introduction) to their nonparametric counterparts. The benchmark in each case is the s -step ahead linear model

$$\pi_t = \alpha + \beta(L)x_{t-s} + \varepsilon_t, \quad (2)$$

with $x_{t-s} = \pi_{t-s}$ for an AR model, and $x_{t-s} = (\pi_{t-s}, \Delta m_{t-s})$ for a VAR model, where m_t can be any one of the six monetary aggregates: $M1_t$, $M2_t$, $M3_t$, $M1D_t$, $M2D_t$, and $M3D_t$. Forecasts for each linear model are compared to those of a general nonparametric model

$$\pi_t = f(\pi_{t-s}, \pi_{t-s-1}, \Delta m_{t-s}) + \varepsilon_t. \quad (3)$$

The nonparametric model relaxes any assumptions about functional form, so that the only restrictions in equation (3) are the variables included in x and the lag length. This model encompasses all nonlinear models that have been proposed, including threshold, smooth transition, and Markov switching models [see, e.g., Granger and Teräsvirta (1993) or Hamilton (1994)]. By contrast, even if equation (2) is not correctly specified, as is almost certainly the case, it might still provide a better approximation than equation (3) in practice. The nonparametric model converges more slowly than the linear model, so that there is no a priori reason to expect one model to forecast better out-of-sample. In fact, for a nonparametric model to be of use with a sample of several hundred observations, which is true for this paper, it is necessary that the linear model be severely misspecified.

We estimate the nonparametric models using the Nadaraya-Watson kernel estimator [Nadaraya (1965), Watson (1964)]. The forecast of inflation at time $T+s$ is given by

$$E(\pi_{T+s} | \pi_T, \pi_{T-1}, \Delta m_T) = \frac{\sum_{t=1}^{T-s} \pi_{t+s} K[(x_t - x_T)/h]}{\sum_{t=1}^{T-s} K[(x_t - x_T)/h]}, \quad (4)$$

where $x_t = (\pi_t, \pi_{t-1}, \Delta m_t)$ and $K(\cdot)$ is the product normal kernel function. A practical difficulty associated with any nonparametric estimation is the choice of bandwidth h , and the recursive nature of our analysis makes matters more difficult, as we need to choose the bandwidth thousands of times. Each time a forecast was made, out-of-sample inflation forecasts were calculated for the previous 50 observations using many different bandwidth choices, and we set h equal to the value that yielded the lowest MSPE for those 50 forecasts. For example, to make a one-step ahead forecast of inflation for January 1995, we use the value of h that produced the best forecasts over the period November 1990 to December 1994. There is little theoretical guidance as to the selection of bandwidth for out-of-sample forecasts. The intuition behind our procedure is that we are interested in producing out-of-sample forecasts, so we should use the bandwidth that has produced the best forecasts in the past. To the extent that this procedure is not optimal, our nonparametric forecasts can be improved further.

Table 2 offers a comparison of the linear and nonlinear models. DM test statistics are reported in parenthesis.^{5,6} In nearly every case, the nonparametric model does better. The only way we can observe the nonparametric models consistently outperforming linear models with the same regressors is if there is a relationship between money growth and inflation, and if that relationship is far from linear. In the Introduction, we cited two motivations for this paper—the inflation forecast-targeting literature, and the monetary VAR literature. The results in

TABLE 2. Relative MSPE (Nonparametric VAR/Linear VAR)

Horizon	AR	M1	M2	M3	M1D	M2D	M3D
1 Month	0.88 (1.75)	0.94 (0.67)	0.84 (2.38)	0.84 (2.18)	0.88 (1.69)	0.87 (1.83)	0.86 (1.89)
6 Months	0.93 (2.12)	0.86 (1.15)	0.86 (2.43)	0.84 (2.22)	0.90 (2.08)	0.91 (1.92)	0.88 (2.22)
12 Months	0.93 (2.21)	0.99 (0.32)	0.91 (2.26)	0.88 (2.40)	0.94 (1.66)	0.95 (1.23)	0.91 (2.12)
24 Months	1.09 (-5.84)	1.00 (0.00)	0.94 (0.80)	0.91 (1.08)	1.07 (-3.91)	1.10 (-1.88)	1.05 (-0.83)

Notes: Entries in Table 2 are the MSPE of the nonparametric AR model of inflation divided by the MSPE of the linear AR model of inflation, for the period January 1994 to April 2002. Numbers in parentheses are DM test statistics [Diebold and Mariano (1995)].

Table 2 are most relevant for the latter, as they show that a linear model omits relevant information, and may cause expected movements to be mistakenly labeled as shocks, resulting in a price puzzle. Given that it is not straightforward to modify structural VAR models to incorporate nonparametric estimation, and that there are controversies over which variables to include, changes in the policy rule, and so on, we leave this to future work.

Table 3 reports the MSPE of each nonparametric model with money growth relative to the nonparametric AR(2) model at each horizon. There are two things to note here. First, the VAR model usually does better than the AR model, with the MSPE ratio in most cases less than one. This is quite different from Table 1, where the MSPE ratio was greater than 1 in 12 cases, especially in light of the fact that the nonparametric AR model is already outperforming the linear AR model. Second, most of the gains appear to be a result of allowing for a nonlinear functional form, rather than from the inclusion of money growth.

TABLE 3. Relative MSPE (Nonparametric VAR/Nonparametric AR)

Horizon	M1	M2	M3	M1D	M2D	M3D
1 Month	1.12 (-0.97)	0.94 (1.78)	1.00 (0.06)	1.00 (0.01)	0.99 (0.46)	0.99 (0.78)
6 Months	1.00 (0.05)	0.98 (0.88)	1.02 (-0.66)	0.94 (1.80)	0.98 (0.79)	1.00 (0.24)
12 Months	0.95 (1.20)	0.90 (2.76)	0.98 (0.43)	0.99 (0.49)	0.99 (0.40)	0.98 (0.93)
24 Months	0.96 (0.69)	0.66 (6.43)	0.89 (1.10)	0.98 (1.46)	0.88 (4.33)	0.91 (1.86)

Notes: Table 3's entries are the MSPE of the nonparametric model of inflation containing money divided by the MSPE of the nonparametric AR model of inflation, for the period January 1994 to April 2002. Numbers in parentheses are DM test statistics [Diebold and Mariano (1995)].

For the nonparametric AR model, we use the same number of lags (two lags) as that used in the linear AR model. Using the same set of explanatory variables enables us to determine whether relaxing the linear functional form is responsible for the improved forecasts. Gao and Tong (2004) have recently proposed a procedure for selecting the number of lags in a nonparametric AR model framework. For the data we used, Gao and Tong’s procedure for selecting the number of lags in the nonparametric model led to a slight improvement in the forecasting accuracy. In the remaining part of the paper, we continue to consider the cases of using two lags in the nonparametric AR model, and adding one lag of money growth (in Section 3.5, a lag of velocity) for the nonparametric VAR model to better focus on the effects of functional form.

3.3. Specification Tests

These forecasting results suggest that the linear AR and VAR models are misspecified. In this section we formally test the correctness of linear AR and VAR models. We test the null hypothesis of a linear specification against a general nonparametric (nonlinear) model. The consistent model specification test proposed by Fan and Li (1996), Li (1999), and Zheng (1996) will be used to test the following null hypotheses for the inflation model: (i) a linear AR model: $\pi_t = \beta_0 + \beta_1\pi_{t-s} + \beta_2\pi_{t-s-1} + u_t$ against a nonparametric AR model: $\pi_t = g(\pi_{t-s}, \pi_{t-s-1}) + u_t$; (ii) a linear VAR model, $\pi_t = \beta_0 + \beta_1\pi_{t-s} + \beta_2\pi_{t-s-1} + \beta_3\Delta m_{t-s} + u_t$ against a nonparametric VAR model: $\pi_t = g(\pi_{t-s}, \pi_{t-s-1}, \Delta m_{t-s}) + u_t$; and (iii) a nonparametric AR model: $\pi_t = g(\pi_{t-s}, \pi_{t-s-1}) + u_t$ against a nonparametric VAR model: $\pi_t = g(\pi_{t-s}, \pi_{t-s-1}, \Delta m_{t-s}) + u_t$.

Briefly, the testing procedure is implemented in the following manner. Start by introducing some notation: $w_t = (\pi_{t-s}, \pi_{t-s-1})$, $z_t = \Delta m_{t-s}$. Then the above hypotheses can be tested based on $E(u_t | x_t) = 0$, where $u_t = \pi_t - \beta_0 - w_t\beta$ and $x_t = w_t$ for (i); $u_t = \pi_t - \beta_0 - w_t\beta - z_t\gamma$ and $x_t = (w_t, z_t)$ for (ii); $u_t = \pi_t - g(w_t)$ and $x_t = (w_t, z_t)$ for (iii). The test statistic proposed by Fan and Li (1996), and Zheng (1996) is a kernel estimate of $I = E[u_t E(u_t | x_t) f(x_t)]$. This is because $I = E\{[E(u_t | x_t)]^2 f(x_t)\} \geq 0$, and $I = 0$ if and only if the null hypothesis is true. Therefore, I serves as a proper candidate for testing the null hypothesis of $E(u_t | x_t) = 0$. A feasible test statistic is given by

$$\hat{I}_n = \frac{1}{n(n-1)} \sum_{t=1}^n \sum_{s \neq t}^n \hat{u}_t \hat{u}_s K_{h,ts},$$

where \hat{u}_t is the residual (estimated error) from the null model, $K_{h,ts} = \prod_{j=1}^d h_j^{-1} k[(x_{s,j} - x_{t,j})/h_j]$ is the product kernel function, and h_j is the smoothing parameter associated with x_j ($j = 1, \dots, d$), d is the dimension of x_t ($d = 2$ for (i), and $d = 3$ for (ii) and (iii)). A standardized test is given by

$$\hat{J}_n = n(h_1 \dots h_d)^{1/2} \hat{I} / \sqrt{\hat{\sigma}_0^2},$$

where $\hat{\sigma}_0^2 = h_1 \cdots h_d / n(n-1) \sum_t \sum_{s \neq t} \hat{u}_t^2 \hat{u}_s^2 K_{h,t,s}^2$. Under the null hypothesis and some regularity conditions, \hat{J}_n has an asymptotic standard normal distribution.

Li and Wang (1998), and Hsiao and Li (2001) show that in finite sample applications, the \hat{J}_n test is significantly undersized (there is finite sample negative bias under the null hypothesis). They recommend using bootstrap procedures to better approximate the finite sample null distribution of the test statistic \hat{J} . Hsiao and Li (2001) proposed a bootstrap procedure for time series data. We adopt the bootstrap procedure suggested by Hsiao and Li (2001) to obtain the critical values for the test statistic \hat{J}_n . The number of bootstrap replications is 1,000. The bootstrap critical values differ for each case, so rather than reporting all of the test statistics and bootstrap critical values, we summarize the testing results.

For case (i) of testing a linear AR model, we reject the null of a linear AR model for all $t = 1, 6, 12, 24$ at the 5% level based on bootstrap critical values.

For case (ii) of testing a linear VAR model, we reject the null of a linear VAR model for all money growth models, and for $s = 1, 6, 12, 24$ at the 5% level based on bootstrap critical values.

For case (iii) of testing a nonparametric AR model, the results are mixed. For $s = 1$, we *do not* reject the null hypothesis of a nonparametric AR model at the 5% level for all money growth models. For $s = 6$, we reject the null hypothesis of a nonparametric AR model at the 5% level for $m = M1$ and $M3$, but we *do not* reject the null hypothesis at the 5% level for $m = M2, M1D, M2D$ and $M3D$. For $s = 12$, we reject the null hypothesis of a nonparametric AR model at the 5% level for $m = M2$ and $M3$, but we *do not* reject the null hypothesis at the 5% level for $m = M1, M1D, M2D$ and $M3D$. For $s = 24$, we reject the null hypothesis of a nonparametric AR model at the 5% level for $m = M1, M2, M2D$, and $M3D$, but we *do not* reject the null hypothesis at the 5% level for $m = M3$ and $M1D$.

These in-sample testing results are largely consistent with the out-of-sample prediction results. The linear AR and linear VAR models are strongly rejected, and we conclude that there is significant nonlinearity in the relationship between inflation and its lagged values, and between inflation and money growth. Also, there is some weak evidence that a nonparametric VAR model fits better than a nonparametric AR model in in-sample fit. Our primary interest is in producing out-of-sample forecasts, so in the next section we compare forecast performance of a nonparametric model with velocity replacing the money growth variable.

3.4. Forecasting Inflation with Velocity

These results suggest the importance of functional form, but with a few exceptions, out-of-sample causality from money growth to inflation is still hard to find. There are, however, alternative ways to incorporate information on the behavior of monetary aggregates into the model. We now evaluate the information content of velocity, motivated by the P* inflation forecasting model that has been studied by many authors [see, e.g., Gerlach and Svensson (2003) and the references contained

TABLE 4. Relative MSPE (Nonparametric VAR/Parametric VAR)

Horizon	V1	V2	V3	V1D	V2D	V3D
1 Month	0.74 (2.24)	0.72 (3.04)	0.88 (1.62)	0.77 (2.05)	0.76 (2.42)	0.77 (2.10)
6 Months	0.84 (1.62)	0.90 (1.29)	0.93 (1.42)	0.87 (1.56)	0.85 (1.84)	0.84 (2.02)
12 Months	0.73 (3.10)	0.88 (1.77)	0.90 (2.07)	0.75 (2.87)	0.76 (3.56)	0.79 (3.14)
24 Months	0.83 (2.00)	0.82 (1.29)	1.20 (-3.14)	0.99 (0.09)	0.98 (0.54)	0.99 (0.16)

Notes: Table 4's entries are the MSPE of the nonparametric model of inflation containing velocity divided by the MSPE of the linear VAR model of inflation containing velocity, for the period January 1994 to April 2002. Numbers in parentheses are DM test statistics [Diebold and Mariano (1995)].

therein]. Despite the lack of a formal theoretical basis for the P^* model [Gerlach and Svensson (2003)], it is nevertheless popular as a tool for forecasting inflation. In keeping with our use of monthly data, the measure of output is industrial production.

Table 4 shows that the forecast improvements from using nonparametric VAR models versus linear VAR models are more pronounced than for money growth. With only one exception, the nonparametric (nonlinear) VAR model always outperforms its linear counterpart, and in many cases the differences are substantial. The *only* difference between the models is that the linear model imposes restrictions on functional form, suggesting that it is crucial to allow for nonlinearity when estimating P^* -type models. To the best of our knowledge, this is the first paper to demonstrate important nonlinearities in the (out-of-sample) P^* inflation forecasting model. Table 5 is particularly interesting. Our earlier results suggested only modest improvements in the forecast performance of a nonparametric VAR

TABLE 5. Relative MSPE (Nonparametric VAR/Nonparametric AR)

Horizon	V1	V2	V3	V1D	V2D	V3D
1 Month	0.94 (0.74)	0.90 (1.85)	0.99 (0.16)	0.97 (0.43)	0.87 (2.25)	0.87 (2.03)
6 Months	1.01 (-0.14)	0.93 (0.94)	0.93 (1.62)	1.03 (-0.43)	0.84 (2.01)	0.84 (2.29)
12 Months	0.93 (0.85)	0.83 (2.41)	0.83 (4.17)	0.93 (0.80)	0.71 (3.87)	0.71 (4.02)
24 Months	0.83 (2.57)	0.63 (4.12)	0.70 (3.31)	1.00 (-0.01)	0.60 (4.60)	0.59 (4.79)

Notes: Table 5's entries are the MSPE of the nonparametric model of inflation containing money divided by the MSPE of the nonparametric AR model of inflation, for the period January 1994 to April 2002. Numbers in parentheses are DM test statistics [Diebold and Mariano (1995)].

model over the autoregressive benchmark when money growth is included. When money growth is replaced by velocity, we find strong evidence of causality. Even at a one-month forecast horizon, the velocity of three of the aggregates reduces the MSPE by 10% or more. The forecast improvement grows with the forecast horizon, and at a 24-month horizon including velocity of Divisia M2 or Divisia M3 reduces the MSPE by 40%.

Note that the relative MSPE comparison between the nonparametric VAR model with velocity and the nonparametric VAR model with money growth can be obtained as a ratio of the results of Table 5 to that of Table 3 (as they both have the MSPE of the nonparametric AR model in the denominator). We observe that the nonparametric VAR models with velocity generally perform much better (have a smaller MSPE) than the nonparametric VAR models with money growth.

To sum up our results, we have shown that the relationship between money growth and inflation is nonlinear, with a linear VAR model forecasting no better (and usually worse) than an autoregressive benchmark, and the nonparametric model producing small forecasting gains. Tables 4 and 5 show that the relationship between inflation and velocity is also nonlinear, and when we relax the strict assumptions of the linear model, velocity serves as a very important source of information about inflation. This suggests that the P^* model is still useful for predicting U.S. inflation.

We also have carried out in-sample specification tests for testing (i) a parametric VAR model (with velocity) versus a nonparametric VAR model, and (ii) a nonparametric AR model versus a nonparametric VAR model (with velocity). The test statistic is the same as discussed in Section 3.3 with \hat{u}_i being the estimated error from the null model, $x_t = (\pi_{t-s}, \pi_{t-s-1}, v_{t-s})$, where $v = V1, V2, V3, V1D, V2D, V3D$. The testing results are as follows.

For case (i) of testing a linear VAR, we reject the null hypothesis of a linear VAR model for all $s = 1, 6, 12$, and 24 at the 5% level based on bootstrap critical values (for all measures of velocity). For case (ii) of testing a nonparametric AR model versus a nonparametric VAR model, the results are mixed. For $s = 1$, we reject the null for all measures of velocity. For $s = 6$, we reject the null for $V1, V2, V3, V2D$, and $V3D$, but we do not reject the null with $V1D$. For $s = 12$ and $s = 24$, we reject the null for all measures of velocity. Thus, the in-sample testing results suggest that there is significant nonlinear interaction between inflation and money growth.

One difficulty with using a nonparametric estimation approach is that the interpretation of the model is more difficult. We therefore consider next the forecasts from a parametric nonlinear model, the threshold model.

Threshold autoregressive (TAR) models [see, e.g., Hansen (1996, 2000)] are similar to the linear AR but allow for multiple regimes. In the leading case of two regimes, the TAR model can be written

$$y_t = \begin{cases} \alpha_0 + \alpha_1\pi_{t-s} + \alpha_2\pi_{t-s-1} & \text{if } q_{t-s} \geq c \\ \beta_0 + \beta_1\pi_{t-s} + \beta_2\pi_{t-s-1} & \text{if } q_{t-s} < c, \end{cases}$$

where q_{t-s} is a “switching variable” observable at the time a forecast is made, equal to money growth or velocity, and c is the “threshold.” It is straightforward to compute multivariate threshold forecasts by simply adding additional variables to each equation. When q represents money growth, there is an intuitive interpretation of the TAR model. During periods of high money growth (such as during the 1970s) the money supply might be a dominant factor in the determination of the price level, whereas in other periods money growth is only a minor factor in the determination of the price level. There is no reason to believe that inflation will behave in the same way across the two regimes. When q represents velocity, a threshold model can be motivated by observing that changes in velocity may be reflecting changes in the inflation generating process itself, possibly due to differences in the predictability of inflation in different regimes.

We have examined both full sample tests for threshold nonlinearity as well as the out-of-sample forecast gains from using a threshold model. A number of published papers have addressed the choice between in-sample testing and out-of-sample forecast comparison when the goal is to evaluate theories (i.e., to explain the data rather than just produce forecasts). Authors such as West (1996) begin with the assumption that out-of-sample forecast evaluation is of interest, without providing any underlying motivation for evaluating the forecasts, whereas authors such as Kilian and Taylor (2003) and Clements et al. (2004) have provided reasons why an estimated nonlinear model may not forecast as well as simple benchmark models, even when the true data generating process is the nonlinear model being used to compute forecasts. Potter (1999) gives reasons why it is difficult to forecast using threshold models, because small errors in classification of observations into regimes can dramatically increase the mean squared error of the forecasts.

We do not take sides in this debate. Instead, we consider both approaches to model evaluation. If in-sample tests uncover evidence favoring a nonlinear specification, or if the out-of-sample forecasts of that specification are better than those of a linear model, the nonlinear model deserves further consideration. The main difficulty with tests for threshold nonlinearity is that the value of the threshold is not identified under the null hypothesis. Hansen (1996, 2000) has provided a simulation procedure for inference that controls the size of the test.

Our findings for the threshold model can be summarized as follows. In-sample tests, following the strategy outlined by Hansen (1996), *do not* reject linearity for any of the monetary aggregates, but *do* reject in all cases for each of the six velocity variables. This is strong evidence that the large improvements in forecast performance from including velocity in the nonparametric model (Tables 4 and 5) is a result of a simple type of nonlinearity. In contrast, the threshold model seldom yields better out-of-sample forecast performance than the linear AR model, even when velocity is used as the switching variable. This might be explained by problems with classifying observations into different regimes as described by Potter (1999). Tables with detailed results can be obtained from the authors on request.

4. CONCLUSION

This paper has evaluated the performance of nonparametric forecasting models of inflation. When relaxing assumptions about functional form, money growth and velocity contain information about inflation for horizons as short as one month. It is particularly interesting that the forecast improvement from including velocity is so large—over 40% in some cases. We have found these results in spite of the fact that we focus on the period from 1994 to 2002, a time period for which it is widely believed that the demand for money was unstable. Our results also suggest that the nonlinearity may be captured by a threshold model, even though the threshold model often does not provide good forecasts, possibly because of difficulties in assigning observations to one of the regimes. We conclude that arguments that the Federal Reserve can learn nothing by monitoring the behavior of monetary aggregates may be premature, as they are based on analysis of an overly restrictive set of models.

As suggested in the Introduction, our findings of nonlinearity have important implications for VAR models of monetary policy. More generally, our results indicate that nonparametric methods are sometimes a useful tool for macroeconomic forecasting, with the benefits of relaxing assumptions on functional form substantially outweighing any efficiency losses. It would be worthwhile to see whether our findings for inflation carry over to forecasts of other macroeconomic variables. The univariate results of Stock and Watson (1999b) may provide a useful starting point.

NOTES

1. See, for example, Leeper and Roush (2002, 2003), Stock and Watson (1999a), and the references contained therein. Recent empirical papers on inflation forecasting include, among many others, Clark and McCracken (in press), Gerlach and Svensson (2003), and Stock and Watson (2003). See these papers for an extensive list of references.

2. See, for example, Diebold and Nason (1990). This is one of the few papers to have analyzed nonparametric forecasts of macroeconomic variables, and the conclusion of the paper was actually that nonparametric models are not useful. Note that we are referring specifically to nonparametric forecasts here (as opposed to other types of nonlinear models).

3. For examples of in-sample nonlinearity tests, see, for example, Michael, Nobay, and Peel (1997), Taylor (2001), and Hamilton (2003). Granger (2001) concludes that the evidence for nonlinearity in macroeconomic data is weak, and finds it troubling that few of the papers he reviews look at forecast performance. Diebold (1998) takes a more pessimistic view, arguing that nonlinearities “require large amounts of high quality data” and that many nonlinearities “simply don’t appear to be important in macroeconomics.” See Clements et al. (2004) for additional discussion. Chen, Racine, and Swanson (2001) find some evidence of nonlinearity in U.S. inflation using a neural network-based semiparametric model with 1948 to 1995 quarterly data.

4. An alternative would be to select the lag length each time a forecast is made [see, e.g., Stock and Watson (2003)]. Given that we use an increasing window of data for estimation, and that the number of observations ranges from approximately 400 for the initial forecasts to 500 for the final forecast, the optimal lag length choice will not change much through time.

5. It is known that the DM statistic does not have an asymptotic normal distribution when the models are nested [see, e.g., Clark and McCracken (2001)]. We nevertheless view it to be both a useful available alternative and better than not reporting any significance tests.

6. Corradi and Swanson (2002, 2004) have developed a formal test for out-of-sample nonlinear predictive accuracy, but their test does not allow for comparison of nonparametric models. Fan and Li's tests (1996) allow for comparison of nonparametric models, but they only consider the case of in-sample tests.

REFERENCES

- Barnett, William A. and Apostolos Serletis (2000) *The Theory of Monetary Aggregation*. New York: Elsevier Science, North-Holland.
- Belongia, Michael T. (1996) Measurement matters: Recent results from monetary economics reexamined. *Journal of Political Economy* 104, 1065–1083.
- Belongia, Michael T. and James A. Chalfant (1989) The changing empirical definition of money: Some estimates from a model of the demand for money substitutes. *Journal of Political Economy* 97, 387–397.
- Bernanke, Ben S. and Michael Woodford (2005) *Inflation Targeting*. Chicago: University of Chicago Press.
- Boivin, Jean (2001) The Fed's Conduct of Monetary Policy: Has It Changed and Does It Matter? Working Paper, Columbia University.
- Chao, John, Valentina Corradi, and Norman R. Swanson (2001) An out of sample test for Granger causality. *Macroeconomic Dynamics* 5, 598–620.
- Chen, Xiaohong, Jeffrey Racine, and Norman R. Swanson (2001) Semi-parametric ARX neural-network models with an application to forecasting inflation. *IEEE Transactions on Neural Networks* 12, 674–683.
- Clarida, Richard, Jordi Gali, and Mark Gertler (2000) Monetary policy rules and macroeconomic stability: Evidence and some theory. *Quarterly Journal of Economics* CXV, 147–180.
- Clark, Todd E. and Michael W. McCracken (2001) Tests of equal forecast accuracy and encompassing for nested models. *Journal of Econometrics* 105, 85–110.
- Clark, Todd E. and Michael W. McCracken (in press) The predictive content of the output gap for inflation: Resolving in-sample and out-of-sample evidence. *Journal of Money, Credit, and Banking*.
- Clements, Michael P., Phillip Hans Franses, and Norman R. Swanson (2004) Forecasting economic and financial time-series with non-linear models. *International Journal of Forecasting* 20, 169–183.
- Corradi, Valentina and Norman R. Swanson (2002) A consistent test for nonlinear out of sample predictive accuracy. *Journal of Econometrics* 110, 353–381.
- Corradi, Valentina and Norman R. Swanson (2004) Nonparametric Bootstrap Procedures for Predictive Inference Based on Recursive Estimation Schemes. Working Paper, Rutgers University.
- Diebold, Francis X. (1998) The past, present, and future of macroeconomic forecasting. *Journal of Economic Perspectives* 12, 175–192.
- Diebold, Francis X. and Roberto S. Mariano (1995) Comparing predictive accuracy. *Journal of Business and Economic Statistics* 13, 253–263.
- Diebold, Francis X. and James A. Nason (1990) Nonparametric exchange rate prediction? *Journal of International Economics* 28, 315–332.
- Fan, Yanqin and Qi Li (1996) Consistent model specification tests: Omitted variables and semiparametric functional forms. *Econometrica* 64, 865–890.
- Gao, Jiti and Howell Tong (2004) Semiparametric nonlinear time series model selection. *Journal of Royal Statistics Series (B)* 66, 321–336.
- Gerlach, Stefan and Lars E.O. Svensson (2003) Money and inflation in the Euro area: A case for monetary indicators? *Journal of Monetary Economics* 50, 1649–1672.
- Granger, Clive W.J. (2001) Overview of nonlinear macroeconomic empirical models. *Macroeconomic Dynamics* 5, 466–481.
- Granger, Clive W.J. and Timo Teräsvirta (1993) *Modelling Non-Linear Economic Relationships*. New York: Oxford University Press.
- Hamilton, James D. (1994) *Time Series Analysis*. Princeton, N.J.: Princeton University Press.

- Hamilton, James D. (2003) What is an oil shock? *Journal of Econometrics* 113, 363–398.
- Hansen, Bruce E. (1996) Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* 64, 413–430.
- Hansen, Bruce E. (2000) Sample splitting and threshold estimation. *Econometrica* 68, 575–603.
- Hanson, Michael S. (2004) The “price puzzle” reconsidered. *Journal of Monetary Economics* 51, 1385–1413.
- Hsiao, Cheng and Qi Li (2001) A consistent test for conditional heteroskedasticity in time series regression models. *Econometric Theory* 17, 188–221.
- Kilian, Lutz and Mark P. Taylor (2003) Why is it so difficult to beat the random walk forecast of exchange rates? *Journal of International Economics* 60, 85–107.
- Leeper, Eric M. and Jennifer E. Roush (2002) Putting “M” back into monetary policy. NBER Working Paper 9552.
- Leeper, Eric M. and Jennifer E. Roush (2003) Putting “M” back into monetary policy. *Journal of Money, Credit, and Banking* 35, 1217–1256.
- Li, Qi (1999) Consistent model specification tests for time series econometric models. *Journal of Econometrics* 92, 101–147.
- Li, Qi and Suojin Wang (1998) A simple bootstrap test for a parametric regression functional form. *Journal of Econometrics* 87, 145–165.
- Michael, Panos, A. Robert Nobay, and David A. Peel (1997) Transactions costs and nonlinear adjustment in real exchange rates: An empirical investigation. *Journal of Political Economy* 105, 862–879.
- Nadaraya, E.A. (1965) On estimating regression. *Theory of Probability and its Applications* 9, 141–142.
- Potter, Simon M. (1999) Nonlinear time series modelling: An introduction. *Journal of Economic Surveys* 13, 505–528.
- Stock, James H. and Mark W. Watson (1999a) Forecasting inflation. *Journal of Monetary Economics* 44, 293–335.
- Stock, James H. and Mark W. Watson (1999b) A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. In R. Engle and H. White (eds.), *Cointegration, Causality and Forecasting: A Festschrift in Honour of Clive W.J. Granger*, pp. 1–44. Oxford: Oxford University Press.
- Stock, James H. and Mark W. Watson (2003) Forecasting output and inflation: The role of asset prices. *Journal of Economic Literature* XLI, 788–829.
- Svensson, Lars E.O. (2003) What is wrong with Taylor rules? Using judgment in monetary policy through targeting rules. *Journal of Economic Literature* 41, 426–477.
- Svensson, Lars E.O. and Michael Woodford (2003) Indicator variables for optimal policy. *Journal of Monetary Economics* 50, 691–720.
- Taylor, Alan M. (2001) Potential pitfalls for the purchasing-power parity puzzle? Sampling and specification biases in mean-reversion tests of the law of one price. *Econometrica* 69, 473–498.
- Watson, G.S. (1964) Smooth regression analysis. *Sankhya Series A* 26, 359–372.
- West, Kenneth W. (1996) Asymptotic inference about predictive ability. *Econometrica* 64, 1067–1084.
- Woodford, Michael (2003) *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton, N.J.: Princeton University Press.
- Zheng, John Xu (1996) A consistent test of functional form via nonparametric estimation technique. *Journal of Econometrics* 75, 263–289.