Guido Ascari
University of Pavia

Emanuela Marrocu
University of Cagliari and CRENoS

FORECASTING INFLATION: A COMPARISON OF LINEAR PHILLIPS CURVE MODELS AND NONLINEAR TIME SERIE MODELS

Abstract

The aim of this paper is to analyze the forecasting performance of alternative model for the US inflation rate over the period 1950.1-2002.7. NAIRU Phillips curve models forecasts are contrasted with those obtained by a special class of nonlinear time series models, the threshold autoregressive models. The forecast evaluation is conducted on point and density forecasts. The results show that overall the non linear specification are better able to capture the distributional features of the series, although in terms of MSFE the Phillips curve specification can yield noticeable forecasting gains for medium and long term horizons. Previous finding on the forecasting superiority of the simple naïve model are confuted.

Keywords: Forecasting, inflation, threshold models, Phillips curve.

JEL classifications: C22, C53, E31

Corresponding author: Emanuela Marrocu, Dipartimento di Ricerche Economiche e Sociali, Università di Cagliari, Viale S. Ignazio, 78 – 09123 Cagliari – Italy
Email: emarrocu@unica.it

Acknowledgements: Financial support from CNR-Agenzia 2000 (cod. CNRG000721) is gratefully acknowledged.

1. Introduction

The availability of a reliable and accurate econometric-quantitative method for forecasting the behaviour of inflation is of primary importance, given the emphasis that the European Central Banks – in particular the European one - pose on to price stability. In the literature there is no consensus and the debate is very lively as emphasised for example by the recent symposium on "The Return of the Phillips Curve" in The Journal of Monetary Economics, 1999. Even more important, and maybe more difficult, is to build a quantitative model to forecast inflation, in order to give guidance to the monetary policy authorities.

The traditional approach to the analysis of inflation dynamics is based on the Phillips Curve and on the concept of non-accelerating-inflation-rate-of-unemployment (NAIRU). The fundamental relation is estimated by regressing the inflation change on a constant, its own lagged values, lagged unemployment rate, on a set of exogenous control variables and supply shock variables. Different versions of a regression of the type described above have been often used in the literature both to estimate the level of the NAIRU and to describe and forecast the dynamics of inflation (see Staiger et al., 1997, and reference therein).

From a theoretical perspective, both the Phillips Curve and the concept of NAIRU have been the subject of very well known critiques. Besides, also from an empirical perspective, both the stagflation of the '70s and the simultaneous occurrence in the second half of the '90s in US of low and falling inflation and low unemployment (Brayton et al., 1999) appear to be at odds with the properties of a standard Phillips Curve.

With respect to the theoretical modelling, new models have been proposed. These models are based on the synthesis between typical elements of the real business cycle literature (microfoundations and intertemporal optimisation) and typical elements of the new-keynesian literature, such as monopolistic competition and nominal rigidities (Ascari, 2000). In particular, by introducing in a stochastic dynamic general equilibrium model the staggering of price decisions by monopolistic competitive firms, the so-called *New-Keynesian Phillips Curve* (Galí-Gertler, 1999) is obtained: the inflation rate depends on marginal cost and on the expected inflation rate for next period. Under some conditions it exists a log-linear relationship between marginal costs and the output gap. Thus including the latter in the expression for the inflation rate yields a form similar to a traditional Phillips Curve. The main and substantial difference between a standard Phillips Curve and the *New-Keynesian Phillips Curve* rests on the forward-looking nature of the latter. Current inflation depends on future expected inflation and not from the previous inflation level. The attempts to use such formulation for an empirical model of inflation dynamics, however, have encountered serious problems, substantially due to the strong persistence properties of the inflation time-series.

From an empirical perspective and following the traditional Phillips Curve approach, Staiger et al. (1997) use a specification that allows for time variation in the NAIRU. However, one questions the value added and significance of a concept as NAIRU, if this sort of "natural" rate shifts largely in time. The suspect is that the shifts in the NAIRU are necessary to catch the strong persistence and non-linearity (or structural instability) typical of the inflation time series (Stock-Watson, 1996). As suggested by Stock and Watson (1996), the class of linear and log-linear models reported above proved inadequate to capture inflation dynamics, due to the high degree of non-linearity and to the presence of "jumps" in correspondence with some crucial dates and historical episodes.

This aspects of inflation behaviour have been also documented for other countries. With reference to the Italy, for example, Gallo and Otranto (1997) stress how the previous works, which analysed the dynamics of inflation in Italy through the use of the VARs, had to conclude that the problem of regime shifts will cause great difficulties in the inference because the resulting models exhibit structural instability. In a linear methodology as the VAR one, the problem is unsatisfactorily circumvent or by the use of dummy variables to try to catch the shifts (Favero-Spinelli, 1996) or by dividing the sample in sub-samples (Marcellino-Mizon, 1997). It hence seems that the use of non-linear models to describe adequately the dynamic behaviour of inflation rate in Italy and in the others OECD countries has great potential. Such models in fact have been proved successful in the identification of different regimes and in uncover the determinants of the transition between them.

This paper is an attempt to contribute to the international debate by carrying out a forecasting exercise in order to compare the forecasting performance of alternative models for the US inflation over the period 1950.1-2002.7. NAIRU Phillips curve models forecasts are contrasted with those obtained by a special class of nonlinear time series models, namely the threshold autoregressive models. Time series analysis has recently witnessed interesting developments in the modelling of nonlinearities, asymmetries and cyclical phenomena. The need to apply this class of models, as emphasized above, derives from the consideration that traditional econometric models and vector autoregressive models have rarely proved successful in describing and forecasting inflation dynamics; this is mainly due to the intrinsic characteristics of the inflation rate series, consistently documented across different countries and over different time periods, namely high persistence and presence of several structural breaks. Nonlinear models are therefore expected to provide noticeable gains in terms of higher forecast accuracy.

The forecasting performance of the estimated models is evaluated at horizons as far as two years ahead. Forecasts of inflation are intrinsically relevant not only for monetary policy purposes, but also to assess the usefulness and applicability of the new forecast evaluation tools, namely density forecasts, in relatively small samples as those available for most macroeconomic time series (see Diebold et al., 1998 and Wallis, 2003). The evaluation based on such tools will be accompanied by the one based on traditional Mean Square Forecast Error (MSFE) metric and the Diebold and Mariano (1995) test.

The rest of the paper is organised as follows. In section 2 we present the statistical properties of the data and the results of the linearity tests. In section 3 we report the results from the modelling procedures. The forecasting exercise is reported and discussed in section 4, while in section 5 we summarise the results and make some concluding remarks.

2. Data description and linearity tests

In this paper we analyse the dynamic behaviour of the US Consumer Price Index (CPI) for the period 1950.1-2002.7. More precisely the study focuses on the 12-month changes in the 12-month inflation rate. The 12-month inflation rate, p_t^{12}, is defined as the 12-month change of the price index, p_t (CPI)[1]:

$$p_t^{12} = \ln(p_t) - \ln(p_{t-12})$$

The 12-month changes in the 12-month inflation, $\Delta_{12}\pi_t^{12}$, is graphed in Figure 1. It is evident that the series is characterized by a changing degree of volatility, with the later two decades featuring very low volatility. This is coincident with the change in the monetary policy implemented in the mid-1980s and might be a possible explanation for the results reported in Atkenson-Ohanian (2001), who argue that modern inflation forecasting models are not able to provide forecasts more accurate than the ones obtained from a simple naïve model.

The statistical properties of the series are reported in table 1 for entire sample period, the estimation period and the forecasting period. The estimation sample refer to the period 1950.1-1985.12, while the forecasting sample cover the period 1986.1-2002.7.

The splitting of the entire sample between estimation and forecasting period allows us to withhold around 30% of the total number of observations in order to evaluate the forecasting performance of the nonlinear models, as suggested by Granger (1993).

The 12-month changes in the 12-month inflation series $\Delta_{12}\pi_t^{12}$ is mean stationary, kurtosis is particularly high over the entire period, while the forecasting period exhibits a lower variance with respect to the estimation one.

In order to detect nonlinearities in the $\Delta_{12}\pi_t^{12}$ series, we performed the RESET test and the S2 test proposed by Luukkonen-

[1] DATASTREAM is the source of all the variables included in this study.

Saikkonen-Teräsvirta (1988a, 1988b). Both tests are devised for the null hypothesis of linearity. While the RESET test is devised for a generic form of misspecification, the S2 test is formulated for a specific alternative hypothesis, i.e. smooth transition autoregressive (STAR)-type nonlinearity. Luukkonen-Saikkonen-Teräsvirta, however, show that the S2 test has reasonable power even when the true model is a SETAR one. The RESET test has been computed in the traditional version and in the modified version found to be superior by Thursby and Schmidt (1977)[2]. The S2 test is performed assuming that the variable governing the transition from one regime to the other is y_{t-d} with the delay parameter d in the range [1,6][3].

Table 2 reports the results of the linearity tests computed for the whole sample period, the estimation period and the forecast period. The selected lag order p ranges from 3 to 5 in order to check for the effects of different dynamic structures. The tests applied to the entire sample period and to the estimation period lead to the rejection of the null in a large number of cases, indicating that there is strong evidence of nonlinear components for the data. However, when the tests are applied to the forecast period the evidence based on the RESET tests indicates that nonlinearities are present with less intensity, while the S2 test (d=3,4), on the other hand, is highly significant at almost all lags.

---

[2] In the traditional form, the RESET test is computed by running a linear autoregression of order p, followed by an auxiliary regression in which powers of the fitted values obtained in the first stage are included along with the initial regressors. The modified RESET test requires that all the initial regressors enter linearly and up to a certain power h in the auxiliary regression; Thursby and Schimdt suggest using $h=4$. The Lagrange Multiplier form (Granger and Teräsvirta, 1993) of the test is adopted in this study, thus the test is distributed as a c^2 with up to $3p$ degrees of freedom for the modified version.

[3] The auxiliary regression for the LM S_2 test is computed as follows:

$\hat{e}_t = b_0 + \sum_{i=1}^{p} b_i y_{t-i} + \sum_{i=1}^{p} x_i y_{t-i} y_{t-d} + \sum_{i=1}^{p} y_i y_{t-i} y_{t-d}^2 + \sum_{i=1}^{p} k_i y_{t-i} y_{t-d}^3$ where ε_t are the estimated residuals from a linear regression of order p. Under the null hypothesis the test has a χ^2 distribution with $3p$ degrees of freedom.

3. Models specification

As mentioned in the introduction the forecasting exercise is based on the evaluation of the performance of linear models of the Phillips curve type and nonlinear threshold models. A naïve and a simple autoregressive model, AR(12), are also estimated as alternative benchmarks.

3.1 The Phillips curve linear models

In a recent paper Atkenson and Ohanian (2001) analysis whether NAIRU Phillips curve models could be appropriate to forecast US inflation in periods characterised by low and stable inflation, as it was the case in the US in the 1990s. They compare the NAIRU models which include the unemployment gap with a simple naïve model over the period 1984-1999 and found that the naïve models is always superior in terms of RMSE. AO replicate the estimation in SW (1999) in which the change in the inflation rate depends on lagged values and on some demand side indicators (unemployment rate and activity index). Overall they specify 132 different models which turn out to be outperformed by the naïve models in the forecasting exercise. This result calls for a more adequate specification of the NAIRU models which accounts also for supply-side determinants of the inflation and for possible nonlinear features present in the data. However, Fisher-Liu-Zhou (2002) point out that the results in Atkenson and Ohanian (2001) are forecasting-period dependent, as the naïve model prove to be adequate only in period of low volatile inflation, in other periods (for instance 1977-84) Phillips curve models provide forecasting gains and exhibit predictive power in anticipating the direction of future inflation changes.

The Phillips curve models used by Atkenson and Ohanian (2001) include only lagged inflation and the unemployment rate or the activity index as excess demand variables, ignoring the supply-side of the economic system. In this research the traditional Phillips curve model is respecified following the triangular model in Gordon (1997) which includes three different kinds of variables: inertia (lagged inflation), excess demand, and supply shocks. The general representation is as follows:

$$p^{12}_{t+12} - p^{12}_{t} = a(L)(p_{t} - p_{t-1}) + b(L)x_{t} + g(L)z_{t} + e_{t+12}$$

where x_t is a vector of excess demand proxies variables, z_t collects supply-side variables and *a(L)*, *b(L)* and *g(L)* are the usual polynominals in the lag operator and the error term ε_t is assumed n.i.d.

The aim is to check whether the triangular models turn out to be superior from a forecasting perspective. As already mentioned the excess demand is represented, in turn, by the unemployment rate, the change in industrial production, capacity utilisation rate and the Chicago activity index; the latter is a weighted average of 85 monthly indicators of real economic activity. The oil price and the import prices should capture the supply component. By using the US data we estimate 12 models, four belong to the class of traditional PC models, while 8 are triangular specifications. In general the dynamics was specified including twelve lags of the dependent variable and of the demand variable and 13 lags for the supply variable[4]. The "best" model was selected according to the Akaike and Schwarz criteria.

3.2 Threshold autoregressive models

The threshold autoregressive models were first proposed by Tong (1978), Tong and Lim (1980) and Tong (1983). The basic idea of the TAR models is that the behaviour of a process is described by a finite set of linear autoregressions. The appropriate AR model that generates the value of the time series at each point in time is determined by the relation of a conditioning variable to the threshold values. If the conditioning variable is the dependent variable itself after some delay *d* (y_{t-d}), the model is known as *self-exciting threshold autoregressive* (SETAR) model.

The SETAR model is piecewise-linear in the space of the threshold variable, rather than in time. An interesting feature of SETAR models is that the stationarity of y_t does not require the model to be stationary in each regime, on the contrary, the limit cycle

[4] Note that the models which include the Chicago Activity Index, due to data availability are estimated starting from 1967.3

behaviour that this class of models is able to describe arises from the alternation of explosive and contractionary regimes.

In this study we choose a two-regime (SETAR-2) and a three-regime (SETAR-3) SETAR models, which can be represented as follows:

$$\text{SETAR-2:} \quad y_t = \begin{cases} f_0^{(1)} + \sum_{i=1}^{p^{(1)}} f_i^{(1)} y_{t-i} + e_t^{(1)} & if\ y_{t-d} \le r \\ f_0^{(2)} + \sum_{i=1}^{p^{(2)}} f_i^{(2)} y_{t-i} + e_t^{(2)} & if\ y_{t-d} > r \end{cases}$$

$$\text{SETAR-3:} \quad y_t = \begin{cases} f_0^{(1)} + \sum_{i=1}^{p^{(1)}} f_i^{(1)} y_{t-i} + e_t^{(1)} & if\ y_{t-d} \le r_1 \\ f_0^{(2)} + \sum_{i=1}^{p^{(2)}} f_i^{(2)} y_{t-i} + e_t^{(2)} & if\ r_1 < y_{t-d} \le r_2 \\ f_0^{(3)} + \sum_{i=1}^{p^{(3)}} f_i^{(3)} y_{t-i} + e_t^{(3)} & if\ y_{t-d} > r_2 \end{cases}$$

where $\varepsilon_t^{(j)}$ is assumed IID$(0,\sigma^{2(j)})$, r_j represent the threshold values and d the delay parameter.

The models are estimated by following the three-stage procedure suggested by Tong (1983) for the case of a SETAR-2 (p_1, p_2; d) model. For given values of d and r, separate AR models are fitted to the appropriate subsets of data, the order of each model is chosen according to the usual AIC criteria. In the second stage r can vary over a set of possible values while d has to remain fixed, the re-estimation of the separate AR models allows the determination of the r parameter, as the one for which AIC(d) attains its minimum value. In stage three the search over d is carried out by repeating both stage 1 and stage 2 for $d=d_1, d_2, ..., d_p$. The selected value of d is, again, the value that minimises AIC(d).

As potential candidate for the threshold variable in describing the inflation dynamics we allow not only for the lagged variable but also for the demand and supply variables which are included in the Phillips curve specifications, such unemployment rate, the change in industrial production, capacity utilization rate and the Chicago activity index. In these cases we assume that the regime alternation is governed by exogenous variables and the models are labelled TAR-2 and TAR-3.

The models show clear evidence that the annual change in the inflation rate is strongly characterised by nonlinearities as the dynamic structure, the estimated coefficients and the error variance differ across regimes[5].

4. Forecasting comparison results

In this section we conduct two different forecasting exercises intended to evaluate the models on their ability to produce point forecasts and density forecasts.

4.1 Point forecasts evaluation

As reported in section 2, the forecasting sample covers the period 1986.1-2000.7; the models are specified and estimated over the first estimation period, 1950.1-1985.12, and the first set of 1 to 24 steps ahead forecast (h=1, 2,…24) computed. The models are then estimated recursively keeping the same specification but extending the sample with one observation each time. In this way 176 point forecasts are obtained for each forecast horizon. These forecasts can be considered *genuine* forecasts as in the specification stage we completely ignore the information embodied in the forecasting period. The computation of multi-step-ahead forecasts from nonlinear models involves the solution of complex analytical calculations and the use of numerical integration techniques, or alternatively, the use of simulation methods. In this study the forecasts are obtained by applying the Monte Carlo method with

[5] Given the large number, estimated models are not reported in order to save space, but are available from the authors upon request.

regime-specific error variances, so that each point forecast is obtained as the average over 500 replications.

In table 3A we report the MSFEs normalised with respect to the naïve model, while table 3B reports the comparison with respect to the linear AR(12) model. We select these two models as alternative candidates in order to compare our results with those reported in Atkenson-Ohanian (2001) and to check the conclusion reached by Staiger et al. (1997), i.e. that a forecaster who used only lags of inflation would have produced more accurate two-year ahead forecasts of inflation than those based on the unemployment rate.

The values are calculated as the ratio of the competing model (Phillips Curve linear-type or threshold model) MSFE with respect to the MSFE of the benchmark model. Therefore, a value less than 1 denotes a better forecast performance of the competing model. We have also applied the Diebold and Mariano (DM) test for equality of forecasting accuracy, and indicated with stars the cases for which the MSFEs are statistically significantly between the benchmark and the competing model.

In contrast to what we expected the triangular models Phillips curve models proved marginally superior only when the specification includes the activity index and the import prices. All the other specifications provide a forecasting performance which on the bases of the Diebold-Mariano test turned out to be indistinguishable from that of the traditional PC models. This result may be due to an overfitting problem, while the triangular models are more adequate to capture the in-sample behaviour of the inflation variable, this cannot be exploited to improve forecasting accuracy. In table 3A and 3B we focus on the traditional Phillips curve model where the excess demand is represented by the unemployment rate or, in turn, by one of the exogenous variables, namely the capacity utilisation rate, the change in industrial production, the Chicago activity index and the unemployment rate.

Focusing on the Phillips curve models, although there seem to be some forecasting gains for the specification which includes the Chicago Activity Index at all forecasting horizon considered (a reduction of almost 30% for h=24), the Diebold-Mariano test is not significant. Therefore, the evidence reported in table 3A highlights the fact that the forecasting accuracy of Phillips curve models is not significantly different from the one exhibited by the naïve model. This finding is in contrast with previous results presented in Atkenson-Ohanian (2001), who claimed that naïve model was never outperformed by the Phillips curve competitors. It is worth noting that in the study by Atkenson-Ohanian (2001) the difference in RMSEs was not assessed by means of the Diebold-Mariano (1995) test. When the forecasting comparison is carried out with respect to the simple AR(12) model (table 3B), it turns out that all the Phillips curve models yield noticeable gains for medium and long forecasting horizons, i.e for h=6, 9, 12 and 24. The significant reductions in the MSFE metrics range from 49% to 25%.

Turning to nonlinear threshold models some interesting results emerge. This class of models - both in the self-exciting specification, i.e. when the transition variable is dependent variable lagged d times, and in the specifications for which the transition from one regime to the other is governed by one of the exogenous variables – are able to outperform the naïve model only for short run forecasts (h=3, 6). For short term forecasting horizons the accuracy gains, as measured by the reduction in the MSFE, range from 82% to 25%. For medium and long term horizon the threshold models performance deteriorates quite dramatically. However, with respect to the AR(12) evidence of forecasting gains is found starting from the 6 steps-ahead horizon. The only model that is always outperformed by the linear model is the two- and three- regime TAR model which include the Chicago Activity Index as a transition variable. The specification which yields the highest forecasting gains is the one which include the unemployment rate for both the TAR-2 and TAR-3 model.

4.2 Density forecasts evaluation

A density forecast is an estimate of the complete probability distribution of the possible future values of a variable. Density forecasts are becoming increasingly used in real time forecasting in macroeconomics and finance as they provide a full description of the uncertainty which accompanies point forecasts. In particular, central banks are interested in the uncertainty surrounding inflation forecasts in order to devise optimal monetary policies.[6]

Moreover, in a recent paper Clements and Smith (1998) show that the relative performance of non-linear time series models depends on how forecast accuracy is assessed (see also Boero and Marrocu, 20029 for the evaluation of exchange rate forecasts). Traditional measures, such as root mean squared forecast errors, on which evaluation is often based, may mask the superiority of non-linear models with respect to a simple random walk model. Such superiority becomes evident when forecasting accuracy is evaluated in terms of density or interval forecasts.

In this section, we evaluate the one-step-ahead density forecasts of the models by applying the methods suggested by Diebold *et al.* (1998) and surveyed by Tay and Wallis (2000).

The evaluation of the density forecasts is based on the analysis of the probability integral transforms of the actual realisations of the variables with respect to the forecast densities of the models. These are defined as $z_t = F_t(y_t)$, where F(.) is the forecast cumulative distribution function and y_t is the observed outcome. Thus, z_t is the forecast probability of observing an outcome no greater than that actually realized. If the density forecasts correspond to the true density, then the sequence of probability integral transforms $\{z_t\}_{t=1}^{N}$ is i.i.d. uniform (0,1). To check whether the sequence of probability integral transforms departs from the i.i.d. uniform hypothesis, the distributional properties of the z_t series are examined by visual inspection of plots of the empirical distribution function of the z_t series, which are compared with those of a uniform (0,1). To supplement these graphical devices, the Kolmogorov-Smirnov test[7] can be used on the sample distribution function of the z_t series (see Diebold *et al*., 1999, and Tay and Wallis, 2000). The independence part of the i.i.d. uniform (0,1) hypothesis can be assessed by studying

[6] the Bank of England, for instance, publishes density forecasts of inflation in its quarterly Inflation Report since February 1986 (see Wallis, 2003 for a discussion on alternative methods to evaluate the Bank of England density forecasts).

[7] The maximum absolute difference between the empirical distribution function and the distribution function under the null hypothesis of uniformity.

the correlograms of the z_t series and of powers of this series (to establish the existence of dependence in higher moments) and applying formal tests of autocorrelation.

In our analysis below, we use the Kolmogorov-Smirnov test and the Ljung-Box test for autocorrelation on $(z_t - \overline{z})$, $(z_t - \overline{z})^2$, $(z_t - \overline{z})^3$, $(z_t - \overline{z})^4$. A well known limitation of this approach is that the effects of a failure of independence on the distribution of the tests for unconditional uniformity is unknown. Moreover, failure of the uniformity assumption will affect the tests for autocorrelation.

The one-step-ahead density forecasts of the annual change in the 12-month inflation rate are obtained under the assumption of Gaussian errors, with the appropriate regime-specific variances for the SETAR models. In figure 2 we report some selected plots of the empirical distribution function of the z_t series against the theoretical uniform distribution function. We omit the 45° line to avoid over-crowding the plots. The 95% confidence intervals along side the hypothetical 45° line are calculated using the critical values of the Kolmogorov Smirnov test, reported in Lilliefors (1967, Table 1, p. 400), in the presence of estimated parameters[8]. In table 4 we report the results of the Ljung-Box test for autocorrelation of the z_t series and its powers.

For the class of threshold models with an exogenous transition variable we select the ones which includes the unemployment rate as these models exhibit an adequate performance in terms of MSFE.

As we can see from and figure 2, only the threshold model specifications seem to produce density forecasts which are unconditionally correct as suggested by the Kolmogorov Smirnov test. However, as shown in table 2 the SETAR-2 model and the SETAR-3 fail the location and the skewness moments; the two regime TAR model which includes the unemployment rate do not satisfy the independence part of the joint hypothesis, with the Ljung-Box test showing significant dependencies in the first and higher

[8] The formula reported in Lilliefors (1967) for T>30, level of significance 0.05, is given by $0.886/\sqrt{T}$. The standard critical values of the Kolmogorov-Smirnov test are probably a conservative estimate of the 'correct' critical values when certain parameters of the distribution must be estimated from the sample

moments of the z_t series. The three regime TAR model with the same transition variable passes the Ljung-Box test only for the second moment.

It is interesting to note that both the AR (12), although marginally, and more significantly the Phillips curve models fail the unconditional uniformity test in the evaluation over the entire forecasting sample. The plots of the cdf of the z_t series versus the uniform (0,1) distribution, in figure 2, show that the empirical cdf of the Phillips curve models crosses the bounds in various regions of the distribution. Moreover, table 4 clearly shows that the density forecasts from the PC models violate the independence assumption, violations occur with respect to almost all the powers of z_t transforms.

By combining the information in table 4 and figure 2, overall the threshold models have shown better able to capture the distributional aspects of the annual change in the 12-month inflation rate. This result is in line with previous evidence (Clements and Smith, 2000 and Boero-Marrocu, 2002) that nonlinear time series models are capable to yield forecasting accuracy gains when the forecasting evaluation is conducted by means of density forecasts and not only in terms of MSFEs. Our results also show that nonlinear models are indeed adequate to capture the intrinsic features of the inflation rate series.

5. Conclusions

This paper is an attempt to contribute to the international debate on how and when is possible to forecast inflation by means of traditional Phillips curve models or by adopting new forecasting devices, such as NAIRU models on one hand and nonlinear time series specification on the other. The latter models are expected to prove adequate in forecasting comparison as they device to capture features that are reported to characterize inflation rates across different countries, such jump-fenomena, regime alternation and persistence.

The traditional approach to the analysis of inflation dynamics is based on the Phillips Curve and on the concept of non-accelerating-inflation-rate-of-unemployment (NAIRU). The fundamental relation is estimated by regressing the inflation change on a constant, its own lagged values, lagged unemployment rate, on a set of exogenous control variables and supply shock variables. Different versions of a regression of the type described above have been often used in the literature both to estimate the level of the NAIRU and to describe and forecast the dynamics of inflation (see Staiger et al., 1997, and reference therein).

From both a theoretical and an empirical perspective the Phillips Curve and the concept of NAIRU have been the subject of very well known critiques. Besides, also from an empirical perspective, both the stagflation of the '70s and the simultaneous occurrence in the second half of the '90s in US of low and falling inflation and low unemployment (Brayton et al., 1999) appear to be at odds with the properties of a standard Phillips Curve. The new models – the so called New-Keynesian Phillips Curve models (Galí-Gertler, 1999) - are based on the synthesis between typical elements of the real business cycle literature (microfoundations and intertemporal optimisation) and typical elements of the new-keynesian literature (monopolistic competition and nominal rigidities). The main and substantial difference between a standard Phillips Curve and the New-Keynesian Phillips Curve rests on the forward-looking nature of the latter. Current inflation depends on future expected inflation and not from the previous inflation level. The attempts to use such formulation for an empirical model of inflation dynamics, however, have encountered serious problems, substantially due to the strong persistence properties of the inflation time-series.

From an empirical perspective, Stock and Watson (1996) pointed out the traditional linear Phillips curve models are inadequate to capture inflation dynamics, due to the high degree of non-linearity and to the presence of "jumps" in correspondence with some crucial dates and historical episodes.

In this study we carried out a forecasting exercise in order to compare the forecasting performance of alternative models for the US inflation over the period 1950.1-2002.7. NAIRU Phillips curve models forecasts are contrasted with those obtained by a special class of nonlinear time series models, namely the threshold autoregressive models. The forecasting performance of the estimated models is evaluated at horizons as far as two years ahead by means of traditional metrics, such as the MSFE, and by assessing the whole density forecasts obtained by the competing models.

In terms of MSFE, the Phillips curve models yield noticeable gains for medium and long forecasting horizons when compared to the AR(12) model, while their forecasting accuracy turn out to be indistinguishable from that of the naïve model.

The threshold models outperformed the naïve models only for short term horizons and proved to yield superior forecasts with respect to the linear AR(12) for medium and long term horizons.

The evaluation of one-step-ahead density forecasts highlight the superiority of threshold models with respect to their linear counterparts, both the AR(12) model and the Phillips Curve specifications. This result is in line with previous evidence (Clements and Smith, 2000 and Boero-Marrocu, 2002) that nonlinear time series models are capable of yielding forecasting accuracy gains when the forecasting evaluation is conducted by means of density forecasts and not only in terms of MSFEs. Our results also show that nonlinear models are indeed adequate to capture the intrinsic features of the inflation rate series.

Overall the results reported are promising and the inflation forecasting analysis can be extended in a number of direction. First, it would be interesting to check the robustness of our results by replicating the forecasting exercise for other countries and Union of countries. As a matter of fact, we think that the issue of aggregation deserve attention, in particular in the context of the European Union. Finally, it seems relevant to check whether it is possible to combine forecasts from different forecasting devices in order to improve their accuracy.

References

Atkeson A. and L. Ohanian (2001), Are Phillips Curves Useful for Forecasting Inflation?, *Federal Reserve Bank of Minneapolis, Quarterly Review*, 2-11.

Ascari, G. (2000) Optimising Agents, Staggered Wages and Persistence in the Real Effects of Money Shocks, *The Economic Journal*, 110, 664-86.

Boero, G. and E. Marrocu (2002), "The performance of non-linear exchange rate models: a forecasting comparison", *Journal of Forecasting*, 21, 513-542.

Brayton F., Roberts J.M., e Williams J.C. (1999), What's Happened to the Phillips Curve?, Federal Reserve Board of Governors, Washington D.C..

Clements, M.P. e Smith, J.P. (1998), Non-linearities in exchange rates, Warwick Economics Research Papers, n. 504.

Clements, M. P. and J.P. Smith (2000), "Evaluating the Forecast densities of linear and non-linear models: applications to output growth and unemployment", *Journal of Forecasting*, 19, 255-276.

Diebold, F.X. e Nason, J.A. (1990), Nonparametric exchange rate prediction? *Journal of International Economics*, 28, 315-332.

Diebold, F.X. e Mariano, R.S (1995), Comparing predictive accuracy, *Journal of Business and Economic Statistics*, vol. 13, 3, 253-263.

Diebold, Francis X; Tay, Anthony S; Wallis, Kenneth F. (1998), Evaluating Density Forecasts of Inflation: The Survey of Professional Forecasters. *New York University, Leonard N. Stern School of Business, Department of Economics, Working Papers.*

DIEBOLD, F.X., T.A. GUNTHER and A.S. TAY (1998), "Evaluating density forecasts with applications to financial risk management", *International Economic Review*, 39, 4, 863-883.

Favero, C.A. e Spinelli, F. (1999), Deficits, money growth and inflation in Italy: 1875 – 1994, *Economic Notes*, 28, 43-71.

Fisher, J.D.M., C.T. Liu and R. Zhou (2002), When can we forecast inflation?, Federal Reserve Bank of Chicago Economic Perspectives, 1Q, 30-42.

Galí, J. e Gertler M. (1999), Inflation Dynamics: A Structural Econometric Analysis, *Journal of Monetary Economics*, p. 195-222.

Gallo, G.M e Otranto, E. (1997), Inflazione in Italia (1970-1996): non linearità, asimmetrie e cambiamenti di regime, Ricerche quantitative per la Politica Economica.

Gordon R.J. (1997), The Time Varying NAIRU and its Implications for Economic Policy, *Journal of Economic Perspectives*, Inverno 1997.

Granger, C.W.J. e Teräsvirta, T. (1993), *Modelling non-linear economic relationships*, Advanced Texts in Econometrics, Oxford University Press.

Luukkonen, R., Saikkonen, P. e Teräsvirta, T. (1988a), Testing linearity against smooth transition autoregressive models; Biometrika, 75, 3, 491-499. Luukkonen, R., Saikkonen, P. e Teräsvirta, T. (1988), Testing linearity in univariate time series models, *Scandinavian Journal of Statistics*, 15, 161-175.

Luukkonen, R. e Saikkonen, P. (1988b), Lagrange multiplier tests for testing non-linearities in time series models, *Scandinavian Journal of Statistics*, 15, 55-68.

Marcellino, M. e Mizon, G.E., (1997), Wages, prices, productivity, inflation and unemployment in Italy 1970-1994, mimeo, Istituto Universitario Europeo, Fiesole.

Staiger, D., Stock, J.H., e Watson, M.W. (1997), The NAIRU, Unemployment and Monetary Policy, *Journal of Economic Perspectives*, 11, 33-49.

Stock, J.H., e Watson, M.W. (1996), Evidence on Structural Instability in Macroeconomic Time Series Relations, *Journal of Business and Economic Statistics*, 14, 11-29.

Tay, A.S. and K.F. Wallis (2000), "Density Forecasting: a Survey", *Journal of Forecasting*, 19, 235-254.

Tong, H. e Lim, K.S. (1980), Thresholds autoregression, limit cycles and cyclical data*;* *Journal of the Royal Statistical Society* B, 42, 3, 245-292.

Tong, H. (1983), Threshold models in nonlinear time series analysis, New York, Springer-Verlag.
Tong, H. (1990), *Nonlinear time Series, a dynamical system approach*; Oxford Statistical Science Series, 6, Claredon Press Oxford.
Wallis, Kenneth F. (2003), Chi-Squared Tests of Interval and Density Forecasts, and the Bank of England's Fan Charts, *International Journal of Forecasting,* forthcoming.

TABLES AND FIGURES

TABLE 1 DESCRIPTIVE STATISTICS

	Entire sample *1950.1-2002.7* T=631	Estimation sample *1950.1-1985.12* T=432	Forecasting sample *1986.1-2002.7* T=199
Mean	-0.1171	-0.1205	-0.1103
Median	-0.0171	-0.0170	-0.0679
Maximum	5.7513	5.7513	3.2470
Minimum	-6.9933	-6.9933	-3.2204
Std. Dev.	1.9500	2.2237	1.2143
Skewness	-0.2988	-0.2934	-0.0462
Kurtosis	4.1373	3.5003	3.3458
Jarque-Bera	41.749	10.108	1.0626
Probability	0.0000	0.0000	0.0000

TABLE 2 LINEARITY TESTS - P-VALUES

	Entire sample *1950.1-2002.7* T=631			Estimation sample *1950.1-1985.12* T=432			Forecasting sample *1986.1-2002.7* T=199		
p	*3*	*4*	*5*	*3*	*4*	*5*	*3*	*4*	*5*
RESET, *h*=2	0.1479	0.2394	0.1034	0.1112	0.2266	0.1096	0.8586	**0.8343**	0.8867
RESET, *h*=3	0.2222	0.3376	**0.0766**	0.1007	0.2014	**0.0396**	0.8461	0.8621	0.8605
RESET, *h*=4	0.3876	0.5338	0.1395	0.1987	0.3529	**0.0842**	0.8961	0.8442	0.8915
Mod. RESET, *h*=2	**0.0019**	**0.0073**	**0.0004**	**0.0031**	**0.0147**	**0.0013**	0.3142	0.4459	0.6092
Mod. RESET, *h*=3	**0.0158**	**0.0538**	**0.0034**	**0.0115**	**0.0562**	**0.0034**	0.5728	0.6482	0.6248
Mod. RESET, *h*=4	**0.0568**	0.1780	**0.0050**	**0.0428**	**0.2024**	**0.0080**	0.1344	0.1883	0.2073
S_2, *d*=1	0.2206	**0.0580**	0.1883	0.1882	0.0824	0.1959	0.7078	0.3652	0.3233
S_2, *d*=2	**0.0954**	**0.0137**	**0.0425**	**0.0947**	**0.0292**	**0.0648**	0.1203	**0.0436**	0.1236
S_2, *d*=3	**0.0094**	**0.0079**	**0.0019**	**0.0055**	**0.0079**	**0.0025**	**0.0203**	**0.0186**	**0.0084**
S_2, *d*=4	**0.0005**	**0.0028**	**0.0015**	**0.0004**	**0.0041**	**0.0038**	**0.0199**	**0.0357**	**0.0265**
S_2, *d*=5	**0.0035**	**0.0055**	**0.0076**	**0.0036**	**0.0117**	**0.0176**	**0.0693**	0.1048	0.1804
S_2, *d*=6	**0.0002**	**0.0012**	**0.0025**	**0.0004**	0.0029	**0.0058**	**0.0992**	0.1741	0.2896

p denotes the lag order under the null hypothesis of linearity

TABLE 3A NORMALIZED MSFE WITH RESPECT TO THE NAÏVE MODEL

	Steps-ahead	*1*	*3*	*6*	*9*	*12*	*18*	*24*
Linear Models								
	Exogenous variables							
Phillips Curve	manufacturing capacity utilization rate	1.023	1.017	1.072	1.104	1.140	1.077	***0.915***
Phillips Curve	change in industrial production (Δ)	1.126	1.114	1.157	1.170*	1.196*	1.076	1.119
Phillips Curve	unemployment rate	1.073	1.063	1.106	1.142	1.184	1.132	1.202
Phillips Curve	Chicago Activity Index	***0.946***	***0.946***	***0.951***	***0.950***	***0.922***	***0.842***	***0.726***
Non linear models								
	Exog. transition variables							
SETAR-2 (d=6)	--	**0.127****	**0.556****	0.931	1.153**	1.432**	1.108	1.074
TAR-2 (d=6)	manufacturing capacity utilization rate	**0.121****	**0.559****	1.049	1.373*	1.656**	1.600*	1.452*
TAR-2 (d=1)	change in industrial production (Δ)	**0.132****	**0.606****	1.080	1.484**	1.805**	1.604**	2.012**
TAR-2 (d=6)	unemployment rate	**0.146****	**0.652****	1.078	1.205	1.223**	1.057	1.078*
TAR-2 (d=5)	Chicago Activity Index	**0.128****	**0.659****	1.418*	2.165**	2.838**	2.209**	1.910**
SETAR-3 (d=1)	--	**0.126****	**0.546****	***0.924***	1.138**	1.378**	1.085	1.042
TAR-3 (d=6)	manufacturing capacity utilization rate	**0.119****	**0.548****	1.032	1.360*	1.658**	1.524*	1.387*
TAR- 3 (d=4)	change in industrial production (Δ)	**0.127****	**0.562****	***0.994***	1.251	1.522*	1.359*	1.243
TAR-3 (d=5)	unemployment rate	**0.152****	**0.745****	1.167	1.094	1.069*	1.033	1.011
TAR-3 (d=2)	Chicago Activity Index	**0.136****	**0.624****	1.231	1.661*	2.201**	1.670	1.379

*, ** indicates significance of the Diebold-Mariano (1995) test at 10% and 5% respectively.

TABLE 3B NORMALIZED MSFE WITH RESPECT TO THE AR(12) MODEL

	Steps-ahead	*1*	*3*	*6*	*9*	*12*	*18*	*24*
Linear Models								
	Exogenous variables							
Phillips Curve	manufacturing capacity utilization rate	8.732**	1.873**	***0.950***	**0.707****	**0.633****	***0.952***	***0.871***
Phillips Curve	change in industrial production (Δ)	9.610**	2.052**	1.025	**0.750****	**0.665****	***0.951***	1.065
Phillips Curve	unemployment rate	9.161**	1.958**	***0.980***	**0.732****	**0.658****	1.001	1.145
Phillips Curve	Chicago Activity Index	8.080**	1.743**	***0.842***	**0.608****	**0.512****	**0.745***	**0.691***
Non linear models								
	Exog. transition variables							
SETAR-2 (d=6)	--	1.087**	1.024**	**0.825***	**0.739****	**0.796***	***0.980***	1.022
TAR-2 (d=6)	manufacturing capacity utilization rate	1.029	1.029	***0.930***	***0.880***	***0.920***	1.414	1.383
TAR-2 (d=1)	change in industrial production (Δ)	1.125**	1.117	***0.957***	***0.951***	1.003	1.418**	1.915**
TAR-2 (d=6)	unemployment rate	1.247**	1.201	***0.955***	**0.772****	**0.679****	***0.935***	1.026
TAR-2 (d=5)	Chicago Activity Index	1.089**	1.215**	1.256*	1.387**	1.577**	1.953**	1.818**
SETAR-3 (d=1)	--	1.080*	1.005	**0.818***	**0.729****	**0.766***	***0.959***	***0.992***
TAR-3 (d=6)	manufacturing capacity utilization rate	1.020	1.010	***0.914***	***0.871***	***0.921***	1.347	1.321
TAR- 3 (d=4)	change in industrial production (Δ)	1.083**	1.035	***0.881***	***0.802***	***0.846***	1.202	1.183
TAR-3 (d=5)	unemployment rate	1.301**	1.373*	1.034	**0.701***	**0.594***	***0.913***	***0.962***
TAR-3 (d=2)	Chicago Activity Index	1.165**	1.150	1.091	1.064	1.223	1.476	1.313
*, ** indicates significance of the Diebold-Mariano (1995) test at 10% and 5% respectively.								

TABLE 4 P-VALUES OF THE LJUNG-BOX Q-STATISTICS FOR SERIAL CORRELATION (FIRST TWENTY-FOUR AUTOCORRELATIONS)

		Moments			
		$(z-\overline{z})$	$(z-\overline{z})^2$	$(z-\overline{z})^3$	$(z-\overline{z})^4$
	Exogenous variables				
AR(12)	--	0.000	0.009	0.000	0.006
Phillips Curve	manufacturing capacity utilization rate	0.000	0.000	0.000	0.000
Phillips Curve	change in industrial production (Δ)	0.000	0.000	0.000	0.000
Phillips Curve	unemployment rate	0.000	0.000	0.000	0.000
Phillips Curve	Chicago Activity Index	0.000	0.000	0.000	0.000
SETAR-2 (d=6)	--	0.000	0.126	0.000	0.114
TAR-2 (d=6)	unemployment rate	0.000	0.002	0.000	0.000
SETAR-3 (d=6)	--	0.000	0.215	0.000	0.208
TAR-3 (d=5)	unemployment rate	0.000	0.208	0.000	0.000

FIGURE 1

ANNUAL CHANGE OF THE US INFLATION RATE

1950.1- 2002.7

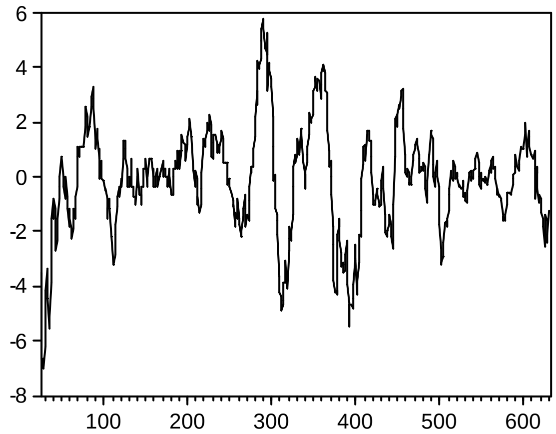
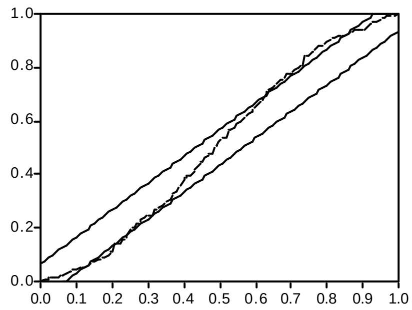
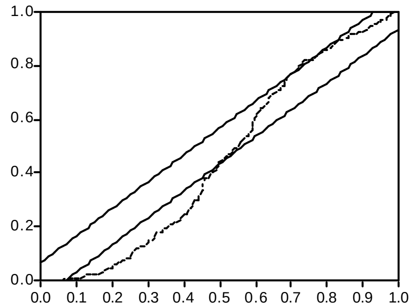


FIGURE 2

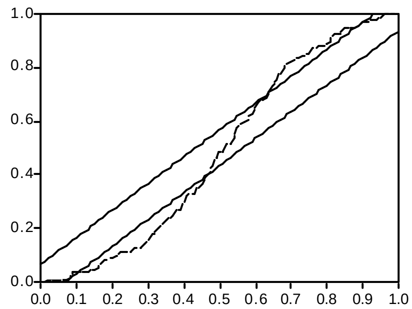
DENSITY FORECASTS



AR(12) model



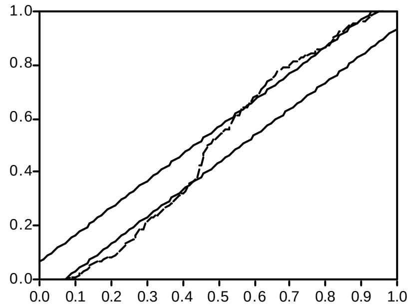
Phillips curve - capacity utilization rate



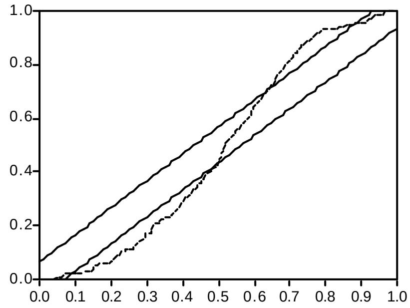
Phillips curve - change in industrial production

FIGURE 2

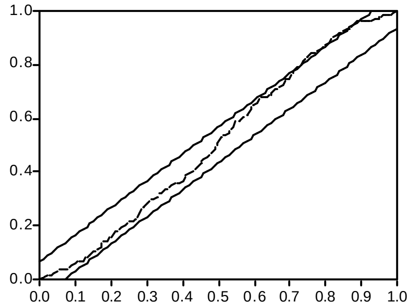
DENSITY FORECASTS *(CONT.ED)*



Phillips curve - Chicago activity index



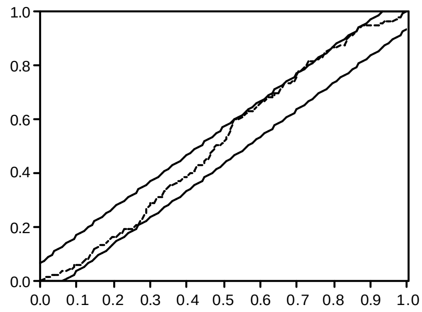
Phillips curve - unemployment rate



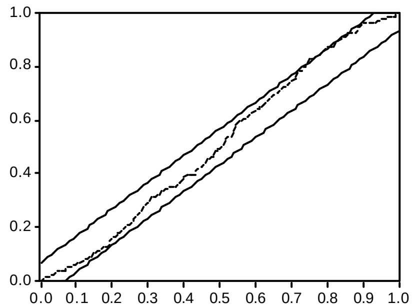
SETAR-2 model

FIGURE 2

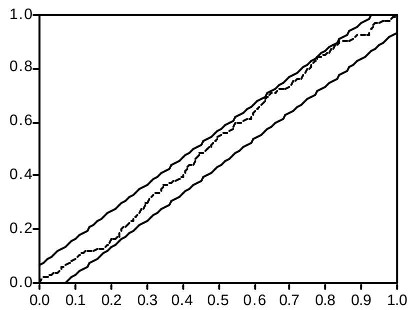
DENSITY FORECASTS *(CONT.ED)*



TAR-2 unemployment rate



SETAR-3 model



TAR-3 - unemployment rate