ELSEVIER

Predictable non-linearities in U.S. inflation

Jane M. Binner [a],[*], C. Thomas Elger [b], Birger Nilsson [b], Jonathan A. Tepper [c]

[a] *Economics and Strategy Group, Aston Business School, Aston University,*
Aston Triangle, Birmingham B4 7ET, UK
[b] *Lund University, Sweden*
[c] *Nottingham Trent University, UK*

Abstract

This paper compares the out-of-sample inflation forecasting performance of two non-linear models; a neural network and a Markov switching autoregressive (MS-AR) model. We find that predictable non-linearities in inflation are best accounted for by the MS-AR model.
© 2006 Elsevier B.V. All rights reserved.

Keywords: Inflation forecasting; Markov switching models; Recurrent neural networks

JEL classification: C45; C52; C53; E37

1. Introduction

There is a large literature comparing the forecasting performance of neural network (NN) models and linear autoregressive (AR) models; see Zhang et al. (1998). Recently, Stock and Watson (1999) addressed the question of forecasting performance of linear and non-linear models for a large number of macroeconomic time-series, including inflation. In their spirit, Nakamura (2005) evaluates the NN model forecasting U.S. inflation. She finds that the NN model outperforms univariate benchmark AR models for short horizons of one and two quarters.

However, with the exception of Binner et al. (2004) this literature does not compare inflation forecasts obtained from Markov switching autoregressive models with forecasts from NN models. This note aims at

---

filling this gap. We provide out-of-sample evidence, using the data in Nakamura (2005: henceforth EN), that predictable non-linearities in U.S. inflation are equally well captured by the NN and MS-AR models for short horizons. In contrast, for longer horizons of three and four quarters, the MS-AR model is the best forecaster. It is a non-linear generalization of the standard AR model with a simple closed form multi-period forecast equation and, therefore, a viable alternative to the NN model.

2. Forecasting models

Our multi-period inflation forecasts are *dynamic* forecasts, implying an iterative forecasting procedure in which a multi-period forecast is built up by a sequence of one-period forecasts. The general structure of the models and a description of how forecasts are obtained are provided below.

2.1. Benchmark models

Assume that k is the number of lags in a particular univariate *benchmark* AR model with coefficients $a_0,a_1,\ldots,a_k$ (estimated using ordinary least squares). By iterating forward, the dynamic forecast of inflation h periods ahead, π_{t+h}, is given by:

$$E_t[\pi_{t+h}] = a_0 \sum_{r=1}^{h} \left(\sum_{\ell=1}^{k} a_\ell\right)^{r-1} + \left(\sum_{\ell=1}^{k} a_\ell\right)^{h-1} \cdot \sum_{\ell=1}^{k} a_\ell \pi_{t+1-\ell}, \tag{1}$$

where the notation $E_t[.]$ indicates that the expectation is taken conditional on observed inflation up to and including time t.

2.2. Markov switching models

In the MS-AR models, an unobserved discrete state Markov chain governs the endogenous switches between different AR processes over time. Following Hamilton (1994, p. 692), the MS-AR models are estimated by maximum likelihood using the Hamilton-filter. To escape local optimums, the log-likelihood functions are maximized with simulated annealing (Goffe et al., 1994). The general MS-AR forecast equation is:

$$E_t[\pi_{t+h}] = \sum_{i=1}^{s} \sum_{j=1}^{s} \sum_{p=1}^{s^{h-1}} \Pr_t(S_t = i)\Pr_{pt}(S_{t+h} = j|S_t = i)E_t[\pi_{pt+h}|S_t = i], \tag{2}$$

where s is the number of states, $\Pr_t(S_t=i)$ is the probability of state i prevailing at time t and $\Pr_{pt}(S_{t+h}=j|S_t=i)$ is the probability of moving from state i at time t to state j at time $t+h$ along path p in the probability tree generated by the underlying Markov chain.[1] The former probabilities are obtained directly from the iterations through the Hamilton-filter while the latter are calculated using the estimated transition probability matrix. The multiplication by $\Pr_t(S_t=i)$ reflects the uncertainty of the prevailing state today, or, the fact that the Markov chain is unobserved even *ex post*. $E_t\lfloor\pi_{pt+h}|S_t=i\rfloor$ is the

---

[1] The index $p=1,\ldots,s^{h-1}$ represents any sequential ordering of the s^{h-1} different paths that begin in state i and end in state j.

forecasted inflation at time $t+h$ along path p and these expectations are evaluated as in Eq. (1), i.e. by iterating forward along the relevant path. Therefore, Eq. (2) states that dynamic inflation forecasts from the MS-AR models are calculated as probability weighted averages of future inflation by traversing the non-recombining tree generated by the Markov chain along all possible paths.

2.3. Neural network model

We employ a recurrent neural network (RNN) model that distinguishes itself from standard feedforward network models in that activations are allowed to feed back to units within the same or preceding layer(s). This forms an internal memory system that enables an RNN to construct temporally sensitive internal representations in response to temporal features found within a data set (Palmer-Brown et al., 2002).

We use a combination of the Jordan (1986) network and Elman's (1990) simple recurrent network to obtain forecasts:

$$E_t[\pi_{t+1}] = g(f(\boldsymbol{c}_{t-1}, \boldsymbol{x}_t, \boldsymbol{\theta}_f), \boldsymbol{\theta}_g), \tag{3}$$

where $\boldsymbol{c}_{t-1}$ (the context vector) is the concatenation of the previous hidden state vector and the previous external output vector, $\boldsymbol{x}_t$ is the external vector of input variables (current and lagged inflation), $\boldsymbol{\theta}_f$ is the set of weights connecting the input layer to the hidden layer, $\boldsymbol{\theta}_g$ is the set of weights connecting the hidden layer to the output layer and, finally, functions f and g represent the activation vectors from the hidden and output layers respectively (we apply the hyperbolic tangent function to the inner products performed for f and g).

Our RNN model employs backpropagation-through-time, an efficient gradient-descent learning algorithm for recurrent networks. To identify the onset of over-fitting during training we use a standard variant of cross-validation referred to as the leave-one-out method (Haykin, 1999, p. 218) as a stopping criterion suitable for estimation problems with sparse data.

Dynamic $t+h$ forecasts from the RNN are generated, consistent with the AR and MS-AR-models, using an iterative procedure. Thus, the RNN is functionally equivalent to a non-linear regression model used for time-series forecasting (Zhang et al., 1998).

3. Data and model specification

Our quarterly inflation data, calculated from the U.S. GDP deflator, covers the period 1960:1 to 2003:1. The last 100 observations are used for forecast evaluation only. Dynamic out-of-sample inflation forecasts are constructed for forecast horizons from 1 to 4 quarters.

Both AR and MS-AR models are estimated using lag lengths from 1 to 8. For the MS-AR model we assume an underlying two-state time-homogenous discrete Markov chain. Following Binner et al. (2004), the intercept and the volatility are allowed to switch between the two states while all autoregressive parameters are equal across states.[2] An empirical advantage of allowing the variance to change across regimes is that a higher mean could be associated with higher (idiosyncratic) variance. A regime-dependent variance can, therefore, help in forecasting inflation.

---

[2] Binner et al. (2004) find that allowing for a different autoregressive structure in different regimes leads to a deterioration of out-of-sample forecasting performance. They interpret this deterioration as in-sample overfitting.

Table 1
Out-of-sample MSE ratios

	Forecast horizon (quarters)			
	1	2	3	4
MS models				
AR1	<u>86.17</u>	<u>76.12</u>	<u>68.96</u>	<u>64.66</u>
AR2	<u>88.61</u>	<u>81.11</u>	<u>75.11</u>	<u>71.54</u>
AR3	<u>93.48</u>	<u>87.81</u>	<u>83.00</u>	<u>81.32</u>
AR4	97.01	93.30	<u>88.09</u>	<u>86.94</u>
AR5	102.65	98.50	92.86	90.06
AR6	95.39	<u>89.44</u>	<u>85.16</u>	<u>82.24</u>
AR7	<u>90.16</u>	<u>88.97</u>	<u>84.47</u>	<u>82.30</u>
AR8	<u>91.88</u>	<u>89.09</u>	<u>85.22</u>	<u>82.75</u>
RNN model				
AR1	<u>87.19</u>	<u>76.44</u>	<u>80.95</u>	<u>89.45</u>
AR2	<u>93.42</u>	<u>86.10</u>	<u>93.67</u>	102.81
AR3	94.73	93.81	106.01	118.45
AR4	95.53	90.93	104.58	121.14
AR5	99.69	94.56	108.79	124.00
AR6	94.04	95.37	111.20	127.46
AR7	<u>84.03</u>	91.29	104.19	119.84
AR8	<u>85.14</u>	<u>88.99</u>	102.18	116.05

In the first panel, MSEs from the MS-AR models are compared to MSEs from AR models with matching lag structures ($100 \times \text{MSE}_{\text{MS-AR}}/\text{MSE}_{\text{AR}}$). In the second panel, the MSEs from an RNN with two lags are compared to MSEs from AR models with different lag structures. Underlined numbers indicate a significant one-sided Diebold–Mariano test statistic based on mean squared errors at the 5% level.

The numbers of parameters in the AR and MS-AR models depend on the particular lag-length. There are always 4 more parameters in the MS-AR model; one additional mean parameter, one additional variance parameter, and two transition probabilities.

Our RNN model is based on two lags of inflation.[3] We normalize the input variables prior to training to accelerate the learning process. Further, our RNN consists of 10 hidden units (a number determined by preliminary forecasting experiments), one external input node and one external output node.

4. Results

Mean squared error (MSE) ratios are presented in Table 1. They suggest that both the MS-AR models and the RNN model outperform the benchmark AR models on short forecast horizons of 1 and 2 quarters.

More specifically, the MSE-ratios suggest that the MS-AR outperforms the linear specification in 15 out of 16 cases. The RNN, on the other hand, yields lower MSE's than the AR models in all 16 cases. In addition, one-sided tests based on Diebold and Mariano (1995) typically suggest that the differences are statistically significant whenever the MSE-ratios are low (as indicated by underlined MSE-ratios in Table 1). More

---

[3] EN estimates NN models with up to four lags, but only reports findings from an NN with two lags. She finds that this modification has essentially no impact on forecast performance. The inclusion of additional lags in our NN model confirms this finding.

specifically, forecasts from the MS-AR are significantly better in 11 out of 16 cases. For the RNN, they are significantly better in 7 cases.

For longer forecast horizons of 3 and 4 quarters, MSE-ratios indicate that MS-AR models outperform both the RNN model and the benchmark models. The forecasting performance of the RNN model declines rapidly with increasing horizon and the linear benchmark models with more than two lags are superior to the RNN model for longer horizons. A similar result is reported by EN. An explanation is that with gradient descent-based learning the error gradients vanish early and therefore longer temporal dependencies in the time-series data are difficult to learn (Bengio et al., 1994). The Diebold–Mariano test statistics show that the MS-AR models are significantly better than the AR models in 14 out of 16 cases. The RNN is, on the other hand, significantly better than the AR model in only 3 cases.

For completeness, we also investigate the cases when MSE-ratios suggest that the linear models outperform the nonlinear models using Diebold–Mariano tests. These tests reveal that the MSE for the AR model is significantly lower than the MSE for the RNN in 5 cases out of 32.[4] Finally, the AR model never significantly outperforms the MS-AR model.

The fact that the MS-AR(1) and the RNN tend to perform better (in relative terms) when compared to the AR(1) can be interpreted in at least 2 ways. Firstly, the additional parameters in the MS-AR(1) and the RNN are more valuable, i.e. provide more out-of-sample information, when the AR model is simple compared to the situation when the AR model is more complex (contains more lags). Secondly, when a large number of lags are included, the additional parameters create too much in-sample flexibility which hurts the (relative) out-of-sample forecasting performance.

To summarize, our empirical findings suggest that both the RNN model and the MS-AR models are well-suited for capturing predictable non-linearities in U.S. inflation. However, for longer forecast horizons, the MS-AR model stands out as the preferred model.

Acknowledgements

References

Bengio, Y., Simard, P., Frasconi, P., 1994. Learning long-term dependencies with gradient descent is difficult. IEEE Transactions on Neural Networks 5, 1550–1560.

Binner, J.M., Elger, T., Nilsson, B., Tepper, J.A., 2004. Tools for non-linear time series forecasting in economics: an empirical comparison of regime switching vector autoregressive models and recurrent neural networks. Advances in Econometrics 19, 71–92.

Elman, J., 1990. Finding structure in time. Cognitive Science 14, 179–211.

Goffe, W., Ferrier, G., Rogers, J., 1994. Global optimization of statistical functions with simulated annealing. Journal of Econometrics 60, 65–99.

Hamilton, J., 1994. Time Series Analysis. Princeton University Press.

Haykin, S., 1999. Neural Networks: A Comprehensive Foundation. Prentice Hall.

---

[4] These cases are the AR model with 5 lags on the 3 quarters forecast horizon and the AR models with 3, 5, 6, and 7 lags on the 4 quarters horizon.

Jordan, M., 1986. Attractor Dynamics and Parallelism in a Connectionist Sequential Machine. Proceedings of the 8th Annual Conference of the Cognitive Science Society, pp. 531–545.

Nakamura, E., 2005. Inflation forecasting using a neural network. Economics Letters 86, 373–378.

Palmer-Brown, D., Tepper, J.A., Powell, H.M., 2002. Connectionist natural language parsing. Trends in Cognitive Sciences 6, 437–442.

Stock, J.H., Watson, M.W., 1999. A comparison of linear and non-linear univariate models for forecasting macroeconomic time series. In: Engle, R., White, R. (Eds.), Cointegration, Causality and Forecasting: A Festschrift in Honor of Clive W.J. Granger. Oxford University Press, pp. 1–44.

Zhang, G., Patuwo, B.E., Hu, M.Y., 1998. Forecasting with artificial neural networks: the state of the art. International Journal of Forecasting 14, 35–62.