Does Money Matter?
An Artificial Intelligence Approach

JM Binner [1] B Jones [2] G Kendall [3] J Tepper [4] PTino [5]

[1] Aston University, Birmingham, UK
[2] State University of New York, Binghamton, USA
[3] University of Nottingham, Nottingham, UK
[4] The Nottingham Trent University, Nottingham, UK
[5] The University of Birmingham, Birmingham, UK

Abstract

This paper provides the most complete evidence to date on the importance of monetary aggregates as a policy tool in an inflation forecasting experiment. Every possible definition of ‘money’ in the USA is being considered for the full data period (1960 – 2006), in addition to two different approaches to constructing the benchmark asset, using the most sophisticated non-linear artificial intelligence techniques available, namely, recurrent neural networks, evolutionary strategies and kernel methods. Three top computer scientists in three top UK universities (Dr Peter Tino at the University of Birmingham, Dr Graham Kendall at the University of Nottingham and Dr Jonathan Tepper at Nottingham Trent University) are competing to find the best fitting US inflation forecasting models using their own specialist artificial intelligence techniques. Results will be evaluated using standard forecasting evaluation criteria and compared to forecasts from traditional econometric models produced by Dr Binner. This paper therefore addresses not only the most controversial questions in monetary economics - exactly how to construct monetary aggregates and to what level of aggregation, but also addresses the ever increasing role of artificial intelligence techniques in economics and how these methods can improve upon traditional econometric modelling techniques. Lessons learned from the experiment will have direct relevance for monetary policymakers around the world and econometricians/forecasters alike. Given the multidisciplinary nature of this work, the results will also add value to the existing knowledge of computer scientists in particular and more generally speaking, any scientist using artificial intelligence techniques.

Keywords: Divisia, Inflation, Evolutionary Strategies, Recurrent Neural Networks, Kernel Methods.

1. Introduction

If macroeconomists ever agree on anything, it is that a relationship exists between the rate of growth of the money supply and inflation. According to the Quantity Theory tradition of economics, the money stock will determine the general level of prices (at least in the long term) and according to the monetarists it will influence real activity in the short run. This relationship has traditionally played an important role in macroeconomic policy as governments try to control inflation.

Measuring money supply is no easy task, however. Component assets range from “narrow money,” which includes cash, non interest bearing demand deposits and sight deposits on which cheques can be drawn, to “broader” money, which includes non-checkable liquid balances and other liquid financial assets such as Certificates of Deposit. In practice, official measures have evolved over time according to policy needs. Obviously many of these assets yield an interest rate and could thus be chosen as a form of savings as well as being available for transactions. Financial innovation, in particular liberalisation and competition in banking, has led to shifts in demand between the components of “money” which have undermined earlier empirical regularities and made it more difficult to distinguish money which is held for transactions purposes from money which is held for savings purposes [1].

The objective of current monetary policy is to identify indicators of macroeconomic conditions that will alert policy makers to impending inflationary pressures sufficiently early to allow the necessary action to be taken to control and remedy the problem. Given that traditional monetary aggregates are constructed by simple summation, their use as a macroeconomic policy tool is highly questionable. More specifically, this form of aggregation weights equally and linearly assets as different as cash and bonds. This form of aggregation involves the explicit assumption that the excluded financial assets provide no monetary services and the included balances provide equal levels of monetary services, whether they are cash or ‘checkable’ deposits or liquid savings account balances. It is clear that all components of the monetary aggregates do not contribute equally to the economy’s monetary services flow. Obviously, one hundred dollars currency provides greater transactions services than the equivalent value in bonds. Thus, this form of aggregation is therefore producing a theoretically unsatisfactory definition of the economy’s monetary flow. From a micro-demand perspective it is hard to justify adding together component assets having differing yields that vary over time, especially since the accepted view allows only perfect substitutes to be combined as one “commodity”.

In the last two decades many countries have relegated the role of monetary aggregates from macroeconomic policy control tool to indicator variables. Barnett [2,3] attributes to a great extent the downgrading of monetary aggregates to the use of Simple Sum aggregates as they have been unable to cope with financial innovation. By drawing on statistical index number theory and consumer demand theory, Barnett advocates the use of chain-linked index numbers as a means of constructing a weighted Divisia index number measure of money. The potential advantage of the Divisia monetary aggregate is that the weights can vary over time in response to financial innovation. [3] provides a survey of the relevant literature, whilst [4] reviews the construction of Divisia indices and associated problems.

This paper is the first of its kind to investigate the forecasting performance of every definition of the US money supply in an inflation forecasting experiment. The novelty of this paper lies in the use of the most sophisticated artificial intelligence techniques available to examine the USA’s recent experience of inflation. Our previous experience in inflation forecasting using state of the art approaches give us confidence to believe that significant advances in macroeconomic forecasting and policymaking are possible using techniques such as those employed in this paper. As in our earlier work, [5], results achieved using artificial intelligence techniques are compared with those using traditional econometric methods.

2. Data and Forecasting Model

Monthly seasonally adjusted CPI data spanning the period 1960:01 to 2006:01 were used in this analysis. Inflation was constructed from CPI for each month as year on year growth rates of prices. Sixteen alternative definitions of the US money supply were evaluated in terms of their inflation forecasting potential. We employed the standard Federal Reserve Bank of St Louis simple sum and Divisia MSI formulations at four levels of aggregation, namely, M1, M2, MZM and M3. These were downloaded directly from the FRED database at St Louis. Next, experimental Divisia indices were constructed to determine whether or not the St Louis Fed constructions could be improved upon. Thus our Divisia measures were constructed using alternative benchmark rates, namely, the BAA interest rate as a benchmark and with an upper envelope of assets used as a benchmark. Out of the sixteen available measures of money, we restricted the choice of monetary aggregate to just one for each of our model selections. We also experimented with the inflation forecasting potential of interest rates in our study, given that short run interest rates are commonly used as a tool to control inflation. Thus, a short interest rate, (three month treasury bills), a long interest rate, (the BAA rate), both a short and long rate were added, or finally no interest rates at all were added alongside the chosen measure of money defined above. Lags of each variable and orders of differencing of each variable were permitted and left to the discretion of the individual modeller.

Each of the three computer scientists were asked to find the best inflation forecasting model based on their own preferred artificial intelligence models. Of the 541 data points available, the first 433 were used for training, (January 1960 - February 1997), the next 50 data points were used for validation, (March 1997 – April 2001) and the next 50 data points were used for forecasting (May 2001 – June 2005).

Individual models compete against one another with the top four being selected based on their performance on the validation set. The winning network models are subsequently evaluated individually and as an ensemble to ascertain performances across horizons of both six months and a year.

3. Models

3.1 Evolutionary Neural Networks

The first approach we will investigate is the use of evolutionary neural networks. The other methodologies reported in this paper typically require the formation of a set of systematic experiments to iterate over the various control parameters. By utilising evolutionary neural networks we aim to find a good set of parameters using evolutionary pressure alone. Some of our previous work has shown success with this approach (see, for example, [6,7]) but the dataset under consideration here is a lot larger and we have never compared the results against as many methodologies as we are proposing here.

Using the evolutionary paradigm we create a population of neural networks. Each network uses inputs from one of the 16 available measures of money along with zero, one or two of the rates of interest. Each of these measures is *lagged*, with the amount of lag being subject to evolutionary pressure. We also allow the network to be recurrent, so that the values which are output from the hidden layer are fed back into the input layer.

The networks compete with one another for survival, with half of the best performing networks surviving to the next generation. The worst performing networks are killed off, and are replaced by mutated versions of the best networks. Mutations are performed by changing either the weights within the network or its architecture. For example, one mutation would simply adjust the weights by normally distributed random variables. Another will add/delete the number of hidden neurons. Yet another will adjust the amount of time lag for the inputs being used.

Our previous work [6,7], as well as the work other researchers (e.g. [8]), has shown that this approach can evolve highly competitive neural networks and our preliminary results (not reported here) show promise for this dataset.

3.2 Recurrent Neural Networks

Recurrent neural networks (RNNs) are typically adaptations of the traditional feed-forward multi-layered perceptron (FF-MLP) trained with the back-propagation algorithm[1] [9]. This particular class of RNN extends the FF-MLP architecture to include recurrent connections that allow network activations to feedback (as internal input at subsequent time steps) to units within the same or preceding layer(s). Such internal memories enable the RNN to construct internal representations of temporal order and dependencies which may exist in the data. Also, assuming non-linear activation functions are used, the universal function approximation properties of FF-MLPs naturally extends to RNNs. These properties have led to wide appeal of RNNs for modelling linear time-series data.

We build upon our previous successes with RNNs [10,11] for modelling financial time-series data and assess the efficacy of applying a discrete-time recurrent neural network to the problem of forecasting US inflation rates. The architecture used employs recurrent connections from the output layer back to the input layer, as found in *Jordan networks* [12], and also from the hidden layer back to the input layer, as found in Elman's Simple Recurrent Networks (SRNs) [13]. An RNN with this type of recurrency can represent auto-regressive with moving average (ARMA) estimators [14]. Research using such networks for time-series prediction tasks has so far proven encouraging [10,11,15,16,17] and warrants further investigation. It is accepted, however, that careful selection of appropriate input variables, lag structures, and network model is crucial to achieving an RNN capable of adequately dealing with the high levels of noise, nonstationarity and nonlinearity which may be prevalent within the time-series data.

The internal input or *context* layer consists of activations fed back via the feedback connections. *Backpropagation-through-time* (BPTT) [9,18,19] is then used to train the individual networks. The specific lag structure, model structure and size are determined through empirical evaluation.

3.3 Kernel-based Regression Techniques

In addition to the above models, we applied Kernel Recursive Least Squares (KRLS) [20]. This is basically a kernelized version of the classical Recursive least-squares technique. Recursive least-squares are performed in the feature space defined by the kernel. We used spherical Gaussian kernels; kernel width is set on the validation set. Regularization parameter controlling the feature space collinearity threshold is set on the validation set as well.

We also considered Kernel Partial Least Squares (KPLS) [21]. This is a kernelized version of the Partial Least Squares technique. Again, spherical Gaussian kernels were used. Kernel width as well as the number of latent factors were set on the validation set.

3.4 Naïve Models

Naïve 1: If the prediction horizon is T months (in our case T=6 or T=12), predict that in T months we will observe the current inflation rate. In other words, if the current inflation rate is R(t), we predict that the inflation rate at time t+T will be R(t), i.e. r(t+T) = R(t), where R(t) is the actual observed inflation rate at time t and r(t) is the inflation rate predicted to occur at time t by our model. This model corresponds to the random walk hypothesis with moves governed by a symmetrical zero-mean density function, and thus measures "the degree to which the efficient market hypothesis applies".

Naïve 2: calculate the mean, Rm, of the inflation rates on the training set and then always predict Rm as the future inflation rate. This should work well if the series is stationary and variance is low.

1. From hereon simply referred to as FF-MLP.

4. Results

Results from recurrent neural networks and evolutionary strategies will follow in the final paper. For the Kernel based regression methods over the 12 month forecast horizon, the performances of both KRLS and KPLS were comparable; hence just results for KRLS are reported. The best performing KRLS on validation set was a non-linear autoregressive model on inflation rates alone. Input data was normalized to zero mean and unit standard deviation (the mean and variance are determined on the training set alone). The input lag was set at 24 months and kernel width was 1.8. The mean square error of the best performing KRLS was 0.000135, compared with 0.000159 for the naïve 1, selected as the representative of model class "Naïve" of simple, but potentially powerful models. This represents an Improvement Over Naïve (ION) model of 15.1%. A flat ensemble of the four best performing models on the validation set produced a mean square error ensemble of 0.000109, with an ION of 31.4%.

5. Evaluation

The "market" is efficient in the sense that it is difficult to beat the Naïve 1 model. The ION measure is best suited for our purposes, because rather than worrying about the precise MSE values, we should be concerned about how much we can beat the rather obvious strategy Naïve 1 by, based on the Efficient Market Hypothesis. We find that the longer the prediction horizon, the harder it is to accurately predict the inflation rate. It seems that inclusion of historical inflation rates alone works quite well. We may speculate that the value of money is implicitly represented in the inflation rates and thus their explicit inclusion as input variables does not help, rather, it can actually make things worse because of the under sampling problems.

Because the validation set is just a sample, we cannot rely solely on picking the best model on the validation set. Slightly worse models on the validation set may be slightly better on the test set. Therefore we constructed flat ensembles of several best performing models on the validation set. The performance of the ensemble on the test set is always better than that of the single winner picked on the validation set.

It is interesting that the ION measures do not change drastically with increasing prediction horizon. That indicates that for longer prediction horizons, it is more difficult to predict the inflation rates, but the naive model is less suited as well, so things cancel out. Future results on the relative performance of the respective Divisia indices will follow in the final paper.

References

[1] A.W. Mullineux, (1996). Financial Innovation, Banking and Monetary Aggregates, Chapter 1, pp. 1–12. Cheltenham, UK: (Eds) Edward Elgar.

[2] W.A. Barnett, (1980) *Journal of Econometrics* 14(1), 11–48.

[3] W.A. Barnett, D. Fisher, and A. Serletis (1992), *Journal of Economic Literature*, 30, 2086-119.

[4] L. Drake, A.W. Mullineux, and J. Agung (1997), *Applied Economics* 29: 775-786.

[5] J.M. Binner, G. Kendall, and A.M. Gazely, (2005), 4th Int. Conf. on Computational Intelligence in Economics and Finance, 871-874.

[6] J.M. Binner, Kendall G. and Gazely A.M. (2004). *Advances in Econometrics*, 19:127-143.

[7] J.M. Binner and G. Kendall, (2002), Proc. Int. Conf. on Artificial Intelligence (IC-AI'02), CSREA Press. Arabnia H.R and Youngsong M. (eds), pp 884-889.

[8] D. Fogel (2002) Blondie24: Playing at the Edge of AI. Morgan Kaufmann.

[9] D. Rumelhart, G. Hinton, R. Williams (1986). Learning Internal Representations by Error Propagation, in: Parallel Distributed Processing, Explorations in the Microstructure of Cognition, Vol. 1. Princeton University Press, pp. 318-362.

[10] J.M. Binner, T. Elger, B. Nilsson and Tepper, J.A. (2006) *Economic Letters*, forthcoming.

[11] J.M. Binner, T. Elger, B. Nilsson and J. Tepper, (2004), *Advances in Econometrics: Applications of Artificial Intelligence in Finance and Economics*, 19: 71-91, Elsevier Science Ltd.

[12] M. Jordan, (1986). Proc of the 8th Annual Conf. of the Cognitive Science Society, 531-545.

[13] J. Elman, (1990), *Cognitive Science* 14, 179-211.

[14] G. Dorffner (1996), *Neural Network World*, 6(4)447-468.

[15] S. Moshiri, N. Cameron and D. Scuse (1999), *Computational Economics*, 14, 219-235.

[16] P. Tenti, (1996). *Applied Artificial Intelligence* 10, 567-581.

[17] C. Ulbricht, (1994), Proc. of the 12th Nat. Conf. in Artificial Intelligence, AAAI Press/MIT Press, Cambridge, MA 883-888.

[18] P.J. Werbos, (1990), in Proc. of IEEE Special Issue on Neural Networks, Vol. 78, pp. 1550-1560.

[19] R.J Williams and J. Peng, (1990), *Neural Computation* 2, 490-501.

[20] Y. Engel, S. Mannor, R. Meir, (2004) *IEEE Transactions on Signal Processing*, 52, (8) 2275-2285.

[21] R. Rosipal, L.J. Trejo (2001) *Journal of Machine Learning Research*, 2, 97-123.