Evolutionary Strategies vs. Neural Networks; An Inflation Forecasting Experiment

Graham Kendall
Department of Computer Science
The University of Nottingham
Nottingham NG8 1BB UK
EMAIL : gxk@cs.nott.ac.uk
Tel : +44 (0)115 846 6514
Fax : +44 (0) 115 951 4249

Jane M Binner and Alicia M Gazely
Department of Finance and Business Information
The Nottingham Trent University
Nottingham NG1 4BU UK
EMAIL : jane.binner|alicia.gazely@ntu.ac.uk
Tel: +44 (0) 115 848 2429; + 44(0) 115 848 2416
Fax: +44 (0) 115 848 6512

Abstract

Previous work has used neural networks to predict the rate of inflation in Taiwan using four measures of 'money' (simple sum and three divisia measures). In this work we develop a new approach that uses an evolutionary strategy as a predictive tool. This approach is simple to implement yet produces results that are favourable with the neural network predictions. Computational results are given.

Keywords: Evolutionary Strategy, Divisia, Inflation Forecasting, Divisia

1. Introduction

In recent years the relationship between 'money' and the macroeconomy has assumed prominence in the academic literature and in Central Banks circles. Although some Central Bankers have stated that they have formally abandoned the notion of using monetary aggregates as indicators of the impact of their policies on the economy, research into the link between some kind of monetary aggregate and the price level is still prevalent. Attention is increasingly turning to the method of aggregation employed in the construction of monetary indices. The most sophisticated index number used thus far relies upon the formulation devised by Divisia (1925). The construction has it roots firmly based in microeconomic aggregation theory and statistical index number theory.

Our hypothesis is that measures of money constructed using the Divisia index number formulation are superior indicators of monetary conditions when compared to their simple sum counterparts. Our hypothesis is reinforced by a growing body of evidence from empirical studies around the world which demonstrate that weighted index number measures may be able to overcome the drawbacks of the simple sum, provided the underlying economic weak separability and linear homogeneity assumptions are satisfied. Ultimately, such evidence could reinstate monetary targeting as an acceptable method of macroeconomic control, including price regulation.

We offer an exploratory study of the relevance of the Divisia monetary aggregate for Taiwan over the period 1978 to date. In this way, we begin with a banking system that was heavily regulated by the Central Bank and the Ministry of Finance until 1989, which saw the introduction of the revised Banking Law in July. At the beginning of the 1980s, drastic economic, social and political changes took place creating a long-term macroeconomic imbalance. Rising oil prices caused consumer prices to rise by 16.3 per cent in 1981, followed by a period of near zero inflation in the mid eighties. From the nineties onwards, inflation has been fluctuating around the 5 per cent mark and hence the control of inflation has not been the mainstay of recent economic policy in Taiwan, unlike the experience of the western world. Rather, policy has focused more on achieving balanced economic and social development.

There have been major financial innovations in Taiwan as transactions technology has progressed and new financial instruments have been introduced, such as interest-bearing retail sight deposits. Although it is difficult to make a distinction between the various types of financial innovation, the effects on the productivity and liquidity of monetary assets are almost certainly different. The question we ask, in keeping with Ford et al (1992) is do the Divisia aggregates adequately capture all the financial innovations?

We adopt the principles of Ford (1997) by allowing both for a period of gradual learning by individuals as they adapt to the financial changes and secondly by incorporating a mechanism to accommodate the changing perceptions of individuals to the increased productivity of money. Individuals are thus assumed to adjust their holdings of financial assets until the diffusion of financial liberalisation is complete.

The novelty of this paper lies in the use of evolutionary strategies (ES) to examine Taiwan's recent experience of inflation. This is an unusual tool in this context and represents the first known application of its kind. Results are compared to those already produced for Taiwan using the Artificial Intelligence technique of neural networks (Binner et al. 2001) to compare the explanatory power of both Divisia and simple sum measures of broad money as indicators of inflation.

The paper concludes with a discussion of the promise of evolutionary strategies as a new tool in the macroeconomic forecasting arena.

2. Evolutionary Strategies

Evolutionary strategies (ES) are closely related to genetic algorithms. Originally they used only mutation, only used a population of a single individual and were used to optimise real valued variables. More recently, ES's have used a population size greater then one, they have used crossover and have also been applied to discrete variables (Bäck, 1991) and (Herdy, 1991). However, their main use is still in finding values for real variables by a process of mutation, rather than crossover.

An individual in an ES is represented as a pair of real vectors, v = (x,σ). The first vector, x, represents a point in the search space and consists of a number of real valued variables. The second vector, σ, represents a vector of standard deviations.

Mutation is performed by replacing x by

$$x^{t+1} = x^t + N(0, \sigma)$$

where N(0, σ) is a random Gaussian number with a mean of zero and a standard deviation of σ. This mimics the evolutionary process that small changes occur more often than larger ones.

In evolutionary computation there are two variations with regard to how the new generation is formed. The first, termed (μ + λ), uses μ parents and creates λ offspring. Therefore, after mutation, there will be μ + λ members in the population. All these solutions compete for survival, with the μ best selected as parents for the next generation. An alternative scheme, termed (μ, λ), works by the μ parents producing λ offspring (where λ > μ). Only the λ compete for survival. Thus, the parents are completely replaced at each new generation. Or, to put it another way, a single solution only has a life span of a single generation. In this work, we use a 1+1 strategy and plan to develop other strategies in the future.

Good introductions to evolutionary strategies can be found in (Bäck et al, 1997), (Fogel, 1998), (Fogel, 2000), (Michalewicz, 1996) and (Michalewicz and Fogel, 2000).

3. Data and Model Specification

Four different monetary measures were used independently to predict future movements in the inflation rate. Monetary data consisted of three Divisia series provided by Ford (1997), one conventional Divisia, (DIVM2), Innovation1 and Innovation2, together with a simple sum series, constructed from component assets obtained from the Aremos-Financial Services database in Taiwan. The Divisia M2 aggregate is constructed by weighting each individual component by its own interest rate whilst Innovation1 (INN1) and Innovation2 (INN2) seek to improve upon the weighting system by capturing the true monetary services flow provided by each component asset more accurately. Thus INN1 is a development of DIVM2 inspired initially by Hendry and Ericsson (1990) and used subsequently by Ford et al (1992). A learning adjustment of the retail sight-deposit interest rate is applied to reflect the adaption of agents to the introduction of interest-bearing sight deposits in 1984. Note: the Innovation1 series does not diverge from the conventional Divisia measure until the late 1980s. The second modified Divisia series, INN2 assumes a period of gradual and continuous learning by agents as they adapt to the increased productivity of money throughout the period and corrects, at least partially, for the distortion arising from technological progress. Individuals are thus assumed to adjust their holdings of financial assets until the diffusion of financial liberalisation is complete.

Inflation was constructed for each quarter as year-on-year growth rates of prices. Quarterly data over the sample period 1970Q1 to 1995Q3 was used as illustrated in Figure 1. Our preferred price series, the Consumer Price Index (CPI), was obtained from Datastream. The four monetary series were subjected to a smoothing process by taking three quarter averages to reduce noise. Finally, to avoid the swamping of mean percent error by huge values during a period of very low inflation from 1983 to 1986, the entire series was translated upwards by 5 percent and results are presented on this basis. Of the total quarterly data points available, after loss of data points due to the smoothing process and the time lag implicit in the model of up to four quarters, 96 quarters remained, of which the first 85 were used for training and the last 7 for were used as a validation set. The first 4 items were only used as a basis for the first prediction.

The aim of the evolutionary strategy is to predict the future inflation rate based on the inflation rate at the previous quarter and the previous four quarters money measure. That is

$$\Pi_t = \mathrm{f}(M_{t\text{-}1}, M_{t\text{-}2}, M_{t\text{-}3}, M_{t\text{-}4}, \Pi_{t\text{-}1})$$

where Π is inflation
t represents time on a quarterly basis
and M is the current money measure being considered (i.e. M2, DIVM2, INN1 or INN2).

The future inflation rate was predicted via a linear function with each term having a weight. Each of the five factors used to predict the next quarter's inflation rate was given an initial, random weight and the task given to the ES was to find a set of weights that allowed future inflation to be predicted. Hence, the function given to the ES was

$$\Pi_t = \mathrm{f}(w_1 M_{t\text{-}1}+w_2 M_{t\text{-}2}+w_3 M_{t\text{-}3}+w_4 M_{t\text{-}4}+w_5 \Pi_{t\text{-}1})$$

where w_n is the weight associated with each term that contributes to the inflation rate for the next quarter.

A given solution (set of weights) is evaluated by predicting the inflation rate for each quarter of the training set. The output from each quarter is compared against the actual value and the absolute sum of those differences is calculated. This figure represents the quality of a given solution. The aim, of course, is to minimise the evaluation function.

If a solution is found that is better than its parent, it becomes the new parent. Each weight in the parent is then mutated as described above (section 2). After a period of training (1,000,000 iterations in this work) the weights were used to predict the inflation rate of the validation set.

4. Testing and Results

The four money measures (M2, DIVM2, INN1 and INN2) were tested independently and these results compared against previous results obtained on the same data using a neural network. The results reported are the arithmetic means calculated over ten individual runs of the ES and are divided between in-sample (the training set) and out-of-sample (the validation set). Within these two categories, three standard forecasting evaluation measures were used to compare the predicted inflation rate with the actual inflation rate, namely, Root Mean Squared Error (RMS), Mean Absolute Difference (MAD) and Mean Percent Error (MPE). The in-sample (a) and out-of-sample (b) results produced by the evolutionary strategies are presented in table 1 alongside previous results (Binner et. al 2001) using neural networks in table 2 below.

(a)	**M2**	**DIVM2**	**INN1**	**INN2**
RMS	0.055654	0.052813	0.052574	0.054517
MAD	0.029751	0.024886	0.024735	0.027855
MPE	25%	19%	18%	22%
(b)	**M2**	**DIVM2**	**INN1**	**INN2**
RMS	0.035367	0.016644	0.016481	0.059012
MAD	0.031768	0.013897	0.013389	0.050765
MPE	36%	16%	15%	59%

Table 1. ES results for full sample, s=85

(a)	**M2**	**DIVM2**	**INN1**	**INN2**
RMS	0.032106	0.022578	0.018764	0.026806
MAD	0.024703	0.017493	0.013878	0.018186
MPE	30%	22%	16%	21%
(b)	**M2**	**DIVM2**	**INN1**	**INN2**
RMS	0.014801	0.016043	0.010715	0.011575
MAD	0.013540	0.014983	0.008021	0.009034
MPE	16%	17%	9%	10%

Table 2. NN results for full sample, s=85

The neural network clearly has superior forecasting capabilities, producing more accurate forecasts using all forecasting evaluation methods both in and out-of sample. The four RMS errors are, on average, twice as great as those produced by the neural network. The best inflation forecast is achieved using the

INN1 monetary aggregate, where the neural network RMS error is 35% lower than the ES equivalent. Figures 1 and 2 below illustrate the best fitting (INN1) forecasts for each technique. Although the neural network is undoubtedly more powerful, the ES method clearly has potential for learning future movements in inflation. INN2 is undoubtedly the worst performing monetary aggregate, producing out-of-sample RMS errors some 3.5 times greater than INN1 on average.

Figure 1: Taiwan Inflation 1972-1995: ES

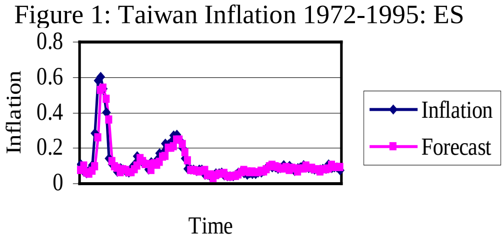
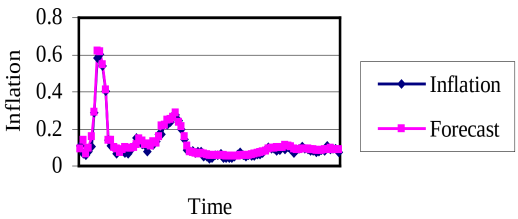


Figure 2: Taiwan Inflation 1972-1995: NN



To investigate and improve the fit of the algorithm further, the training data was split into smaller sample sizes. That is, the data was split into *n* subsets and each subset had a set of weights associated with that data. Initial experiments showed that using smaller subsets does produce results closer to the expected values. Following this insight a series of experiments were conducted where the training data was split into a number of subsets. The size of the subsets, *s*, took the values, *s*={10, 20, 30, 50, 70, 85}.

Tables 3, 4, 5 and 6 show the results when applying this approach to the four data sets. It is interesting to note that, in general, as the value of *s* increases, so does the error. This can be seen in figure 3, which plots the error of the M2 training data set. This graph is indicative of all the data sets. It is also interesting to record that *s*=1 was tested but is not shown as this training data is learnt perfectly, but its predictive capabilities are terrible. This is to be expected as each data sample has its own set of weights, however, not one of them is able to predict accurately.

(a)	**s=10**	**s=20**	**s=30**	**S=50**	**S=70**	**s=85**
RMS	0.029596013	0.0488479	0.0522658	0.0542072	0.054768	0.055654
MAD	0.015665693	0.0251071	0.026575	0.0276298	0.0278926	0.029751
MPE	14%	21%	22%	23%	23%	25%
(b)	**s=10**	**s=20**	**s=30**	**S=50**	**S=70**	**s=85**
RMS	0.03460674	0.0468069	0.0333721	0.0298592	0.0320764	0.035367
MAD	0.029368693	0.0408344	0.0282566	0.026615	0.0277392	0.031768
MPE	33%	48%	31%	30%	32%	36%

Table 3. Different Sample Sizes for M2 data

(a)	**s=10**	**s=20**	**s=30**	**S=50**	**S=70**	**s=85**
RMS	0.029077811	0.0407076	0.0456028	0.0500104	0.0526802	0.052813
MAD	0.013509228	0.0191466	0.0214196	0.0232269	0.0241488	0.024886
MPE	13%	16%	18%	17%	18%	19%
(b)	**s=10**	**s=20**	**s=30**	**S=50**	**S=70**	**s=85**
RMS	0.017450212	0.013862	0.010434	0.01599	0.008025	0.016644
MAD	0.011356283	0.0095	0.0082	0.0136117	0.0061	0.013897
MPE	12%	10%	9%	16%	6%	16%

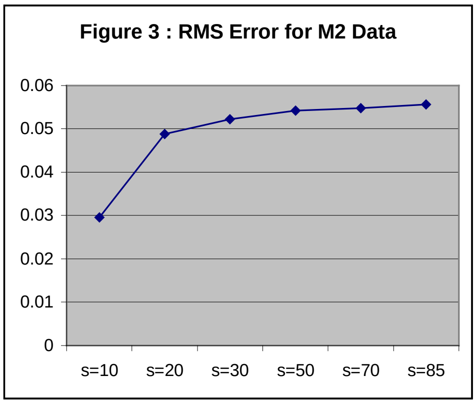
Table 4. Different Sample Sizes for DIVM2 data

(a)	s=10	s=20	s=30	S=50	S=70	s=85
RMS	0.028637794	0.0408193	0.0455279	0.0495153	0.0530051	0.052574
MAD	0.013415464	0.0191138	0.0216154	0.0233427	0.0240428	0.024735
MPE	13%	16%	18%	17%	18%	18%
(b)	**s=10**	**s=20**	**s=30**	**S=50**	**S=70**	**s=85**
RMS	0.019071346	0.023874	0.0158909	0.0162886	0.0160835	0.016481
MAD	0.01558306	0.0207543	0.0136517	0.0138929	0.0131833	0.013389
MPE	17%	23%	15%	16%	14%	15%

Table 5. Different Sample Sizes for INN1 data

(a)	s=10	s=20	s=30	S=50	S=70	s=85
RMS	0.029615464	0.0416456	0.0474273	0.0516671	0.0537119	0.054517
MAD	0.015299391	0.0208679	0.0241918	0.0252718	0.0262228	0.027855
MPE	15%	18%	21%	19%	20%	22%
(b)	**s=10**	**s=20**	**s=30**	**S=50**	**S=70**	**s=85**
RMS	0.104725328	0.1188458	0.04745	0.038448	0.0631411	0.059012
MAD	0.087689038	0.1061943	0.039716	0.032923	0.0539396	0.050765
MPE	103%	124%	46%	37%	63%	59%

Table 6. Different Sample Sizes for INN2 data



Using smaller sample sizes to allow the training data to be learnt more accurately presents a problem in that we need to decide which weights to use to predict the future inflation rate. If the sample size includes the entire training set then there is only one set of weights. If we split the training sample into two sets then there are two sets of weights. We can continue this until the sample size is one, in which case there are *n* sets of weights (where *n* is the size of the training set). We have tried averaging the weights from all the samples, but this yielded very poor results. Therefore, it was decided to use the final set of weights, under the premise that these weights reflect inflation rates that are temporally closer to the prediction set. These results can be seen in part b of tables 3, 4, 5 and 6. However, using the final set of weights may not be the best approach. For example, if the sample size is 20 and the training set, *n*, is 85 then the last set of weights will only include 5 samples on which it has been trained. It may be better to use one of the other set of weights that is evolved using a sample size of 20. Table 7, summarises the predictions for all the weights for out-of-sample data for the INN1 data set with *s*=10. This gives a set of 9 weights with the first 8 being evolved having access to 10 samples. The last set of weights has only been trained on a sample size of 5. It can be seen that the best predictions are from the final set of weights. Furthermore, as we get temporally closer to the figures we are predicting the predictions improve. Table 8 shows that this observation is also true when varying *s* for the INN1 data set. It is interesting to note that when *s*=20, this leads to the worst predictions overall. This is likely to be due to the fact that the fifth set of weights has only been trained on five samples, whereas when *s*=30 the final set of weights has been trained on 25 samples, when *s*=50 the final sample size is 35 and when *s*=70 the sample size, although small, at 15, is still large enough to beat a larger sample size that is temporally disjoint from the data. From these results we conclude that to predict out-of-sample data the weights have to be trained on data that is temporally closer to the data being predicted but, in addition, the sample size on which those weights have been trained has to be large enough to allow the weights to learn the underlying trend. Table 9, shows that these observations are also true of the DIVM2 data. The other data sets also exhibit this behaviour.

	1	2	3	4	5	6	7	8	9
RMS	0.848952	0.213233	0.134902	0.146802	0.035492	0.021814	0.020376	0.014171	0.019071
MAD	0.7798	0.2071	0.1344	0.1417	0.0426	0.0283	0.0290	0.0212	0.0155
MPE	868%	232%	151%	156%	50%	33%	35%	25%	17%

Table 7. INN1 Prediction from set of weights when s=10

	1	2	3	4	5
S=20					
RMS	0.141752	0.076644	0.016340	0.012454	0.023874
MAD	0.1402	0.0796	0.0237	0.0201	0.0207543
MPE	158%	91%	29%	25%	23%
S=30					
RMS	0.057670	0.015914	0.0158909		
MAD	0.0541	0.0229	0.0136517		
MPE	63%	28%	15%		
S=50					
RMS	0.032925	0.0162886			
MAD	0.0391	0.0138929			
MPE	46%	16%			
S=70					
RMS	0.017343	0.0160835			
MAD	0.0241	0.0131833			
MPE	29%	14%			

Table 8. INN1 Prediction from set of weights when s=20, 30, 50 and 70

	1	2	3	4	5	6	7	8	9
S=10									
RMS	0.017435	0.147206	0.105138	0.166307	0.049582	0.026448	0.016889	0.012001	0.017450
MAD	0.0147	0.1447	0.1032	0.1613	0.0480	0.0227	0.0150	0.0105	0.011356
MPE	15%	161%	114%	175%	54%	25%	16%	12%	12%
S=20									
RMS	0.202970	0.044502	0.018806	0.013915	0.013862				
MAD	0.2020	0.0408	0.0147	0.0137	0.0095				
MPE	224%	45%	16%	15%	10%				
S=30									
RMS	0.059102	0.011872	0.010434						
MAD	0.0495	0.0094	0.0082						
MPE	56%	10%	9%						
S=50									
RMS	0.021789	0.01599							
MAD	0.0170	0.013611							
MPE	19%	16%							
S=70									
RMS	0.016074	0.008025							
MAD	0.0126	0.0061							
MPE	14%	6%							

Table 9. DIVM2 Prediction from set of weights when s=10, 20, 30, 50 and 70

5. Concluding Remarks

This research provides the first evidence of its kind to our knowledge to compare the predictive performance of evolutionary strategies with neural networks. Neural networks clearly provide more accurate forecasts over larger sample sizes as they can learn the data perfectly. Evolutionary strategies are found to compete very favourably with neural networks over smaller sample sizes and can also learn the data perfectly in extreme cases. There is clearly scope for further research into the trade off between sample size and predictive accuracy of evolutionary strategies and for the development of this technique as a new macroeconomic forecasting tool.

Evidence presented here also supports the view that Divisia indices appear to offer advantages over simple sum indices as macroeconomic indicators. It may be concluded that a money stock mismeasurement problem exists and that the technique of simply summing assets in the formation of monetary aggregates is inherently flawed. The role of monetary aggregates in the major economies today has largely been relegated to one of a leading indicator of economic activity, along with a range of other macroeconomic variables. However, further empirical work on Divisia money and, in particular, close monitoring of Divisia constructs that have been adjusted to accommodate financial innovation, may serve to restore confidence in former well established money-inflation links. Ultimately, it is hoped that money may be re-established as an effective macroeconomic policy tool in its own right. This application of evolutionary strategies vs. neural network techniques to examine the money - inflation link is highly experimental in nature and in keeping with the pioneering work conducted by two of the current authors for the UK, USA and Italy, the overriding feature of this research is very much one of simplicity. It is virtually certain in this context that more accurate inflation forecasting models could be achieved with the inclusion of additional explanatory variables, particularly those currently used by monetary authorities around the world as leading indicator components of inflation.

References

- Bäck, T., Hoffmeister, F. and Schwefel, H-P. (1991). A survey of evolution strategies. Proceedings of the Fourth Conference on Genetic Algorithms (eds Belew, R. and Booker, L.), Morgan Kaufmann Publishers, San Mateo, CA, pp 2-9.
- Bäck, T., Fogel, D. B. and Michalewicz. 1997. Handbook of evolutionary computation. Oxford University Press. ISBN 0 7503 0392 1
- Binner, J.M., Gazely A.M. and Chen S.H. (2001) Financial innovation in Taiwan: An application of neural networks to the broad money aggregates. European Journal of Finance, under review.
- Divisia, F. (1925) L'indice monétaire et al théorie de la monnaie, Rev. d'Econ. Polit., 39 980-1008.
- Fogel, D.B. (1998) Evolutionary computation the fossil record, IEEE Press, ISBN 0-7803-3481-7
- Fogel, D.B. (2000) Evolutionary computation: toward a new philosphy of machine intelligence, 2nd Ed., IEEE Press Marketing, ISBN 0-7803-5379-X
- Ford, J.L. (1997) Output, the price level, broad money and Divisia aggregates with and without innovation: Taiwan, 1967(1) – 1995(4). Discussion paper 97-17, Department of Economics, The University of Birmingham.
- Ford J.L., Peng, W.S. and Mullineux A.W. (1992) Financial innovation and Divisia monetary aggregates. Oxford Bulletin of Economics and Statistics, pp 87-102.
- Hendry, D.F. and Ericsson, N.R., (1990) Modelling the demand for narrow money in the UK and the United States. Board of Governors of the Federal Reserve System, International Finance Discussion Papers, No. 383.
- Herdy, M. (1991). Application of the evolution strategy to discrete optimization problems. Proceedings of the First International Conference on Parallel Problem Solving from Nature (PPSN), Lecture Notes in Computer Science (eds Schwefel, H-P and Männer, R), Springer-Verlag, Vol. 496, pp 188-192
- Michalewicz, Z. (1996). Genetic algorithms + data structures = evolution programs (3rd rev. and extended ed.). Springer-Verlag, Berlin
- Michalewicz, Z and Fogel, D.B. (2000). How to solve it. Springer-Verlag. ISBN 3-540-66061-5