Inflation Forecasting using a Neural Network

Emi Nakamura [*]
Harvard University

June 21, 2005

Abstract

This paper evaluates the usefulness of neural networks for inflation forecasting. In a pseudo out-of-sample forecasting experiment using recent U.S. data, neural networks outperform univariate autoregressive models on average for short horizons of 1 and 2 quarters. A simple specification of the neural network model and specialized estimation procedures from the neural networks literature appear to play significant roles in the success of the neural network model.

Keywords: Model Selection, Linearity, Forecasting.

JEL Classifications: C51, C52, C53, E37.

[*]I would like to thank Jón Steinsson, James Stock and one anonymous referee for helpful comments and suggestions on this paper. Contact information: Department of Economics, Harvard University, Littauer Center, Cambridge MA 02138. E-mail: nakamura@fas.harvard.edu. Homepage: http://www.harvard.edu/~ nakamura.

1 Introduction

Recently, there has been considerable interest in applications of neural networks in the economics literature, particularly in the areas of financial statistics and exchange rates.[1] In contrast, relatively few studies have applied neural network methods to macroeconomic time series.[2] A limitation of the small number of papers that apply neural networks to macroeconomic applications is that they do not use standard practice neural networks estimation methods, such as "early-stopping" and "pre-processing".[3] However, the investigation of nonlinearities in time series data is important to macroeconomic theory as well as forecasting, as illustrated, for example, in innovative work by Brock and Hommes (1997), Barnett, Medio and Serletis (2003) and others. The most closely related paper in the literature to the present paper is Moshiri and Cameron (2000).

In this note, I investigate the usefulness of a neural network (NN) model for forecasting inflation. I compare the performance of the NN model with that of univariate autoregression models in a pseudo out-of-sample foreasting experiment. In the process, I evaluate the importance of standard-practice NN estimation methods such as early-stopping and pre-processing. Thus, I follow up on Swanson and White's (1997) conjecture that these types of techniques may be of interest given the relative inability other types of model selection criteria such as the Schwarz Information Criterion to pick out the optimal forecasting model.

2 Background and Methodology

Despite its evocative name, a neural network (NN) is simply a parameterized non-linear function that can be fitted to data for prediction purposes. The non-linear function is constructed as a combination of non-linear building blocks, known as transfer functions. A common examples of a NN transfer function is the hyperbolic tangent function. The structure of the NN is described, in neural networks jargon, by the number of "neurons" and "layers" in the NN. These features determine the number and organization of the non-linear transfer functions. Increasing the numbers of transfer functions (adding more neurons and layers) increases the flexibility of the NN.

The main appeal of NN's is their flexibility in approximating a wide range of functional relationships between inputs and outputs. Indeed, sufficiently complex neural networks are

[1]For example, see recent work on financial applications of neural networks by Fernandez-Rodriguez, Gonzalez-Martel and Sosvilla-Rivero (2000) and Refenes and White (1998).

[2]See Stock and Watson (1998), Swanson and White (1997), Chen, Racine and Swanson (2001) and .

[3]An exception is the work by Gonzalez (2000) on forecasting GDP.

able to approximate arbitrary functions arbitrarily well.[4] Thus, there is a close relationship between the NN approach and the older economics literature on flexible functional forms such as the translog function.

The algorithms used to estimate NN's are known as "training algorithms". These algorithms are much like standard minimization routines used, for example, in non-linear least squares. Loosely speaking, the training algorithms iteratively adjust the parameters in the direction of the negative gradient of mean squared error. However, standard-practice NN estimation approaches differ from econometric estimation techniques in important ways. In order to avoid "overfitting", the NN training algorithm is often stopped before a local minimum is reached. Intuitively, overfitting occurs when the NN provides a near-perfect fit in-sample but poor predictions out-of-sample. NN's are thought to be particularly susceptible to overfitting because of their flexibility in approximating different functional forms.

One of the most common types of early stopping procedures is the following cross-validation based approach. First, the data are divided into a training set and a validation set. Next, the training algorithm is run on the training set until the MSE starts to increase on the *validation* set (which usually occurs long before the minimum MSE is reached on the training set).

I estimate a very simple neural network for inflation,

$$\hat{\pi}_{t+j} = L_1 \cdot \tanh(I_1 \cdot x_{t-1} + b1) + L_2 \cdot \tanh(I_2 \cdot x_{t-1} + b2) + b3, \qquad (1)$$

where $\hat{\pi}_{t+j}$ is the NN inflation forecast j quarters in advance, x_{t-1} is a vector of lagged inflation variables $[\pi_{t-1}\, \pi_{t-2}]$, tanh is the hyperbolic tangent function, and L_1, L_2, I_1, I_2, b_1, b_2, and b_3 are parameters. In NN jargon, L_1 and L_2 are "layer weights", I_1 and I_2 are "input weights" and b_1, b_2 and b_3 are "biases". Notice that the NN depends on two lags of inflation. Thus, like Stock and Watson (1998), I consider only univariate inflation forecasting models. Given the limitations of the data, the simple network "architecture" given by (1) was chosen with very minimal search over alternative network architectures.[5]

I train the NN using the Levenberg-Marquardt algorithm, a standard training algorithm from the NN literature. The algorithm is terminated according to the early stopping procedure described above. The validation set used in the early stopping procedure is chosen in a somewhat unusual manner: first, the dataset used to train the NN is divided into "observa-

[4]For example, see Hornik, Stinchcombe and White (1989) for a discussion of the NN "universal approximation" property.

[5]The only search over network architecture was to compare the fit of this NN with NN's using alternative non-linear functions (instead of the hyperbolic tangent function) and up to 4 lags of the inflation variable. Since these modifications had essentially no impact on the results, I do not report the results here.

tions" each consisting of π_t and x_{t-1}, and then every second observation from this dataset is chosen to be part of the validation set.[6] This procedure is related to the innovative approach used by Lebaron and Weigend (1998).

Since local minima (and the areas surrounding them) are considerable problems for NNs, it is standard in the neural networks literature to train the network using a considerable number of random initial values of the parameters, and select only the most successful of these NN's. However, it is also undesirable to "saturate" the parameter space with too many random initial values since this approach, in effect, finds the globally minimizing parameter values by trial and error– counteracting the effects of the early stopping procedure. In order to balance the concerns of "overfitting" and "over-saturation", I train the network using 100 random initial values and select the NN that yields the lowest MSE on the training and validation sets.[7]

I also estimate linear autoregressive (AR) models with lag lengths between 1 and 8 of the form,

$$\hat{\pi}_{t+j} = a_0 + \sum_{i=1}^{k} a_i \pi_{t-i}, \tag{2}$$

where $\hat{\pi}_{t+j}$ is the AR model's inflation forecast j periods in advance, and k is the number of lags included in the model. Rather than using a model selection criterion to select a particular lag length, I simply present results for all the possible lag lengths.

The data are the U.S. GDP deflator from the first quarter of 1960 to the third quarter of 2003. I use a "fixed scheme" pseudo out-of-sample forecasting approach to compare the NN and AR models.[8] Namely, I set aside the last 100 observations of data (i.e. the data for the period 1978q3-2003q1) for testing purposes, and refrain from using this data at any point in the estimation process described above. I then compare the AR and NN models on the basis of their success at forecasting inflation on the test set. Like Swanson and White (1997), I use a model selection approach as opposed to the more traditional hypothesis testing approach of calculating significance levels and confidence intervals.[9]

[6]For example, the first observation is $[x_1, \pi_2]$, the second is $[x_2, \pi_3]$, and so on.

[7]To be more specific, 100 random initial values are drawn for the parameter vector $(L_1, L_2, I_1, I_2, b_1, b_2, b_3)$. The neural network is then re-estimated for each random draw of the parameter vector– yielding 100 alternative parameter estimates. Finally, the parameter value is chosen that corresponds to the lowest MSE calculated on both the training and validation sets. Notice, however, that the resulting parameter vector depends on the particular values of the 100 random draws of the parameter vector. In this sense, the results of a single run of the NN training algorithm are also random. Section 3 reports the results of a Monte Carlo experiment in which the entire NN training algorithm (including the selection of random initial parameter values) is run 400 times.

[8]See McCracken and West (2001) for an overview of these methods.

[9]See Swanson and White (1997) for a discussion of the advantages of the model selection approach relative to the hypothesis testing approach.

3 Results

Table 1 shows the ratio of the mean MSE of the NN model to the MSE's of the AR models on the test set. The NN model has a lower MSE for forecast horizons of 1 and 2 quarters. The NN model has a similar MSE to the AR model for the 3 quarter horizon, and a higher MSE for the 4 quarter horizon. The MSE ratios on the training and validation sets are generally lower than on the test set as a consequence of the overfitting problem described above.

The NN estimator has the unusual property that it generates a distribution of parameter values and MSE's *for the same dataset*, due to the early stopping and random initialization procedures described above. The results in Table 1 are for the mean MSE in a Monte Carlo experiment in which the training algorithm is run 400 times.[10]

The NN training algorithm plays a significant role in the the success of the NN model. Table 2 shows the ratio of the MSE's for the NN estimated by Nonlinear Least Squares (NLLS) versus the MSE estimated by the NN training algorithm described above. In the case of the NLLS minimization algorithm, I use 5000 random initial values as in Stock and Watson (1998)– a considerably larger number than in the NN algorithm. Table 2 shows that NLLS yields a higher MSE than the NN training algorithm for the 1, 2 and 3 quarter horizons, but a lower MSE for the 4 quarter ahead horizon.

Existing applications of NN's to macroeconomic forecasting, such as Stock and Watson (1998) and Swanson and White (1997) find that NN's perform poorly relative to linear models. Table 2 shows that an important reason for the more positive results presented here is the early stopping procedure. Indeed, Table 2 shows that the early stopping procedure makes an important enough contribution to the fit of the NN model for horizons of 1-3 quarters that we would have found little advantage in using the NN without the early stopping procedure– even for short horizons. The early stopping procedure would probably be even more important for more complicated NN's since these NN's suffer more from overfitting.

However, Table 2 also reminds us that the early stopping procedure is not infalliable, as is sometimes suggested in the NN literature. The MSE associated with the 4 quarter ahead forecast is actually higher with the early stopping procedure. Ultimately, early stopping is only one way of avoiding the overfitting problem, and functional form assumptions– such as those imposed by the linear models– are another. While the NN early stopping approach

[10] The standard deviations of the NN estimates of MSE are fairly small, ranging from 0.05 to 0.28 percentage points.

is preferable for short horizon inflation forecasting, the advantage disappears for longer horizons.

4 Conclusion

The existing applications of NN's to macroeconomic forecasting find that NN's perform poorly relative to linear models. I come to a somewhat more positive conclusion regarding the usefulness of NN's for inflation forecasting: the NN model performs well relative to AR models for horizons of 1 and 2 quarters on the test set 1978-2002. My results suggests that the early stopping procedure contributes considerably to the predictive success of the NN approach, and should be incorporated into future forecasting experiments involving NN's. Moreover, simple (e.g. two lag) specifications of neural networks should not be overlooked when data are limited, as is the case for many macroeconomic variables.

References

Barnett, William A., Alfredo Medio, and Apostolos Serletis, "Nonlinear and Complex Dynamics in Economics," 2003. Working Paper.

Brock, William A. and Cars H. Hommes, "A Rational Route to Randomness," *Econometrica*, 1997, *65* (5), 1059–1095.

Chen, Xiaohong, J. Racine, and N. Swanson, "Semiparametric ARX Neural Network Models with an Application to Forecasting Inflation," *IEEE Transactions on Neural Networks*, 2001, *12*, 674–683.

Fernandez-Rodriguez, Fernando, Christian Gonzalez-Martel, and Simon Sosvilla-Rivero, "On the profitability of technical trading rules based on artificial neural networks:: Evidence from the Madrid stock market," *Economics Letters*, 2000, *69* (1), 89–94.

Gonzalez, Steven, "Neural Networks for Macroeconomic Forecasting: A Complementary Approach to Linear Regression Models," 2000. Finance Canada Working Paper 2000-07.

Hornik, K., M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, 1989, *2*, 359–366.

Lebaron, B. and A.S. Weigend, "A bootstrap evaluation of the effect of data splitting on financial time series," *IEEE Transactions on Neural Networks*, 1998, *9* (1), 213–220.

McCracken, M. W. and K. D. West, "Inference About Predictive Ability," in Michael Clements and David Hendry, eds., *A Companion to Economic Forecasting*, Oxford, U.K.: Blackwell Publishers, 2001.

Moshiri, Saeed and Norman Cameron, "Econometrics Versus ANN Models in Forecasting Inflation," *Journal of Forecasting*, February 2000, *19*.

Refenes, A. P. and H. White, "Neural Networks and Financial Economics," *International Journal of Forecasting*, 1998, *6* (17).

Stock, James H. and Mark W. Watson, "A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series," June 1998. NBER Working Paper 6607.

Swanson, Norman R. and Halbert White, "A Model Selection Approach to Real-Time Macroeconomic Forecasting Using Linear Models and Artificial Neural Networks," *The Review of Economics and Statistics*, November 1997, *79* (4), 540–550.

Table 1: MSE Ratio of Neural Network to AR Models

	Test Set				Training/Validation Set			
	Forecast Horizon (Quarters)				**Forecast Horizon (Quarters)**			
	1	**2**	**3**	**4**	**1**	**2**	**3**	**4**
AR1	0.84	0.77	0.98	1.20	0.85	0.77	0.77	0.72
AR2	0.89	0.84	1.04	1.18	0.86	0.81	0.81	0.73
AR3	0.90	0.83	1.01	1.15	0.90	0.85	0.81	0.73
AR4	0.90	0.82	0.96	1.12	0.90	0.85	0.83	0.72
AR5	0.93	0.80	0.94	1.07	0.90	0.87	0.82	0.72
AR6	0.86	0.79	0.91	1.04	0.92	0.86	0.82	0.72
AR7	0.77	0.75	0.84	0.98	0.94	0.87	0.84	0.74
AR8	0.77	0.73	0.83	0.91	0.93	0.88	0.84	0.76

*Each cell shows the ratio of the MSE of the NN model to the MSE of the AR model

Table 2: MSE Ratio of NN estimated with NLLS versus Early Stopping*

	Test Set				Training/Validation Set			
	Forecast Horizon in Quarters				**Forecast Horizon in Quarters**			
	1	**2**	**3**	**4**	**1**	**2**	**3**	**4**
MSE Ratio	1.17	1.16	1.28	0.82	0.94	0.92	0.92	0.99

*Each cell shows the ratio of the MSE of the NN model estimated by NLLS to the MSE of the NN model estimated using early stopping.